Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. The problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. So, in essence, it is not easy to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of the field of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. In fact, machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Two things machine learning can predict, are classification to predict class membership and regression to show numerical values [12]. While data mining is actually part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in the field of prediction [23]. The

Manuscript received August 17, 2021; revised December 6, 2021.

embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

There are many Data Mining methods, including K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known method used for classification [26].

However, it is essential to know the level of accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study, whether the patient should be hospitalized or self-isolated. Therefore, this study also conducted a further test of the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. Testing the accuracy of predicting the treatment status of Covid-19 patients is carried out on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article conducted.

- Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM methods. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict the severity of pneumonia and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and method compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug users using Forward Chaining and Certainty Factor methods [22]. Mean while, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future by using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. The research in this article is different from previous research; the difference lies in the research objectives and the research methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. The research in this article is not using the DT method but uses the RF and SVM methods.
- Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its research objectives from the research in this article. The

previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.

- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.
- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and the research method compared to the research in this article.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. METHODOLOGY

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process consists of a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses confusion matrix and k-fold cross validation in measuring the classification performance of RF and SVM methods.

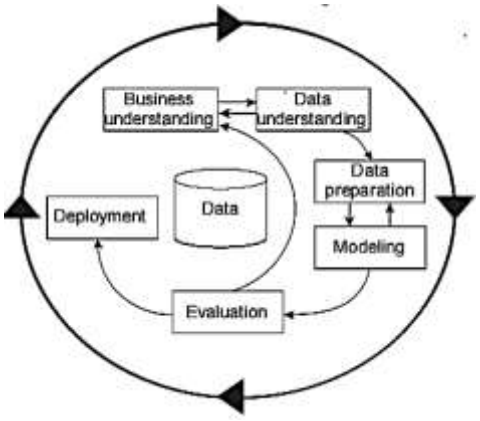

Figure 1. The CRISP-DM Process

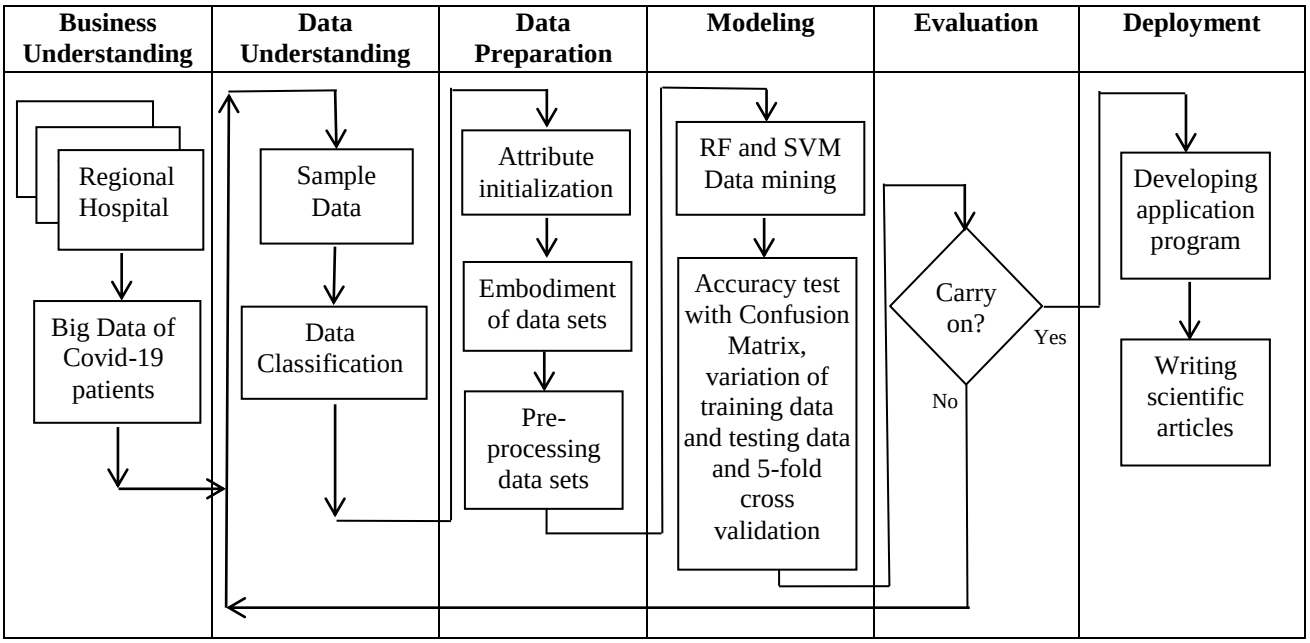

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, symptoms of the disease, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each attribute of the treatment status class. The existing attributes and categories are representations of disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data set of the signs and symptoms and the treatment status of Covid-19 patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from other patients, and some patients have similar diagnoses to others. The number of signs and symptoms of the study or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \text{ x } log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

K-fold cross-validation was also used to measure the performance of the classification method. K-fold cross-validation is one of the cross-validation tests to assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the K k-fold value. In performance testing of this study with k-fold cross validation, the dataset is partitioned into 5 subsets (k = 5) and allows each subset to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The performance evaluation of the model carried out is on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is as shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Performance in predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than using the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

Testing the performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models are presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific articles. Another form of deployment is in the form of making reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestion for further research is to produce a machine learning system model to predict the death and recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.
[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.
[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.
[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.
[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[32] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.
[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check _ Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

By Anthony Anggrawan

WORD COUNT	5582	TIME SUBMITTED	14-JUN-2022 03:54PM
		PAPER ID	87237853

Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. The problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. So, in essence, it is not easy to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of the field of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. In fact, machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Two things machine learning can predict, are classification to predict class membership and regression to show numerical values [12]. While data mining is actually part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in the field of prediction [23]. The

Manuscript received August 17, 2021; revised December 6, 2021.

embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

There are many Data Mining methods, including K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known method used for classification [26].

However, it is essential to know the level of accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study, whether the patient should be hospitalized or self-isolated. Therefore, this study also conducted a further of the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. Testing the accuracy of predicting the treatment status of Covid-19 patients is carried out on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article conducted.

- Askin Kavzoglu, Fusun Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM methods. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict the severity of pneumonia and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and method compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug users using Forward Chaining and Certainty Factor methods [22]. Mean while, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future by using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. The research in this article is different from previous research; the difference lies in the research objectives and the research methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. The research in this article is not using the DT method but uses the RF and SVM methods.
- Sujeet Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its research objectives from the research in this article. The

previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.

- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.
- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and the research method compared to the research in this article.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process consists of a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses confusion matrix and k-fold cross validation in measuring the classification performance of RF and SVM methods.

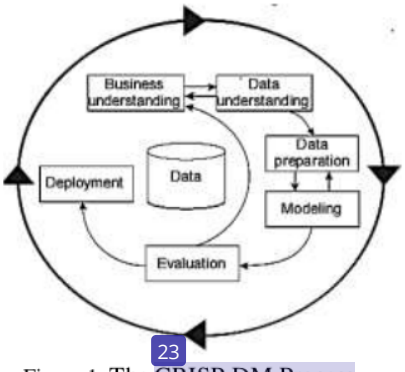

Figure 1. The CRISP-DM Process

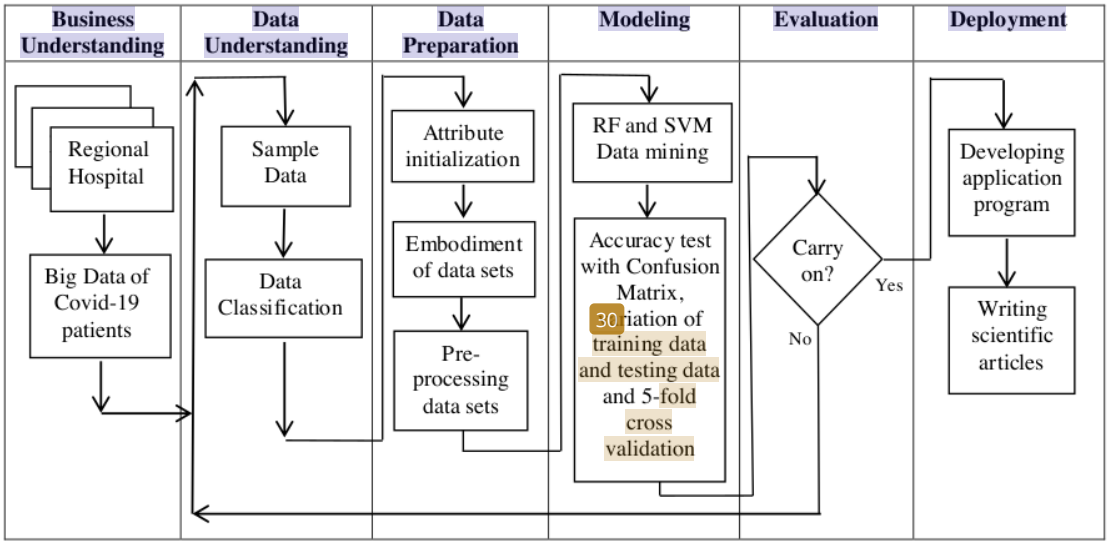

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, symptoms of the disease, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each attribute of the treatment status class. The existing attributes and categories are representations of disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data Set of the Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from other patie and some patients have similar diagnoses to others. The number of signs and symptoms of the study or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \times log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

K-fold cross-validation was also used to measure the performance of the classification method. K-fold cross-validation is one of the cross-validation tests to assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the K k-fold value. In performance testing of this study with k-fold cross validation, the dataset is partitioned into 5 subsets (k = 5) and allows each subset to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The performance evaluation of the model carried out is on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is as shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Performance in predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than using the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

Testing the performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models are presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific articles. Another form of deployment is in the form of making reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestion for further research is to produce a machine learning system model to predict the death and recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al*., "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al*., "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al*., "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al*., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al*., "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al*., "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al*., "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al*., "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[32] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check _ Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

ORIGINALITY REPORT

17%
SIMILARITY INDEX

PRIMARY SOURCES

1	Hairani Hairani, Anthony Anggrawan, M. Ade Candra. "Prediction of Electricity Usage with Back-propagation Neural Network", International Journal of Engineering and Computer Science Applications (IJECSA), 2022 Crossref	66 words — 1%
2	www.jatit.org Internet	66 words — 1%
3	www.ncbi.nlm.nih.gov Internet	66 words — 1%
4	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
5	www.mdpi.com Internet	42 words — 1%
6	Norah Alballa, Isra Al-Turaiki. "Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review", Informatics in Medicine Unlocked, 2021 Crossref	29 words — 1%
7	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision	23 words — 1%

Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018
Crossref

8	thesai.org Internet	23 words — 1%
9	www.researchgate.net Internet	23 words — 1%
10	jurnal.ugm.ac.id Internet	22 words — < 1%
11	"iCAST 2017 proceedings", 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017 Crossref	19 words — < 1%
12	Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 Crossref	19 words — < 1%
13	bura.brunel.ac.uk Internet	19 words — < 1%
14	eprints.utm.my Internet	19 words — < 1%
15	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%

16	annals-csis.org Internet	17 words — < 1%
17	en.wikipedia.org Internet	16 words — < 1%
18	etd.aau.edu.et Internet	13 words — < 1%
19	www.ijicic.org Internet	13 words — < 1%
20	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	12 words — < 1%
21	Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022 Crossref	12 words — < 1%
22	riull.ull.es Internet	12 words — < 1%
23	www.sv-europe.com Internet	12 words — < 1%
24	Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 Crossref	11 words — < 1%

25	arxiv.org Internet	11 words — < 1%
26	dokumen.pub Internet	10 words — < 1%
27	ijece.iaescore.com Internet	10 words — < 1%
28	Zaiwei Song, Yang Hu, Siqian Zheng, Li Yang, Rongsheng Zhao. "Hospital pharmacists' pharmaceutical care for hospitalized patients with COVID-19: Recommendations and guidance from clinical experience", Research in Social and Administrative Pharmacy, 2020 Crossref	9 words — < 1%
29	pdfs.semanticscholar.org Internet	9 words — < 1%
30	123dok.com Internet	8 words — < 1%
31	Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021 Crossref	8 words — < 1%
32	Soham Guhathakurata, Sayak Saha, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A new approach to predict COVID-19 using artificial neural networks", Elsevier BV, 2022 Crossref	8 words — < 1%

33	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
34	annalsofrscb.ro Internet	8 words — < 1%
35	journals.gaftim.com Internet	8 words — < 1%
36	plus.mcmaster.ca Internet	8 words — < 1%
37	www.ajtmh.org Internet	8 words — < 1%
38	www.medrxiv.org Internet	8 words — < 1%
39	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	7 words — < 1%
40	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	7 words — < 1%
41	hdl.handle.net Internet	7 words — < 1%

42 Subrat Kumar Dang, Kulwant Singh. "Predicting tensile-shear strength of nugget using M5P model tree and random forest: An analysis", Computers in Industry, 2021
Crossref

6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

Plagiarism check 18 august 2022_Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT	8082	TIME SUBMITTED	18-AUG-2022 01:00PM
		PAPER ID	89632164

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning predict with high accuracy. For that, or that's why this study aims to propose data mining classification method as a machine learning model to predict the treatment status of Covid-19 nts accurately, whether hospitalized or self-iso d. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—data mining, random forest, support vector machine, prediction, Covid-19, machine learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1]-[3]. It is the cause of the global health crisis [4]-[6], which is not d because of its high-speed transmission [5], [7], but more than 100 million people have died infected worldwide, an re than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the d-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not su sing that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a pat should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both in ent and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16], [17]; moreover, with advances in information and communication technology [18], [19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20]-[22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

accurately [13], and provide high-accuracy prediction results [23]. For this reason (or why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. Machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the treating status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32], [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used in research, including research to classify and predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi *et al*. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this

article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao *et al*. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan *et al*. (2021) implemented machine learning to diagnose drug users and types of drug-using with Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao *et al*. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub *et al*. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata *et al*. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. At the same time, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review.

Pratiyush Guleria *et al*. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev *et al*. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table I compares some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the prior research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table I in the Criteria/Attributes column). So, this study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is on Covid-19 patients from a regional hospital in Mataram, Indonesia. The significant data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19

patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi *et al.* (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao *et al.* (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan *et al.*(2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao *et al.* (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub *et al.* (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata *et al.* (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria *et al.* (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, *et al.* (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev *et al.* (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Fig. 1 [43]. Fig. 2 shows the process carried out at each stage of CRISP-DM.

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the

previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

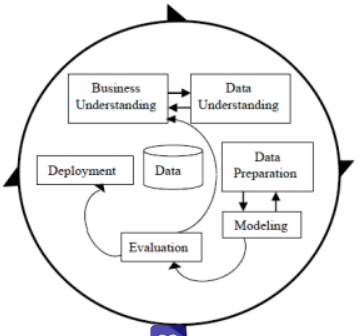

Figure 1. The CRISP-DM process.

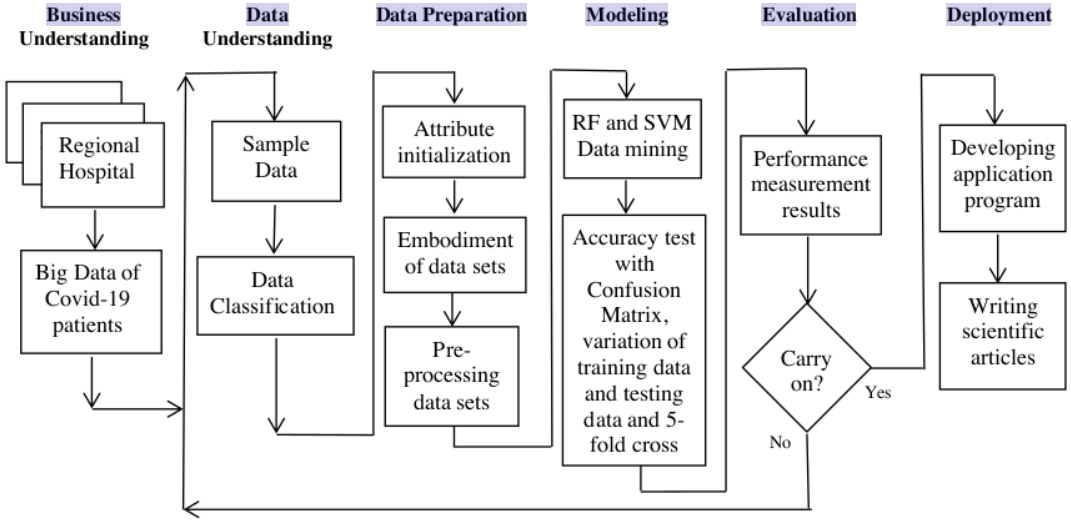

Figure 2. Data mining process of Covid-19 patient big data.

IV. Result and Discussion

A. Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria of patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

TABLE II. Data Set of The Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	.....	....
..	.....	....
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

B. Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data

representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table II shows the association between signs and symptoms of disease and treatment status in the study data set.

C. Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table III).

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table IV. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	...	...	G11	G12	Class
1	Yes	No	No	No	...	...	No	No	Inpatient
2	Yes	Yes	No	Yes	...	...	No	No	Inpatient
3	Yes	No	Yes	No	...	...	No	No	Inpatient
4	Yes	No	No	No	...	...	No	No	Inpatient
5	Yes	No	No	No	...	...	No	No	Inpatient
6	Yes	No	No	No	...	...	Yes	Yes	Inpatient
7	No	No	Yes	No	...	...	No	No	Inpatient
..	...	...	...	...	...	...	...	...	...
..	...	...	...	...	...	...	...	...	...
114	No	No	No	No	...	...	No	Yes	Self-isolation
115	Yes	No	No	No	...	...	Yes	No	Inpatient
116	Yes	No	No	No	...	...	Yes	Yes	Inpatient
117	No	No	No	No	...	...	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table V.

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	...	...	G11	G12	Class
1	1	-1	-1	-1	...	...	-1	-1	-1
2	1	1	-1	1	...	...	-1	-1	-1
3	1	-1	1	-1	...	...	-1	-1	-1
4	1	-1	-1	-1	...	...	-1	-1	-1
5	1	-1	-1	-1	...	...	-1	-1	-1
6	1	-1	-1	-1	...	...	1	1	-1
7	-1	-1	1	-1	...	...	-1	-1	-1
..	...	...	...	...	...	...	...	...	...
..	...	...	...	...	...	...	...	...	...
114	-1	-1	-1	-1	...	...	-1	1	1
115	1	-1	-1	-1	...	...	1	-1	-1
116	1	-1	-1	-1	...	...	1	1	-1
117	-1	-1	-1	-1	...	...	-1	-1	-1

D. Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, known various programming language [44], which has their respective advantages in building application programs [45], [46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction (as show in Table VI).

TABLE VI. USE OF HYPERPARAMETER ON SVM AND RF METHOD

Classifier Method	Hyperparameter	Value
SVM	C	1
	Kernel	Rbf
	Degree	3
	Gamma	Scale
	Coef	0
	Tol	0,000 1
	ax_iter	-1
RF	n_estimators	100
	Criterion	Gini
	Max_depth	None
	Min_samples_split	2
	Min_samples_leaf	1

1) SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: 1) Forming a linear equation from the training data that has gone through the preprocessing stage; 2) Finding the values of w and b by means of elimination and substitution of linear equations; 3) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for $yi = +1, -1$. The formula of the SVM data mining method is: 1) to form a linear equation

from the training data; 2) find the value of w and b, and 3) the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1 \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

2) RF data mining method

The process of realizing the classification using the RF data mining method is as follows: 1) Generating a random subset of data; 2) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and 3) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:

S = Set of cases

pi = the proportion of case i to the Set of cases

3) Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (*TP*) or positive data detected correctly; False Positive (*FP*) or negative data detected is positive; True Negative (*TN*) or negative data detected correctly, and False Negative (*FN*) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcome. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table VII. The results of the predictions of the SVM and RF methods are shown in Tables VIII and IX.

TABLE VII. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

TABLE VIII. CONFUSION MATRIX OF SVM

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	2	12

TABLE IX. CONFUSION MATRIX OF RF

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	1	13

4) K-fold Cross-validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

E. Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

1) Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in

le X and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE X. . SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

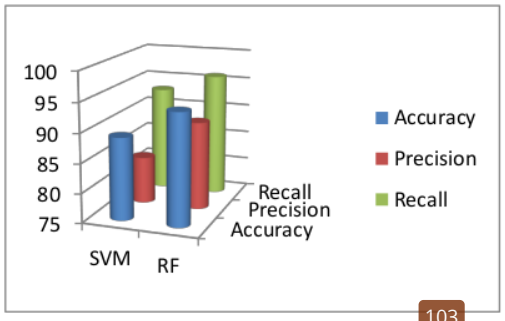

Figure 3. System model performance testing with 85% of training data and 15% of testing data.

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table XI.

TABLE XI. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

2) Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table XII and Fig. 4.

TABLE XII. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

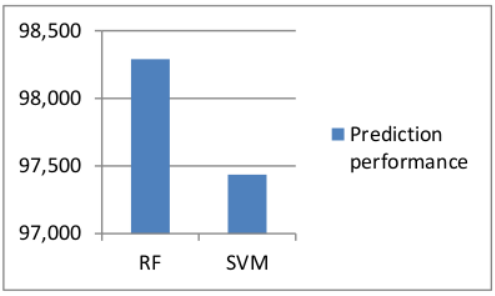

Figure 4. Predictive performance testing using K-fold cross-validation.

F. *Deployment*

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of

various other fields of science including patients' care status, with a combination of machine learning and the Internet of Things.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

References

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV infection from an Asymptomatic contact in germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral load in upper respiratory specimens of infected patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 7, pp. 1–21, September 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2914–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital length of stay and risk of death using artificial intelligence-based modeling," *Front. Med.*, vol. 8, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, pp. 281–286, April 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive flexibility and adaptive decision-making: evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System m for prediction analytics using k-nearest neighbors algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical study on classifiers for earlier prediction of covid-19 infection cure and death rate in the indian states," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, pp. 1–11, May 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of online learning versus face to face learning in english grammar learning," in *The 5th International Conference on Computing Engineering and Design*, 2018, pp. 1–4.

[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[20] H. Zhang *et al.*, "A tool to early predict severe corona virus disease 2019 (COVID-19) : A multicenter study using the risk nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, pp. 1–17, 2 May 020.

[21] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[22] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine learning for diagnosing drug users and types of drugs used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[26] C. W. Yancy, "From mitigation to containment of the COVID-19 pandemic: putting the SARS-CoV-2 genie back in the bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, pp. 50–67, February 2017.

[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data modeling and data analytics: a survey from a big data perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[29] B. Suvarnamukhi and M. Seshashayee, "Big data concepts and techniques in data processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[30] Y. Li, L. Guo, and Y. Guo, "An efficient and performance-aware big data storage system yang," in *Communications in Computer and Information Science*, vol. 367 CCIS, pp. 214–217, April 2018.

[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, pp. 2–6 November 2020.

[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, pp. 1–12, 2017.

[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support vector machine versus random forest for remote sensing image classification: A meta-analysis and systematic review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[34] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–844, 2020.

[36] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[37] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4, pp. 1–14, April 2021.

[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[39] Mehrotra and R. Agarwal, "A review of use of data mining during COVID-19 pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship recipients recommendation system using AHP and moora methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[41] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–4, 2022.
[42] O. Marban, G. Mariscal, and J. Segovia, "A data mining & knowledge discovery process model,"*Austria: Data Mining and Knowledge Discovery in Real Life Applications*, 2009, p. 438.
[43] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, January 2000.
[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Darniani, "Reciprocity effect between cognitive style and mixed learning method on computer programming skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[45] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between cognitive style and gender of two hybrid learning to learning achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based garbage container system using NodeMCU ESP32 microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Anthony Anggrawan currently works as an associate professor in the Department of Computer Science as a lecturer, university Rector, and State Civil Apparatus at Bumigora University, Indonesia. In addition, he is a reviewer of articles in several reputable international scientific journals. He received his Masters in Computer Science Information Technology (M.T) from the 10 November Institute of Technology, Surabaya, Indonesia. After that, he earned his Doctoral degree (Ph.D.) in Accounting Information Systems from Universiti Utara Malaysia. Then, he received his second Doctoral degree (Dr.) from Hasanuddin University, Makassar, Indonesia, in linguistics. Finally, he earned his third Doctorate in Educational Technology from the State University of Jakarta. His research interests include Information Technology, Data Mining, Machine Learning, Online Learning, and the Internet of Things. From 2016 until now, he has been the chairman of the Association for Higher Education in Informatics and Computers (APTIKOM) for the West Nusa Tenggara region, Indonesia.

Mayadi obtained a bachelor's degree (S.Kom) in Computer Science from Bumigora University, Indonesia, and a master's degree (M.Kom) in Informatics Engineering from Amikom University, Indonesia. Currently a lecturer in the Computer Science Study Program at Bumigora University, Indonesia, and a Vice Chancellor III. His research interests are Artificial intelligence, Machine Learning, Data Mining, Big Data, Deep learning, the Internet of Things, and Data Science.

Christofer Satria received a bachelor's degree (S.Sn) in Visual Communication Design from Petra Christian University, Surabaya, Indonesia, and a master's degree (M.Sn) in Visual Communication Design from the Indonesian Art Institute Denpasar, Bali, Indonesia. He is currently a lecturer in the Visual Communication Design Study Program at Bumigora University, Indonesia, and the head of the laboratory in photography, animation, and video. His research interests include animation learning media, video learning media, education method, Data Mining, and experimental Design. He is currently pursuing a doctorate in the same area as his expertise.

Bambang Krismono Triwijoyo currently works as an assistant professor Department of Computer Science as a lecturer and head of the quality assurance institute at Bumigora University, Indonesia. In addition, he is a reviewer of articles in several reputable international scientific journals. He completed his post-graduation in Computer Science from the Sepuluh November Institute of Technology, Surabaya, Indonesia, in 2003. Doctoral degree in Computer Science from Bina Nusantara University, Jakarta, Indonesia, in 2021. His research interests are Computer Vision, Image Processing, and Machine Learning. He is also active in several professional organizations, including IEEE Computer Science, the Indonesian Informatics Experts Association (IAII), and the Indonesian Pattern Recognition Association (INAPR).

Ria Rismayati obtained a bachelor's degree (S.Kom) in Informatics Engineering from the College of Management and Computer Informatics, Mataram, Indonesia, and a master's degree (M.Kom) from AMIKOM University, Yogyakarta, Indonesia. Currently, she is a lecturer in the Computer Science program at Bumigora University, Indonesia. Her research interests include Data Mining, Artificial Intelligence, Enterprise Architecture, and Information Systems.

Plagiarism check 18 august 2022_Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

31%
SIMILARITY INDEX

PRIMARY SOURCES

1	thesai.org Internet	165 words — 2%
2	inass.org Internet	111 words — 1%
3	www.jatit.org Internet	109 words — 1%
4	Anthony Anggrawan, Dwi Kurnianingsih, Christofer Satria. "Sistem Aplikasi Cerdas Klasterisasi Penerima Bantuan Covid-19", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Crossref	80 words — 1%
5	journal.universitasbumigora.ac.id Internet	77 words — 1%
6	umpir.ump.edu.my Internet	54 words — 1%
7	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
8	www.aimspress.com Internet	48 words — 1%

9	ijair.id Internet	45 words — 1%
10	www.bmj.com Internet	44 words — 1%
11	Fermi Dwi Wicaksono, Yusri Bin Arshad, Haeryip Sihombing. "Norm-dist Monte-Carlo integrative method for the improvement of fuzzy analytic hierarchy process", Heliyon, 2020 Crossref	42 words — 1%
12	S V Deshmukh, A Roy. "An Empirical Exploration of Artificial Intelligence in Medical Domain for Prediction and Analysis of Diabetic Retinopathy: Review", Journal of Physics: Conference Series, 2021 Crossref	40 words — < 1%
13	Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021 Crossref	37 words — < 1%
14	egonw.github.io Internet	34 words — < 1%
15	Farshid Rahimibashar, Ladan Sedighi, Alireza Shahriary, Zeljko Reiner et al. "Is there any association between plasma lipid profile and severity of COVID-19?", Clinical Nutrition ESPEN, 2022 Crossref	33 words — < 1%
16	philpapers.org Internet	33 words — < 1%

17	smartcities.ieee.org Internet	33 words — < 1%
18	www.e-journal.unair.ac.id Internet	33 words — < 1%
19	www.hindawi.com Internet	33 words — < 1%
20	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	32 words — < 1%
21	Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021 Crossref	32 words — < 1%
22	jurnal.ugm.ac.id Internet	30 words — < 1%
23	www.ncbi.nlm.nih.gov Internet	29 words — < 1%
24	saucis.sakarya.edu.tr Internet	28 words — < 1%
25	www.mdpi.com Internet	28 words — < 1%
26	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet	27 words — < 1%

27 Fairuz Amalina, Ibrahim Abaker Targio Hashem, Zati Hakim Azizul, Ang Tan Fong, Ahmad Firdaus, Muhammad Imran, Nor Badrul Anuar. "Blending Big Data Analytics: Review on Challenges and a Recent Study", IEEE Access, 2020
Crossref
27 words — < 1%

28 Geng-Xin Xu, Chen Liu, Jun Liu, Zhongxiang Ding et al. "Cross-Site Severity Assessment of COVID-19 from CT Images via Domain Adaptation", IEEE Transactions on Medical Imaging, 2021
Crossref
27 words — < 1%

29 www.gtu.edu.tr
Internet
27 words — < 1%

30 www.researchgate.net
Internet
27 words — < 1%

31 Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A novel approach to predict COVID-19 using support vector machine", Elsevier BV, 2021
Crossref
26 words — < 1%

32 Tawsifur Rahman, Amith Khandakar, Md Enamul Hoque, Nabil Ibtehaz et al. "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access, 2021
Crossref
26 words — < 1%

33 Bhabani Shankar Prasad Mishra, Om Pandey, Satchidananda Dehuri, Sung-Bae Cho. "Unsupervised Functional Link Artificial Neural Networks for Cluster Analysis", IEEE Access, 2020
Crossref
25 words — < 1%

34	Joshua Guzman, Gwen Arianne Recoco, Al Wahid Pandi, Jerico M. Padrones, Jonathan Jared Ignacio. "Evaluating workplace safety in the oil and gas industry during the COVID-19 pandemic using occupational health and safety Vulnerability Measure and partial least square Structural Equation Modelling", Cleaner Engineering and Technology, 2022 Crossref	25 words — < 1%
35	Vimal Kamleshkumar Bhatt. "Assessing the Significance and Impact of Artificial Intelligence and Machine Learning in Placement of Advertisements", 2021 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), 2021 Crossref	25 words — < 1%
36	ijsrcseit.com Internet	25 words — < 1%
37	seaninstitute.org Internet	24 words — < 1%
38	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018 Crossref	23 words — < 1%
39	Tora Fahrudin, Nelsi Wisna, Patrick Adolf Telnoni, Dedy Rahman Wijaya. "Sales Forecasting Web Application in Small and Medium Enterprise", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022 Crossref	23 words — < 1%
40	Navid Rajil, Alexei Sokolov, Zhenhuan Yi, Garry Adams et al. "A fiber optic–nanophotonic	21 words — < 1%

approach to the detection of antibodies and viral particles of COVID-19", Walter de Gruyter GmbH, 2021
Crossref

41 Anindita Sarkar Mondal, Samiran Chattopadhyay, Sarmistha Neogy, Nandini Mukherjee. "Object based schema oriented data storage system for supporting heterogeneous data", 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), 2016
Crossref
20 words — < 1%

42 Qingyun Ren, Bingyin Zhou, Liang Tian, Wei Guo. "Detection of COVID-19 With CT Images Using Hybrid Complex Shearlet Scattering Networks", IEEE Journal of Biomedical and Health Informatics, 2022
Crossref
20 words — < 1%

43 repository.petra.ac.id
Internet
20 words — < 1%

44 Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita. "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review", Computational and Structural Biotechnology Journal, 2021
Crossref
19 words — < 1%

45 eprints.utm.my
Internet
19 words — < 1%

46 harvest.usask.ca
Internet
19 words — < 1%

47 www.rc.auth.gr
Internet
19 words — < 1%

48	Antonia Morita Iswari Saktiawati, Kuwat Triyana, Siska Dian Wahyuningtias, Bintari Dwihardiani et al. "eNose-TB: A trial study protocol of electronic nose for tuberculosis screening in Indonesia", PLOS ONE, 2021 Crossref	18 words — < 1%
49	I Ayuningtias, I Jaya, M Iqbal. "Identification of yellowfin tuna (Thunnus albacares), mackerel tuna (Euthynnus affinis), and skipjack tuna (Katsuwonus pelamis) using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 Crossref	18 words — < 1%
50	files.covid19treatmentguidelines.nih.gov Internet	17 words — < 1%
51	aging-us.com Internet	16 words — < 1%
52	journals.gaftim.com Internet	16 words — < 1%
53	researchprofiles.herts.ac.uk Internet	16 words — < 1%
54	www.iseg.ulisboa.pt Internet	16 words — < 1%
55	jtsiskom.undip.ac.id Internet	15 words — < 1%
56	www.hippocampus.si Internet	15 words — < 1%
57	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020	14 words — < 1%

Crossref

58 Sajana Maharjan, Faisal Mueen Qamer, Mir Matin, Govinda Joshi, Sanjeev Bhuchar. "Integrating Modelling and Expert Knowledge for Evaluating Current and Future Scenario of Large Cardamom Crop in Eastern Nepal", Agronomy, 2019 — 14 words — < 1%
Crossref

59 Roberto Magherini, Elisa Mussi, Michaela Servi, Yary Volpe. "Emotion recognition in the times of COVID19: coping with face masks", Intelligent Systems with Applications, 2022 — 13 words — < 1%
Crossref

60 "Biographies for Volume 25, No. 5", Journal of Accounting and Public Policy, 2006 — 12 words — < 1%
Crossref

61 Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19", Physiological Genomics, 2020 — 12 words — < 1%
Crossref

62 Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 — 12 words — < 1%
Crossref

63 Kuncahyo Setyo Nugroho, Anantha Yullian Sukmadewa, Angga Vidianto, Wayan Firdaus Mahmudy. "Effective predictive modelling for coronary artery diseases using support vector machine", IAES International Journal of Artificial Intelligence (IJ-AI), 2022 — 12 words — < 1%
Crossref

64 S A Pamungkas, I Jaya, M Iqbal. "Segmentation of Enhalus acoroides seagrass from underwater images using the Mask R-CNN method", IOP Conference Series: Earth and Environmental Science, 2021
Crossref
12 words — < 1%

65 Spiridon Vergis, Konstantinos Bezas, Themis P. Exarchos. "Chapter 3 Decision Support System for Breast Cancer Detection Using Biomarker Indicators", Springer Science and Business Media LLC, 2021
Crossref
12 words — < 1%

66 Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022
Crossref
12 words — < 1%

67 ejournal.uin-suska.ac.id
Internet
12 words — < 1%

68 www.monstersjournal.net
Internet
12 words — < 1%

69 Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021
Crossref
11 words — < 1%

70 Yoga Pristyanto, Anggit Ferdita Nugraha, Irfan Pratama, Akhmad Dahlan. "Ensemble Model Approach For Imbalanced Class Handling on Dataset", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020
11 words — < 1%

Crossref

71	arxiv.org Internet	11 words — < 1%
72	docplayer.net Internet	11 words — < 1%
73	epdf.pub Internet	11 words — < 1%
74	github.com Internet	11 words — < 1%
75	www.covprecise.org Internet	11 words — < 1%
76	Chuan Ma, Jun Li, Ming Ding, Kang Wei, Wen Chen, H. Vincent Poor. "Federated Learning With Unreliable Clients: Performance Analysis and Mechanism Design", IEEE Internet of Things Journal, 2021 Crossref	10 words — < 1%
77	F Rahmad, Y Suryanto, K Ramli. "Performance Comparison of Anti-Spam Technology Using Confusion Matrix Classification", IOP Conference Series: Materials Science and Engineering, 2020 Crossref	10 words — < 1%
78	Mutasem Khalil Alsmadi. "Modified SEIR and machine learning prediction of the trend of the epidemic of COVID-19 in Jordan under lockdowns impact", International Journal of Electrical and Computer Engineering (IJECE), 2022 Crossref	10 words — < 1%

79 Qiao Yang, Jixi Li, Zhijia Zhang, Xiaocheng Wu et al. "A Machine Learning Model to Predict Death Outcome in Severe COVID-19 Patients", Research Square Platform LLC, 2020
Crossref Posted Content
10 words — < 1%

80 mdpi-res.com
Internet
10 words — < 1%

81 www.semanticscholar.org
Internet
10 words — < 1%

82 Dennis Peace Ezeobi, William Wasswa, Angella Musimenta, Stella Kyoyagala. "A Neonatal Sepsis Prediction Algorithm Using Electronic Medical Record (EMR) Data from Mbarara Regional Referral Hospital (MRRH)", Research Square Platform LLC, 2022
Crossref Posted Content
9 words — < 1%

83 Gonzalo Mariscal, Óscar Marbán, Covadonga Fernández. "A survey of data mining and knowledge discovery process models and methodologies", The Knowledge Engineering Review, 2010
Crossref
9 words — < 1%

84 I Gede Susrama Mas Diyasa, Agus Prayogi, Intan Yuniar Purbasari, Ariyono Setiawan, Sugiarto, Prismahardi Aji Riantoko. "Data Classification of Patient Characteristics Based on Nutritional Treatment Using the K-Nearest Neighbors Algorithm", 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), 2021
Crossref
9 words — < 1%

85 Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. ""F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability",
9 words — < 1%

2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021
Crossref

86	S H Heriwati*, B H Prilosadoso, B Pujiono, Suwondo, A N Panindias. "3D Puppets Animation for Encouraging Character Education and Culture Preservation in Surakarta", International Journal of Engineering and Advanced Technology, 2019 Crossref	9 words — < 1%
87	Shaker El-Sappagh, Farman Ali, Samir El-Masri, Kyehyun Kim, Amjad Ali, Kyung-Sup Kwak. "Mobile Health Technologies for Diabetes Mellitus: Current State and Future Challenges", IEEE Access, 2019 Crossref	9 words — < 1%
88	docplayer.com.br Internet	9 words — < 1%
89	journalofbigdata.springeropen.com Internet	9 words — < 1%
90	jurnal.iaii.or.id Internet	9 words — < 1%
91	thescipub.com Internet	9 words — < 1%
92	www.medrxiv.org Internet	9 words — < 1%
93	www.science.gov Internet	9 words — < 1%
94	www.upgrad.com Internet	9 words — < 1%

95 "Cloud Computing and Services Science", Springer Science and Business Media LLC, 2013
Crossref
8 words — < 1%

96 Ashutosh Kumar Dubey, Sushil Narang, Abhishek Kumar, Satya Murthy Sasubilli, Vicente Garc鞾-D鞾z. "Performance Estimation of Machine Learning Algorithms in the Factor Analysis of COVID-19 Dataset", Computers, Materials & Continua, 2021
Crossref
8 words — < 1%

97 M. Fatih Aslan, Kadir Sabanci, Ewa Ropelewska. "A CNN-Based Solution for Breast Cancer Detection With Blood Analysis Data: Numeric to Image", 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021
Crossref
8 words — < 1%

98 Rachmadita Andreswari, Dita Aprilia Rahmani, Ria Rahmawati, Achmad Akbar H. "Comparative Analysis of K-Means and K-Nearest Neighbor Algorithm for Telecom Fraud Detection", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022
Crossref
8 words — < 1%

99 Sushil Kumar Mishra, Nitish Kumar Ojha. "A Novel Approach for Face Mask Detection using Tensorflow for Covid-19", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021
Crossref
8 words — < 1%

100 annalsofrscb.ro
Internet
8 words — < 1%

101 collections.plos.org
Internet
8 words — < 1%

102	edulearn.intelektual.org Internet	8 words — < 1%
103	eprints.umm.ac.id Internet	8 words — < 1%
104	etd.aau.edu.et Internet	8 words — < 1%
105	hdl.handle.net Internet	8 words — < 1%
106	ijai.iaescore.com Internet	8 words — < 1%
107	ojs.trp.org.in Internet	8 words — < 1%
108	openaccess.altinbas.edu.tr Internet	8 words — < 1%
109	plus.mcmaster.ca Internet	8 words — < 1%
110	repository.ju.edu.et Internet	8 words — < 1%
111	www.ajtmh.org Internet	8 words — < 1%
112	www.ijisae.org Internet	8 words — < 1%
113	www.ijiser.com Internet	8 words — < 1%
	www.mcsul.c3.furg.br	

114 Internet

8 words — < 1%

115 Anita Desiani, Adinda Ayu Lestari, M Al-Ariq, Ali Amran, Yuli Andriani. "Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification", Pattimura International Journal of Mathematics (PIJMath), 2022

Crossref

7 words — < 1%

116 Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021

Crossref

7 words — < 1%

117 Annisa Nurul Puteri, Dewiani, Zulkifli Tahir. "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019

Crossref

6 words — < 1%

118 David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016

Crossref

6 words — < 1%

119 Li, Dan, Yinghai Ke, Huili Gong, and Xiaojuan Li. "Object-Based Urban Tree Species Classification Using Bi-Temporal WorldView-2 and WorldView-3 Images", Remote Sensing, 2015.

Crossref

6 words — < 1%

120 Mohammed. "Feature Discretization for Individuality Representation in Twins Handwritten Identification", Journal of Computer Science, 2011 — 6 words — < 1%
Crossref

121 Muhammad Daniyal, Kassim Tawiah, Sara Muhammadullah, Kwaku Opoku-Ameyaw. "Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data", Journal of Healthcare Engineering, 2022 — 6 words — < 1%
Crossref

122 Shilei Lu, Yang Liu, Yuwei Li, Ran Wang. "Multidimensional performance-based evaluation method of high-performance cold source in green building", Energy and Buildings, 2021 — 6 words — < 1%
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF

Similarity Check _ Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

By Anthony Anggrawan

WORD COUNT	5582	TIME SUBMITTED	14-JUN-2022 03:54PM
		PAPER ID	87237853

Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. The problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. So, in essence, it is not easy to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of the field of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. In fact, machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Two things machine learning can predict, are classification to predict class membership and regression to show numerical values [12]. While data mining is actually part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in the field of prediction [23]. The

Manuscript received August 17, 2021; revised December 6, 2021.

embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

There are many Data Mining methods, including K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, nam[39] supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known method used for classification [26].

However, it is essential to know the level of accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study, whether the patient should be hospitalized or self-isolated. Therefore, this study also conducted a further [62] of the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. Testing the accuracy of predicting the treatment status of Covid-19 patients is carried out on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article conducted.

- Askin Kavzoglu, Fu[9]n Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms [3]th RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celesti[19]Iwendi et al. (2020) proposed the Random Forest model to predict [4]e disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research[45]ly used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM methods. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study c[4]sified breast cancer as patients with breast cancer. In contrast, [71]esearch in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao [19] et al. (2020) developed a model to predict the severity of pneumonia and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research di[3]s in the research purpose and method compared to the research in this article.
- Anthony Anggrawan et al.[1](2021) implemented machine learning to diagnose drug users and types of drug users using Forward Chaining [45]d Certainty Factor methods [22]. Mean while, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model [55] to predict the number of cases of Covid-19 patients in the future by using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compa[19] to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. The research in this article is different from previous research; the difference lies in the research objectives and the [17]earch methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Cov[4]19 patients should be hospitalized or self-isolated. The research in this article is not using the DT method but uses the RF and SVM methods.
- S[65]m Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its research objectives from the research in this article. The

previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.

- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.
- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and the research method compared to the research in this article.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process consists of a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses confusion matrix and k-fold cross validation in measuring the classification performance of RF and SVM methods.

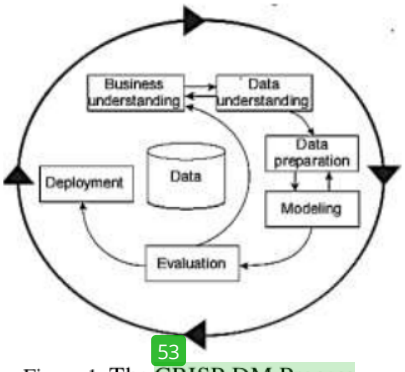

Figure 1. The CRISP-DM Process

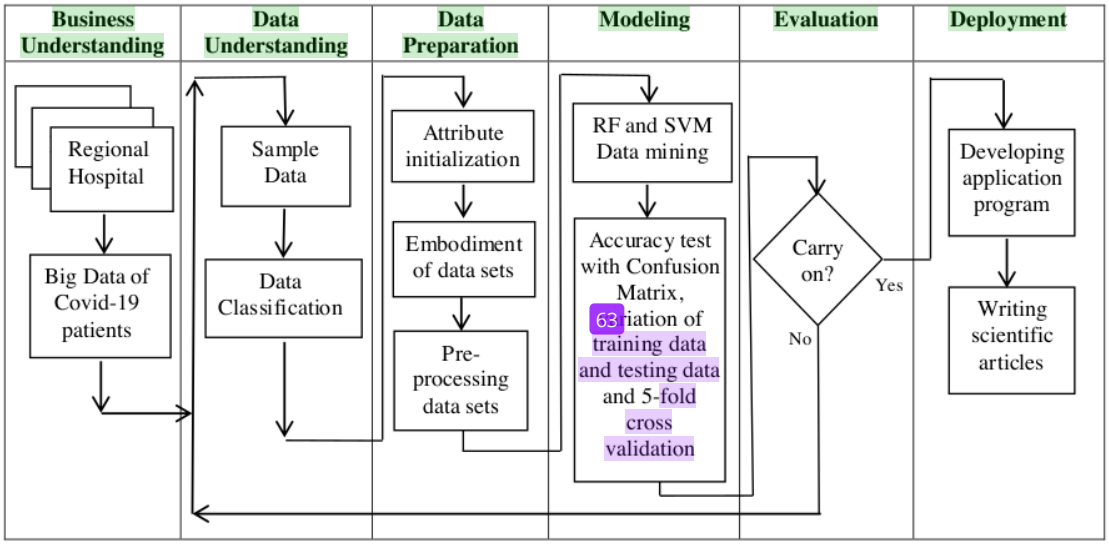

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, symptoms of the disease, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each attribute of the treatment status class. The existing attributes and categories are representations of disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data Set of the Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from other patients and some patients have similar diagnoses to others. The number of signs and symptoms of the study or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \times log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

K-fold cross-validation was also used to measure the performance of the classification method. K-fold cross-validation is one of the cross-validation tests to assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the K k-fold value. In performance testing of this study with k-fold cross validation, the dataset is partitioned into 5 subsets (k = 5) and allows each subset to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The performance evaluation of the model carried out is on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is as shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Performance in predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than using the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

Testing the performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models are presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific articles. Another form of deployment is in the form of making reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestion for further research is to produce a machine learning system model to predict the death and recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–12, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, 12, no. 11, pp. 111–118, 2021.
[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. vember, pp. 1–9, 2020.
[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging urosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. –6325, 2020.
[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, July, pp. 1–9, 2020.
[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815– 2020.
[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in nts with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[32] Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[33] Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipien Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 202
[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, . 1–17, 2022.
[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, ustria: Data Mining and Knowledge Discovery in Real Life plications, 2009, p. 438.
[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining ide*. SPSS, 2000.
[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Lear Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check _ Covid-19 Patient Care Status Using Random Forest and Support Vector Machine Methods

ORIGINALITY REPORT

31%
SIMILARITY INDEX

PRIMARY SOURCES

1	thesai.org Internet	148 words — 3%
2	journal.universitasbumigora.ac.id Internet	126 words — 2%
3	www.jatit.org Internet	79 words — 1%
4	Hairani Hairani, Anthony Anggrawan, M. Ade Candra. "Prediction of Electricity Usage with Back-propagation Neural Network", International Journal of Engineering and Computer Science Applications (IJECSA), 2022 Crossref	59 words — 1%
5	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
6	Haodong Li. "An Overview on Remote Sensing Image Classification Methods with a Focus on Support Vector Machine", 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), 2021 Crossref	45 words — 1%
7	Fermi Dwi Wicaksono, Yusri Bin Arshad, Haeryip Sihombing. "Norm-dist Monte-Carlo integrative	42 words — 1%

method for the improvement of fuzzy analytic hierarchy process", Heliyon, 2020
Crossref

8 www.ijeat.org
Internet
42 words — 1%

9 www.mdpi.com
Internet
42 words — 1%

10 S V Deshmukh, A Roy. "An Empirical Exploration of Artificial Intelligence in Medical Domain for Prediction and Analysis of Diabetic Retinopathy: Review", Journal of Physics: Conference Series, 2021
Crossref
40 words — 1%

11 Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021
Crossref
37 words — 1%

12 www.bmj.com
Internet
35 words — 1%

13 plus.mcmaster.ca
Internet
34 words — 1%

14 Farshid Rahimibashar, Ladan Sedighi, Alireza Shahriary, Zeljko Reiner et al. "Is there any association between plasma lipid profile and severity of COVID-19?", Clinical Nutrition ESPEN, 2022
Crossref
33 words — 1%

15 philpapers.org
Internet
33 words — 1%

www.hindawi.com

16	Internet	33 words — 1%
17	www.researchgate.net Internet	32 words — 1%
18	www.aimspress.com Internet	31 words — 1%
19	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	29 words — < 1%
20	saucis.sakarya.edu.tr Internet	28 words — < 1%
21	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet	27 words — < 1%
22	Geng-Xin Xu, Chen Liu, Jun Liu, Zhongxiang Ding et al. "Cross-Site Severity Assessment of COVID-19 from CT Images via Domain Adaptation", IEEE Transactions on Medical Imaging, 2021 Crossref	27 words — < 1%
23	Tawsifur Rahman, Amith Khandakar, Md Enamul Hoque, Nabil Ibtehaz et al. "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access, 2021 Crossref	26 words — < 1%
24	Tora Fahrudin, Nelsi Wisna, Patrick Adolf Telnoni, Dedy Rahman Wijaya. "Sales Forecasting Web	25 words — < 1%

Application in Small and Medium Enterprise", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022
Crossref

25	ijsrcseit.com Internet	25 words — < 1%
26	smartcities.ieee.org Internet	25 words — < 1%
27	seaninstitute.org Internet	24 words — < 1%
28	www.gtu.edu.tr Internet	24 words — < 1%
29	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018 Crossref	23 words — < 1%
30	jurnal.ugm.ac.id Internet	22 words — < 1%
31	www.ncbi.nlm.nih.gov Internet	22 words — < 1%
32	www.ijppr.humanjournals.com Internet	21 words — < 1%
33	Qingyun Ren, Bingyin Zhou, Liang Tian, Wei Guo. "Detection of COVID-19 With CT Images Using Hybrid Complex Shearlet Scattering Networks", IEEE Journal of Biomedical and Health Informatics, 2022	20 words — < 1%

Crossref

34	Xiang Fang, Yuchong Hu, Pan Zhou, Dapeng Oliver Wu. "V^{3}H: View Variation and View Heredity for Incomplete Multiview Clustering", IEEE Transactions on Artificial Intelligence, 2020 Crossref	20 words — < 1%
35	"iCAST 2017 proceedings", 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017 Crossref	19 words — < 1%
36	Catherine P. Adans-Dester, Stacy Bamberg, Francesco P. Bertacchi, Brian Caulfield et al. "Can mHealth Technology Help Mitigate the Effects of the COVID-19 Pandemic?", IEEE Open Journal of Engineering in Medicine and Biology, 2020 Crossref	19 words — < 1%
37	Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita. "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review", Computational and Structural Biotechnology Journal, 2021 Crossref	19 words — < 1%
38	Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 Crossref	19 words — < 1%
39	bura.brunel.ac.uk Internet	19 words — < 1%
40	eprints.utm.my Internet	19 words — < 1%

41	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%
42	annals-csis.org Internet	17 words — < 1%
43	journals.plos.org Internet	16 words — < 1%
44	www.etasr.com Internet	16 words — < 1%
45	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	14 words — < 1%
46	Anthony Anggrawan, Dwi Kurnianingsih, Christofer Satria. "Sistem Aplikasi Cerdas Klasterisasi Penerima Bantuan Covid-19", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Crossref	13 words — < 1%
47	etd.aau.edu.et Internet	13 words — < 1%
48	www.ijicic.org Internet	13 words — < 1%
49	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for	12 words — < 1%

Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020
Crossref

50 Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022 — 12 words — < 1%
Crossref

51 ejournal.uin-suska.ac.id — 12 words — < 1%
Internet

52 riull.ull.es — 12 words — < 1%
Internet

53 www.sv-europe.com — 12 words — < 1%
Internet

54 Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 — 11 words — < 1%
Crossref

55 arxiv.org — 11 words — < 1%
Internet

56 dokumen.pub — 10 words — < 1%
Internet

57 ijece.iaescore.com — 10 words — < 1%
Internet

58 tojqi.net
Internet

10 words — < 1%

59 Zaiwei Song, Yang Hu, Siqian Zheng, Li Yang, Rongsheng Zhao. "Hospital pharmacists' pharmaceutical care for hospitalized patients with COVID-19: Recommendations and guidance from clinical experience", Research in Social and Administrative Pharmacy, 2020
Crossref
9 words — < 1%

60 journalofbigdata.springeropen.com
Internet
9 words — < 1%

61 kresttechnology.com
Internet
9 words — < 1%

62 pdfs.semanticscholar.org
Internet
9 words — < 1%

63 123dok.com
Internet
8 words — < 1%

64 Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021
Crossref
8 words — < 1%

65 Soham Guhathakurata, Sayak Saha, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A new approach to predict COVID-19 using artificial neural networks", Elsevier BV, 2022
Crossref
8 words — < 1%

66 Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for
8 words — < 1%

COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022
Crossref

67	annalsofrscb.ro Internet	8 words — < 1%
68	ijai.iaescore.com Internet	8 words — < 1%
69	journals.gaftim.com Internet	8 words — < 1%
70	www.ajtmh.org Internet	8 words — < 1%
71	www.medrxiv.org Internet	8 words — < 1%
72	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	7 words — < 1%
73	en.wikipedia.org Internet	7 words — < 1%
74	hdl.handle.net Internet	7 words — < 1%
75	Subrat Kumar Dang, Kulwant Singh. "Predicting tensile-shear strength of nugget using M5P model tree and random forest: An analysis", Computers in Industry, 2021 Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF

Gmail

Anthony Anggrawan <anthony.anggrawan@universitasbumigora.ac.id>

[jait] Manuscript ID: JAIT-5790 – Submission Acknowledgement

Ms. Haylee Lin <haylee.lin@ejournal.net> Tue, Jun 14, 2022 at 9:05 PM
To: Anthony Anggrawan <anthony.anggrawan@universitasbumigora.ac.id>

Dear Anthony Anggrawan:

Thank you for submitting the manuscript, "Machine Learning to Predict the Treatment Status of Covid-19 Patients" to Journal of Advances in Information Technology. With the online journal management system that we are using, you will be able to track its progress through the editorial process by logging in to the journal web site:

Submission URL: http://ojs.ejournal.net/index.php/jait/authorDashboard/submission/5790
Username: aanggrawan

Please use the manuscript ID (JAIT-5790) in all future correspondence. As the corresponding author, you will receive all future communications about this manuscript.

If you have any questions, please contact me. Thank you for considering this journal as a venue for your work.

Ms. Haylee Lin

--
Journal of Advances in Information Technology
Website: http://www.jait.us/
Email: jait@etpub.com

Machine Learning to Predict the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly, researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. However, the problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [23]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

Many Data Mining methods include K-Means,

Manuscript received August 17, 2021; revised December 6, 2021.

Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known technique used for classification [26].

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article.

- Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict pneumonia severity and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and way compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but the RF and SVM methods.
- Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.
- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data

Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.

- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and method compared to this article's research.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. METHODOLOGY

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process comprises a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

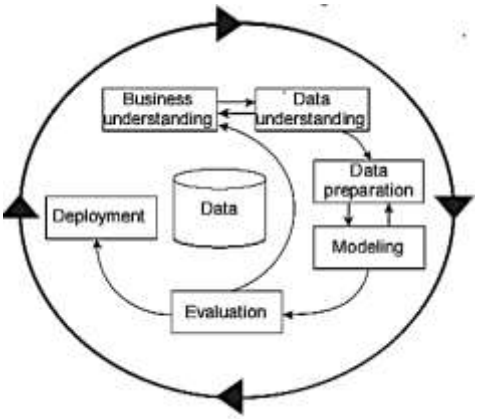

Figure 1. The CRISP-DM Process

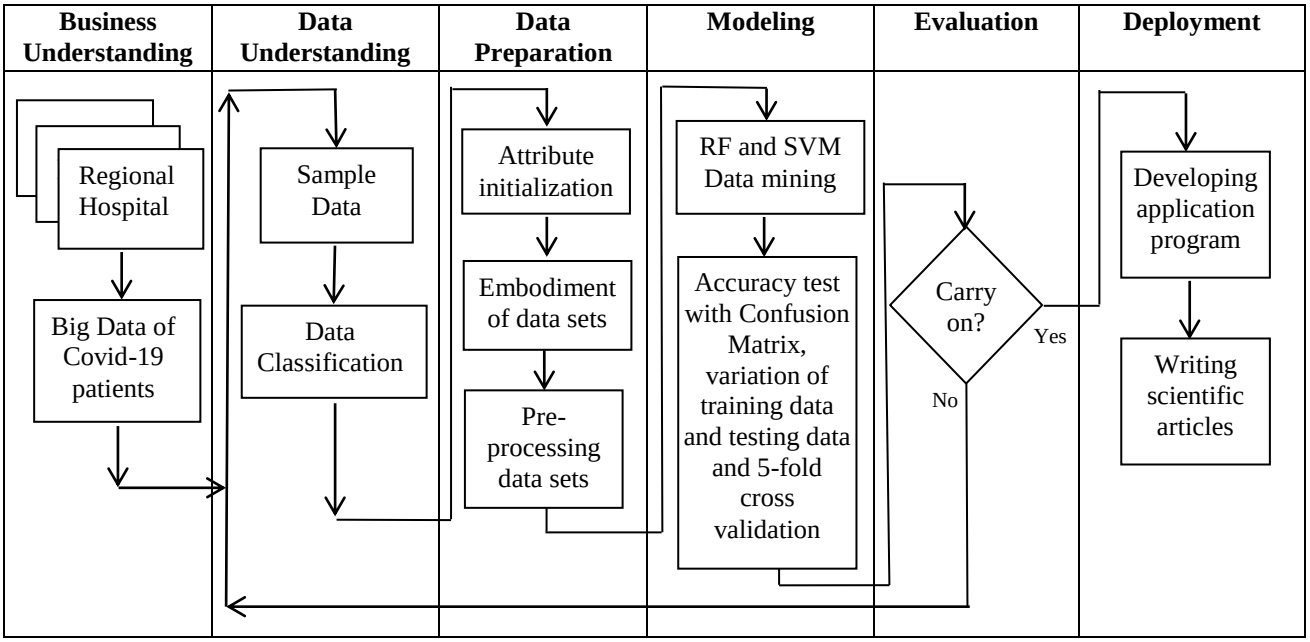

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data set of the signs and symptoms and the treatment status of Covid-19 patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses to others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \text{ x } log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation is helpful in assessing the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestions for further research are the need to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.
[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.
[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.
[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.
[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[32] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.
[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT 5512
TIME SUBMITTED 14-JUN-2022 07:43PM
PAPER ID 87248448

Machine Learning to Predict the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly, researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. However, the problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [23]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

Many Data Mining methods include K-Means,

Manuscript received August 17, 2021; revised December 6, 2021.

Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known technique used for classification [26].

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article.

- Askin Kavzoglu, Fusun Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict pneumonia severity and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and way compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but the RF and SVM methods.
- Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.
- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data

Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.

- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and method compared to this article's research.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process comprises a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

Figure 1. The CRISP-DM Process

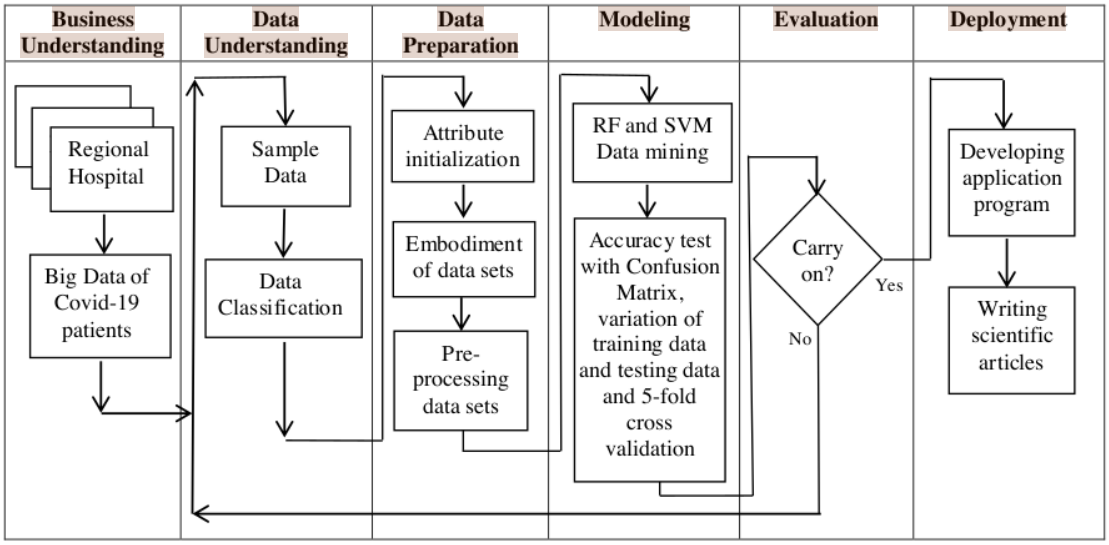

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data Set of the Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses to others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \times log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation is helpful in assessing the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestions for further research are the need to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[32] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

17%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	66 words — 1%
2	www.jatit.org Internet	58 words — 1%
3	Hairani Hairani, Anthony Anggrawan, M. Ade Candra. "Prediction of Electricity Usage with Back-propagation Neural Network", International Journal of Engineering and Computer Science Applications (IJECSA), 2022 Crossref	55 words — 1%
4	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
5	www.mdpi.com Internet	42 words — 1%
6	Norah Alballa, Isra Al-Turaiki. "Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review", Informatics in Medicine Unlocked, 2021 Crossref	31 words — 1%
7	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision	23 words — 1%

Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018
Crossref

8	www.researchgate.net Internet	23 words — 1%
9	jurnal.ugm.ac.id Internet	22 words — < 1%
10	etd.aau.edu.et Internet	21 words — < 1%
11	thesai.org Internet	21 words — < 1%
12	Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 Crossref	19 words — < 1%
13	bura.brunel.ac.uk Internet	19 words — < 1%
14	eprints.utm.my Internet	19 words — < 1%
15	ijece.iaescore.com Internet	19 words — < 1%
16	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%

17	annals-csis.org Internet	17 words — < 1%
18	en.wikipedia.org Internet	15 words — < 1%
19	www.sv-europe.com Internet	13 words — < 1%
20	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	12 words — < 1%
21	Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022 Crossref	12 words — < 1%
22	riull.ull.es Internet	12 words — < 1%
23	Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 Crossref	11 words — < 1%
24	Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021	11 words — < 1%

Crossref

25	arxiv.org Internet	11 words — < 1%
26	www.hindawi.com Internet	11 words — < 1%
27	dokumen.pub Internet	10 words — < 1%
28	link.springer.com Internet	10 words — < 1%
29	Coviki.org Internet	9 words — < 1%
30	Soham Guhathakurata, Sayak Saha, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A new approach to predict COVID-19 using artificial neural networks", Elsevier BV, 2022 Crossref	8 words — < 1%
31	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
32	annalsofrscb.ro Internet	8 words — < 1%
33	journals.gaftim.com Internet	8 words — < 1%
34	plus.mcmaster.ca Internet	8 words — < 1%

35	www.ajtmh.org Internet	8 words — < 1%
36	www.medrxiv.org Internet	8 words — < 1%
37	"Advances in Data Computing, Communication and Security", Springer Science and Business Media LLC, 2022 Crossref	7 words — < 1%
38	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	7 words — < 1%
39	hdl.handle.net Internet	7 words — < 1%
40	"iCAST 2017 proceedings", 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017 Crossref	6 words — < 1%
41	David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 Crossref	6 words — < 1%
42	Melissa A. Hausburg, Jason S. Williams, Kaysie L. Banton, Charles W. Mains, Michael Roshon, David Bar-Or. "C1 esterase inhibitor-mediated immunosuppression in COVID-19: friend or foe?", Clinical Immunology Communications, 2022 Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT 5512
TIME SUBMITTED 14-JUN-2022 07:43PM
PAPER ID 87248448

Machine Learning to Predict the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly, researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. However, the problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [23]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

Many Data Mining methods include K-Means,

Manuscript received August 17, 2021; revised December 6, 2021.

Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known technique used for classification [26].

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article.

- Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict pneumonia severity and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and way compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but the RF and SVM methods.
- Sriram Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.
- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data

Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.

- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and method compared to this article's research.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research methods, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process comprises a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

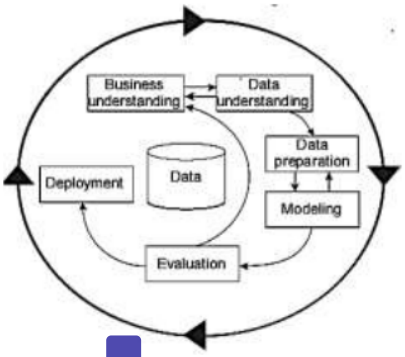

Figure 1. The CRISP-DM Process

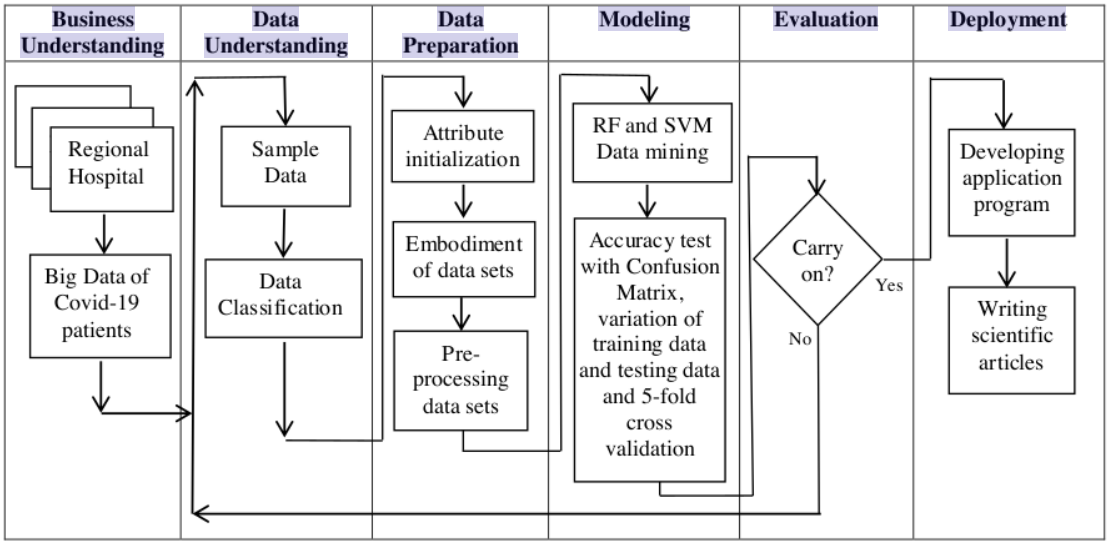

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data Set of the Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses to others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \times log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation is helpful in assessing the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestions for further research are the need to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. ch, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 23, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–8, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 31–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. , no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest ighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, . 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, –11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, 12, no. 11, pp. 111–118, 2021.
[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.
[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.
[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–2020.
[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[32] Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipient Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, p. 1–17, 2022.
[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

31%
SIMILARITY INDEX

PRIMARY SOURCES

1	thesai.org Internet	146 words — 3%
2	journal.universitasbumigora.ac.id Internet	126 words — 2%
3	www.jatit.org Internet	71 words — 1%
4	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
5	Hairani Hairani, Anthony Anggrawan, M. Ade Candra. "Prediction of Electricity Usage with Back-propagation Neural Network", International Journal of Engineering and Computer Science Applications (IJECSA), 2022 Crossref	48 words — 1%
6	Haodong Li. "An Overview on Remote Sensing Image Classification Methods with a Focus on Support Vector Machine", 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), 2021 Crossref	45 words — 1%
7	www.hindawi.com Internet	44 words — 1%

8	Fermi Dwi Wicaksono, Yusri Bin Arshad, Haeryip Sihombing. "Norm-dist Monte-Carlo integrative method for the improvement of fuzzy analytic hierarchy process", Heliyon, 2020 Crossref	42 words — 1%
9	www.ijeat.org Internet	42 words — 1%
10	www.mdpi.com Internet	42 words — 1%
11	Petrisia Widyasari Sudarmadji, Prisca Deviani Pakan, Rocky Yefrenes Dillak. "Diabetic Retinopathy Stages Classification using Improved Deep Learning", 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), 2020 Crossref	40 words — 1%
12	www.ncbi.nlm.nih.gov Internet	40 words — 1%
13	Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021 Crossref	37 words — 1%
14	plus.mcmaster.ca Internet	34 words — 1%
15	Farshid Rahimibashar, Ladan Sedighi, Alireza Shahriary, Zeljko Reiner et al. "Is there any association between plasma lipid profile and severity of COVID-19?", Clinical Nutrition ESPEN, 2022 Crossref	33 words — 1%

16	philpapers.org Internet	33 words — 1%
17	www.researchgate.net Internet	32 words — 1%
18	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	31 words — 1%
19	www.aimspress.com Internet	31 words — 1%
20	saucis.sakarya.edu.tr Internet	28 words — < 1%
21	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet	27 words — < 1%
22	Tawsifur Rahman, Amith Khandakar, Md Enamul Hoque, Nabil Ibtehaz et al. "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access, 2021 Crossref	26 words — < 1%
23	www.bmj.com Internet	26 words — < 1%
24	Mohammad Uzzal Hossain, Ishtiaque Ahammad, Md. Tabassum Hossain Emon, Arittra Bhattacharjee et al. "Whole Genome Sequencing for Revealing the Point Mutations of SARS-CoV-2 Genome in Bangladeshi	25 words — < 1%

Isolates and their Structural Effects on Viral Proteins", Cold Spring Harbor Laboratory, 2020
Crossref Posted Content

25 Tora Fahrudin, Nelsi Wisna, Patrick Adolf Telnoni, Dedy Rahman Wijaya. "Sales Forecasting Web Application in Small and Medium Enterprise", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022
Crossref
25 words — < 1%

26 ijsrcseit.com
Internet
25 words — < 1%

27 smartcities.ieee.org
Internet
25 words — < 1%

28 tj.uettaxila.edu.pk
Internet
25 words — < 1%

29 seaninstitute.org
Internet
24 words — < 1%

30 www.gtu.edu.tr
Internet
24 words — < 1%

31 Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018
Crossref
23 words — < 1%

32 jurnal.ugm.ac.id
Internet
22 words — < 1%

33 etd.aau.edu.et
Internet

21 words — < 1%

34 www.ijppr.humanjournals.com
Internet
21 words — < 1%

35 Qingyun Ren, Bingyin Zhou, Liang Tian, Wei Guo. "Detection of COVID-19 With CT Images Using Hybrid Complex Shearlet Scattering Networks", IEEE Journal of Biomedical and Health Informatics, 2022
Crossref
20 words — < 1%

36 Catherine P. Adans-Dester, Stacy Bamberg, Francesco P. Bertacchi, Brian Caulfield et al. "Can mHealth Technology Help Mitigate the Effects of the COVID-19 Pandemic?", IEEE Open Journal of Engineering in Medicine and Biology, 2020
Crossref
19 words — < 1%

37 Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita. "Artificial Intelligence in clinical care amidst COVID-19 Pandemic: A Systematic Review", Computational and Structural Biotechnology Journal, 2021
Crossref
19 words — < 1%

38 Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022
Crossref
19 words — < 1%

39 bura.brunel.ac.uk
Internet
19 words — < 1%

40 eprints.utm.my
Internet
19 words — < 1%

41	ijece.iaescore.com Internet	19 words — < 1%
42	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%
43	annals-csis.org Internet	17 words — < 1%
44	researchprofiles.herts.ac.uk Internet	16 words — < 1%
45	www.etasr.com Internet	16 words — < 1%
46	Anthony Anggrawan, Dwi Kurnianingsih, Christofer Satria. "Sistem Aplikasi Cerdas Klasterisasi Penerima Bantuan Covid-19", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Crossref	13 words — < 1%
47	www.sv-europe.com Internet	13 words — < 1%
48	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	12 words — < 1%
49	Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under	12 words — < 1%

Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022
Crossref

50	ejournal.uin-suska.ac.id Internet	12 words — < 1%
51	riull.ull.es Internet	12 words — < 1%
52	Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 Crossref	11 words — < 1%
53	Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021 Crossref	11 words — < 1%
54	arxiv.org Internet	11 words — < 1%
55	dokumen.pub Internet	10 words — < 1%
56	link.springer.com Internet	10 words — < 1%
57	tojqi.net Internet	10 words — < 1%

58	Coviki.org Internet	9 words — < 1%
59	kresttechnology.com Internet	9 words — < 1%
60	Soham Guhathakurata, Sayak Saha, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A new approach to predict COVID-19 using artificial neural networks", Elsevier BV, 2022 Crossref	8 words — < 1%
61	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
62	annalsofrscb.ro Internet	8 words — < 1%
63	ijai.iaescore.com Internet	8 words — < 1%
64	journals.gaftim.com Internet	8 words — < 1%
65	www.ajtmh.org Internet	8 words — < 1%
66	www.medrxiv.org Internet	8 words — < 1%
67	"Advances in Data Computing, Communication and Security", Springer Science and Business Media LLC, 2022 Crossref	7 words — < 1%

68	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	7 words — < 1%
69	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	7 words — < 1%
70	hdl.handle.net Internet	7 words — < 1%
71	"iCAST 2017 proceedings", 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017 Crossref	6 words — < 1%
72	David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 Crossref	6 words — < 1%
73	Melissa A. Hausburg, Jason S. Williams, Kaysie L. Banton, Charles W. Mains, Michael Roshon, David Bar-Or. "C1 esterase inhibitor-mediated immunosuppression in COVID-19: friend or foe?", Clinical Immunology Communications, 2022 Crossref	6 words — < 1%
74	en.wikipedia.org Internet	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF

Submit an Article

Section

Article

Articles must be submitted to one of the journal's sections. *

Submission Requirements

You must read and acknowledge that you've completed the requirements below before proceeding.

- [x] The submission has been prepared in accordance with the journal's Instructions for Authors and follows all the journals' policy requirements.
- [x] The submission file is in Microsoft Word format.
- [x] All co-authors must be included at the time of submission with valid e-mail addresses.
- [x] No part of the manuscript has been, or will be, published elsewhere nor is under consideration for publication elsewhere.

Comments for the Editor

Upload

Yes, I agree to have my data collected and stored according to the privacy statement.

Save and continue Cancel

* *Denotes required field*

Submit an Article

Submissions

Submission Files

Search Upload File

18716-2 aanggrawan, JAIT_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients.pdf (2)	June 14, 2022	Manuscript (Word)	

Save and continue Cancel

Add Contributor

Name

Mayadi

Given Name *

Family Name

How do you prefer to be addressed? Salutations, middle names and suffixes can be added here if you would like.

Preferred Public Name

Contact

mayadi.yadot@universitasbumigora.ac.id

Email *

Country

Indonesia

Country *

User Details

Homepage URL

ORCID iD

Department of Computer Science of Bumigora University

Affiliation

Add Contributor

Name

Mayadi

Given Name *

Family Name

How do you prefer to be addressed? Salutations, middle names and suffixes can be added here if you would like.

Preferred Public Name

Contact

mayadi.yadot@universitasbumigora.ac.id

Email *

Country

Indonesia

Country *

User Details

Homepage URL

ORCID iD

Department of Computer Science of Bumigora University

Affiliation

Add Contributor

Name

Mayadi

Given Name *

Family Name

How do you prefer to be addressed? Salutations, middle names and suffixes can be added here if you would like.

Preferred Public Name

Contact

mayadi.yadot@universitasbumigora.ac.id

Email *

Country

Indonesia

Country *

User Details

Homepage URL

ORCID iD

Department of Computer Science of Bumigora University

Affiliation

Homepage URL *ORCID ID*

Department of Visual Communication Design of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Homepage URL *ORCID ID*

Department of Visual Communication Design of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Homepage URL *ORCID ID*

Department of Visual Communication Design of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Homepage URL *ORCID iD*

Department of Computer Science of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Homepage URL *ORCID iD*

Department of Computer Science of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Homepage URL *ORCID iD*

Department of Computer Science of Bumigora University

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Affiliation

Upload

Bio Statement (e.g., department and rank)

Contributor's role *

- (•) Author
- ( ) Translator

- [ ] Principal contact for editorial correspondence.
- [x] Include this contributor in browse lists?

> * *Denotes required field*

Submit an Article

Submissions

Prefix

Examples: A, The

Title *

Machine Learning to Predict the Treatment Status of Covid-19 Patients

Subtitle

Abstract *

Upload

Covid-19 disease spread throughout the world and caused death globally. Accordingly, researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. However, the problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an

List of Contributors

Name	E-mail	Role	Primary Contact	In Browse Lists
Anthony Anggrawan	anthony.anggrawan@universitasbumigora.ac.id	Author	☑	☑
Mayadi	mayadi.yadot@universitasbumigora.ac.id	Author		☑
Christofer Satria	chris@universitasbumigora.ac.id	Author		☑
Krismono	bkrismono@universitasbumigora.ac.id	Author		☑
Ria Rismayati	riris@universitasbumigora.ac.id	Author		☑

Submission Metadata

These specifications are based on the Dublin Core metadata set, an international standard used to describe journal content.

Additional Refinements

Keywords *

Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning ×

Submit an Article

Submissions

Your submission has been uploaded and is ready to be sent. You may go back to review and adjust any of the information you have entered before continuing. When you are ready, click "Finish Submission".

Finish Submission Cancel

Submit an Article

Submission complete

Thank you for your interest in publishing with Journal of Advances in Information Technology.

What Happens Next?

The journal has been notified of your submission, and you've been emailed a confirmation for your records. Once the editor has reviewed the submission, they will contact you.

For now, you can:

- Review this submission
- Create a new submission
- Return to your dashboard

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT 5512
TIME SUBMITTED 14-JUN-2022 07:43PM
PAPER ID 87248448

Machine Learning to Predict the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 disease spread throughout the world and caused death globally. Accordingly, researchers use many research methods to control the Covid-19 pandemic, one of which is getting more attention is the use of prediction methods. However, the problem in medical action is not easy to determine the treatment status. So it is not surprising that previous researchers emphasized that the most frequent mistakes were due to inaccuracies in medical decisions. Therefore, this study proposes a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason, or that is why this study proposes a machine learning system model for decision-making solutions (predictions) of the treatment status of Covid-19 patients, both inpatient and self-isolation using data mining methods.

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [23]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting the new data with high accuracy [22].

Many Data Mining methods include K-Means,

Manuscript received August 17, 2021; revised December 6, 2021.

Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out in order to be able to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [24] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [25] [26], as well as SVM algorithm has a widely known technique used for classification [26].

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some of the related works from the latest scientific articles compared with the work in this research article.

- Askin Kavzoglu, Fusun Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [24]. This previous research is different from the research in the article on the research objectives and the object under study.
- Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [27]. The difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.
- Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [28]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.
- Based on patient clinical data, Boran Hao et al. (2020) developed a model to predict pneumonia severity and the level of care required by Covid-19 patients using statistical methods.[29]. This previous research differs in the research purpose and way compared to the research in this article.
- Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.
- Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson distribution and the gamma distribution [30]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article.
- Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and methods used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but the RF and SVM methods.
- Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [31]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not using the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods.
- Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [32]. This previous research is a literature review study that concludes that the Data

Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from the research in this article because this article is an experimental study, not a literature review research.

- Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods used compared to the research in this article. The difference lies in the fact that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].
- Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [33]. However, the previous research differs in the purpose and method compared to this article's research.
- Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously Covid-19 patients with plasma proteomics [34]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties that previous researchers have not studied. The novelty of the study lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

III. Methodology

This study applies two kinds of data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The development of the application program in this study uses the Python programming language.

The Data Mining Process for Covid-19 patient big data in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining) process; CRISP-DM is standard data mining. The CRISP-DM process comprises a six-stage [35]; see Figure 1 [36]. Figure 2 shows the data mining process in this study. This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

Figure 1. The CRISP-DM Process

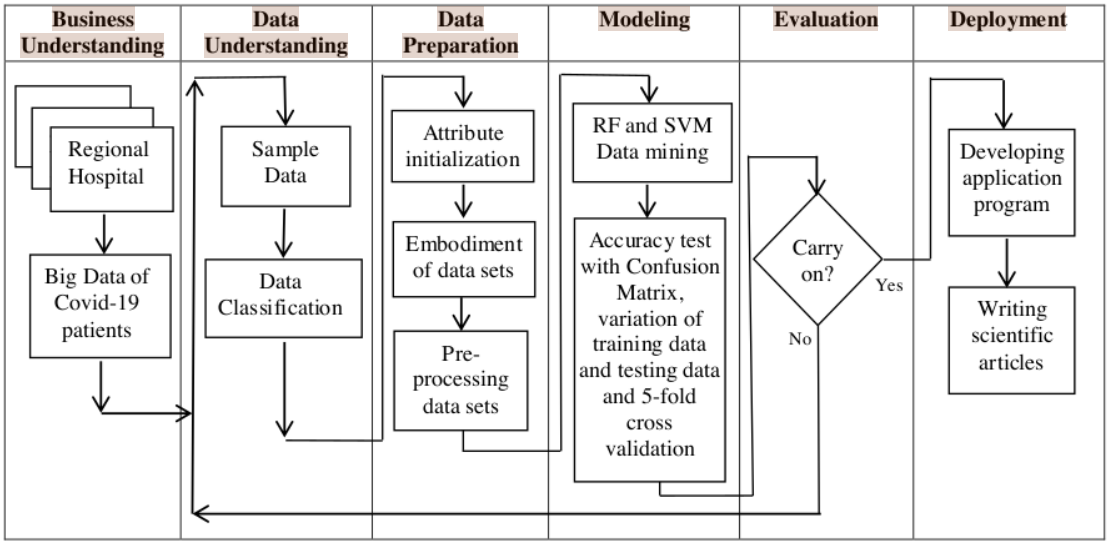

Figure 2. Data Mining Process of Covid-19 Patient Big Data

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected in the form of medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. The big data of Covid-19 patients obtained was 117. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 1 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE I. Data Set of the Signs and Symptoms and the Treatment Status of Covid-19 Patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses to others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 2).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 3. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or

symptoms.

TABLE II. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE III. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	yes	no	no	No	…	…	No	No	Inpatient
2	yes	yes	no	Yes	…	…	No	No	Inpatient
3	yes	no	yes	No	…	…	No	No	Inpatient
4	yes	no	no	No	…	…	No	No	Inpatient
5	yes	no	no	No	…	…	No	No	Inpatient
6	yes	no	no	No	…	…	Yes	Yes	Inpatient
7	no	no	yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	no	no	no	No	…	…	No	Yes	Self-isolation
115	yes	no	no	No	…	…	Yes	No	Inpatient
116	yes	no	no	No	…	…	Yes	Yes	Inpatient
117	no	no	no	No	…	…	No	No	Inpatient

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1.

In addition, preprocessing the data set is also carried out on the class attribute, namely changing the class attribute with y and changing the value of the self-isolation class attribute with the number 1 and the *inpatient class attribute with the number -1, as shown in Table 4.

.TABLE IV. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [37], which has their respective advantages in building application programs [38][39]. The application program built in this research uses the Python programming language to facilitate patient care status prediction.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

The formula of the SVM data mining method is: to form a linear equation from the training data; find the value of w and b, and the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum j^2 \quad (4)$$

$$Entropy(S) = \sum i - pi \times log_2(pi) \quad (5)$$

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

The format of the confusion matrix table is as shown in Table 5.

TABLE V. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation is helpful in assessing the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 6. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VI. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TEST DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 7.

TABLE VII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

4.6 Deployment

One of the deployments in this research is in the form of making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 92%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Suggestions for further research are the need to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion is to conduct further research using other predictive data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first author, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug

Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[23] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[24] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[25] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[26] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[27] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[28] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[29] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[30] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[31] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[32] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[33] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[34] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[35] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[36] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[37] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[38] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[39] A. Anggrawan, S. Hadi, C. Satria, and U. Bumigora, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Similarity Check_Machine Learning Model to Predict the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

17%
SIMILARITY INDEX

PRIMARY SOURCES

1	www.ncbi.nlm.nih.gov Internet	66 words — 1%
2	www.jatit.org Internet	58 words — 1%
3	Hairani Hairani, Anthony Anggrawan, M. Ade Candra. "Prediction of Electricity Usage with Back-propagation Neural Network", International Journal of Engineering and Computer Science Applications (IJECSA), 2022 Crossref	55 words — 1%
4	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
5	www.mdpi.com Internet	42 words — 1%
6	Norah Alballa, Isra Al-Turaiki. "Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review", Informatics in Medicine Unlocked, 2021 Crossref	31 words — 1%
7	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision	23 words — 1%

Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018
Crossref

8	www.researchgate.net Internet	23 words — 1%
9	jurnal.ugm.ac.id Internet	22 words — < 1%
10	etd.aau.edu.et Internet	21 words — < 1%
11	thesai.org Internet	21 words — < 1%
12	Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 Crossref	19 words — < 1%
13	bura.brunel.ac.uk Internet	19 words — < 1%
14	eprints.utm.my Internet	19 words — < 1%
15	ijece.iaescore.com Internet	19 words — < 1%
16	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%

17	annals-csis.org Internet	17 words — < 1%
18	en.wikipedia.org Internet	15 words — < 1%
19	www.sv-europe.com Internet	13 words — < 1%
20	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	12 words — < 1%
21	Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022 Crossref	12 words — < 1%
22	riull.ull.es Internet	12 words — < 1%
23	Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 Crossref	11 words — < 1%
24	Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021	11 words — < 1%

Crossref

25	arxiv.org Internet	11 words — < 1%
26	www.hindawi.com Internet	11 words — < 1%
27	dokumen.pub Internet	10 words — < 1%
28	link.springer.com Internet	10 words — < 1%
29	Coviki.org Internet	9 words — < 1%
30	Soham Guhathakurata, Sayak Saha, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A new approach to predict COVID-19 using artificial neural networks", Elsevier BV, 2022 Crossref	8 words — < 1%
31	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
32	annalsofrscb.ro Internet	8 words — < 1%
33	journals.gaftim.com Internet	8 words — < 1%
34	plus.mcmaster.ca Internet	8 words — < 1%

35	www.ajtmh.org Internet	8 words — < 1%
36	www.medrxiv.org Internet	8 words — < 1%
37	"Advances in Data Computing, Communication and Security", Springer Science and Business Media LLC, 2022 Crossref	7 words — < 1%
38	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	7 words — < 1%
39	hdl.handle.net Internet	7 words — < 1%
40	"iCAST 2017 proceedings", 2017 IEEE 8th International Conference on Awareness Science and Technology (iCAST), 2017 Crossref	6 words — < 1%
41	David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 Crossref	6 words — < 1%
42	Melissa A. Hausburg, Jason S. Williams, Kaysie L. Banton, Charles W. Mains, Michael Roshon, David Bar-Or. "C1 esterase inhibitor-mediated immunosuppression in COVID-19: friend or foe?", Clinical Immunology Communications, 2022 Crossref	6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that, or that's why this study aims to propose a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18][19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20][21][22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [23]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. That is, machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32] [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used together in research, including research to classify and/or predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. Whereas, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review research.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 shows a comparison between some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the last research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The big data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Figure 1 [43]. Figure 2 shows the process carried out at each stage of CRISP-DM.

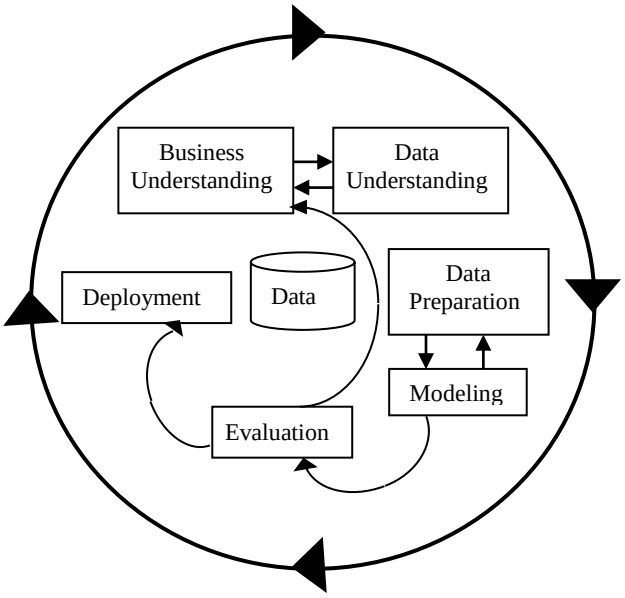

Figure 1. The CRISP-DM Process

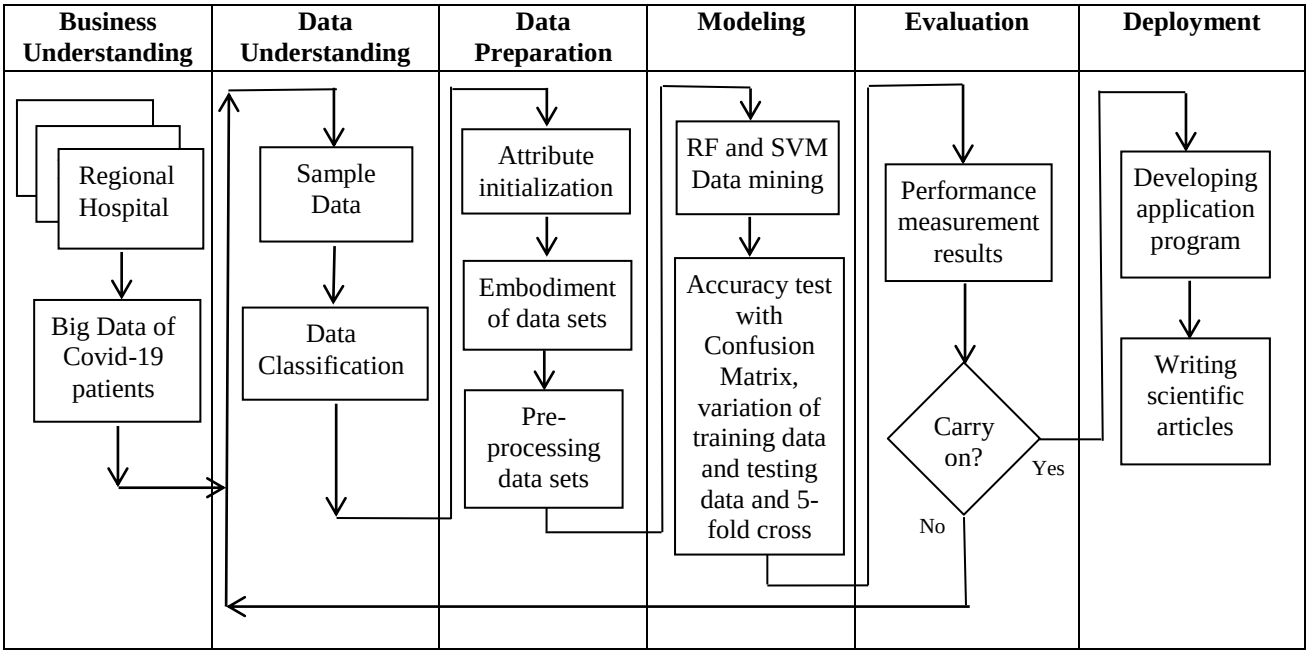

Figure 2. Data Mining Process of Covid-19 Patient Big Data

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [44], which has their respective advantages in building application programs [45][46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction. The hyperparameter used in the SVM method is kernel = linear and Probability = True. Meanwhile, the hyperparameter used for the RF hyperparameter method is n_estimators = 100. The n_estimators is the number of trees in RF.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:

S = Set of cases

pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 6.

TABLE VI. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 7 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VII. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

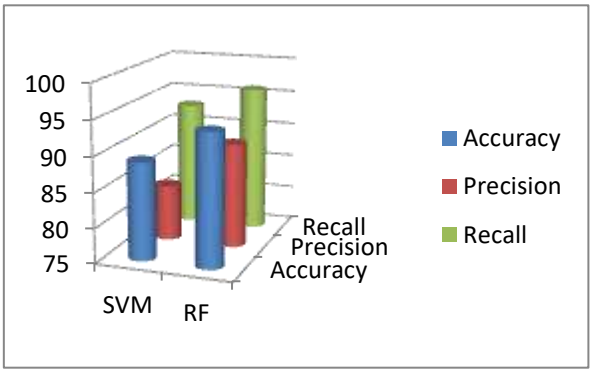

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix. 333

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 9 and Figure 4.

TABLE IX. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

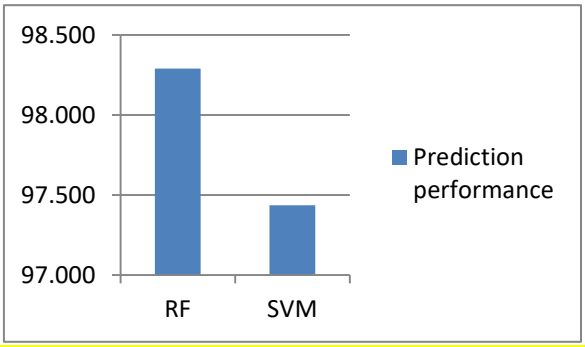

Figure 4. Predictive performance testing using K-fold cross-validation

4.6 Deployment

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of machine learning and the Internet of Things.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li,

"COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.
[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.
[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.
[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.
[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.
[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.
[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.
[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.
[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.
[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.
[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.
[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.
[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.
[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.
[20] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.
[21] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.
[22] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.
[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.
[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.
[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[26] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.
[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.
[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.
[29] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.
[30] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.
[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.
[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[34] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.
[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.
[36] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[37] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[39] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[41] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.
[42] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[43] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[45] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Correction Note

Manuscript ID: JAIT-5790
Title: Machine Learning to Predict the Treatment Status of Covid-19 Patients

Dear Reviewers
We have made improvements to our manuscript following the instructions and suggestions of the reviewers. About which parts we have revised and added writing, it is as explained below:

	Reviewer's Comments (Reviewer A)	The Change/Corrections made (the change is marked with a yellow highlight)	In Page & paragraph
1	Abstract: is vague and does not serve its purpose. Abstract should introduce the domain, discuss the problem statement, present the objective, discuss the possible outcome, and finally present the contribution to the research domain	**Abstract Contents after revision** Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that, or that's why this study aims to propose a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.	On page 1
2	Introduction: this section is again vague. The background on Covid-19 is good but it is expected to discuss more from the research perspective. What are the research implications of the Covid-19 pandemic? How is artificial intelligence contributing to controlling the pandemic? What is the research gap you have identified? Why are you	Following the reviewer's suggestion, the authors have added writing related to the reviewer's direction in the Introduction subsection and Related works subsection as follows: * In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after	On page 2, 2nd paragraph, 2nd column, in the Introduction subsection

	proposing this work? All of these above questions should be answered within the introduction and related works section.	analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care. * For this reason, this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods. The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. That is, machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. However, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [23]. * After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [25]. * By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties (from the last research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms.	Page 2, 1st paragraph, 1st column, in the Introduction subsection Page 2, 2nd paragraph, 1st column, in the Introduction subsection Page 2, 3rd paragraph, 1st column, in the Introduction subsection Page 3, 8th paragraph, 2nd column, in the Related Works subsection
3	Related works: This section summarizes a few research papers on RF and SVM however, there is no real analysis done. Related work study is to identify the relevance of the proposed work and to identify the research gap. I suggest the authors to add a table comparing the state-of-the-art	The authors have added Table 1. Comparison of This Article's Work With Some Previous Related Works in the Related Works subsection following the reviewers' suggestions The authors add sentences related to table 1 as follows: Table 1 shows a comparison between some of the most recent previous related work with the work carried out in this study	Page 4, in the Related Works subsection Page 3, 8th paragraph, 2nd column, in the Related Works subsection

	research on covid-19 prediction with machine learning.		
4	a. Methodology: is very weak. Generally, for a research paper, a methodology paper should be given more importance. It should clearly explain the dataset, introduce the methodology used, and discuss the novelty and algorithms if any. b. I felt the subsections belonging to results and discussions should be added to the methodology section and more detailed information should be added; c. Figure 1 is not readable and of poor image quality; d. SVM and RF hyperparameter details are not discussed. More details of SVM and RF should be given; e. Confusion matrix format is presented but the result is not plotted; f. The need for a recall, and precision is not discussed. The specificity result should be reported which is missing; g. Deployment details should be justified. How scientific article writing can be considered deployment?	The authors have added descriptions to the methodology subsections as follows: This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. a. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [41], as shown in Figure 1 [42]. Meanwhile, figure 2 shows the process carried out at each stage of CRISP-DM. b. Figure 2 shows the process carried out at each stage of CRISP-DM. In figure 2, Business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles. c. The author has improved Figure 1 so that the image quality is legible and clear. d. The hyperparameter used in the SVM method is kernel = linear and probability = True; Meanwhile the hyperparameter used for the RF hyperparameter method is n_estimators = 100. n_estimators is the number of trees in RF.	Page 5, 2nd paragraph, 2nd column, in the methodology subsection Page 5, 2nd paragraph, 2nd column & page 6, 1st paragraph, 1st column On page 5 On page 7, 2nd column, paragraph 4.

		e. Regarding the presentation of the confusion matrix format presented, the author has added the results in Figure 3 and Figure 4 (in Result and Discussion subsection). f. Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity. g. One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers.	Figure 3 on page 8 and Figure 4 on page 9 On page 8, paragraph 3, 1st column On page 9, in subheading 4.6 Deployment
5	Future directions is weak. Should be more clear based on the findings and extension of the work.	As per the reviewer's suggestions, the authors have added the following description (text highlighted in yellow in the Conclusion subsection) as follows: Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of machine learning and the Internet of Things.	On page 9, 2nd column, paragraph 3, in Conclusion subsection
6	In many places, the authors have mentioned big data. When a bigdata is involved the processing and storing techniques of bigdata should also be discussed. Authors must clearly define the type of data they are using in the work.	In accordance with the reviewers' suggestions, the authors have added the following description to the manuscript: Big data demands large storage media [26]. However, big data is no longer traditionally processed [27]. Instead, today's big data processing relies on machines that can provide systematic results [28]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [29]. Cloud facilitates cost-effective big data storage and analysis [29]. Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [29]. Target data and preprocessed data are the processes of extracting raw data from big data [29]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.	On page 2, 1st column, paragraph 4 and 5, in Introduction subsection

		Medical record data of disease symptoms obtained in the form of string data is then converted into numeric data.	On page 5, 1st and 2nd columns, in methodology subsection
7	There were many typos, grammatical and colloquial language found. This should be avoided.	The author has revised according to reviewers' suggestions on many parts of the manuscript (the change is marked with a light blue highlight). For example: * The sentence "The data collected **in the form of** medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. " in the Result and Discussion subsection, it is replaced with the sentence "The data collected **from** medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified.". The phrase "in the form of" is replaced with "from". * the sentence "..... and a precision of 92%; ..." in the Conclusion subsection on lines 8 and 9, has been changed to "and a precision of 90%; ...".	On page 6, 1st column, in Result and Discussion subsection, rows 3 to 7. On page 9, 2nd column, in conclusion subsection, paragraph 1.
	Graphical representation is missing. Authors should consider adding more graphical representations of results.	As per reviewers' suggestions, the authors have added figure 2 and figure 4 charts.	
8			
	Reviewer's Comments (Reviewer B)	**Corrections made (the change is marked with a yellow highlight)**	**In Page & paragraph**
1	In section II. Related works, the review of past work or studies by other researchers should be written in a paragraph. Currently, the authors used bullet style to describe each study. Then, a summary of the description should be described in table form. this would be easier for the reader to compare the previous studies with what has been proposed by the authors. Perhaps the table should provide details about the study, the authors, the limitation, etc.	The authors have followed the reviewer's suggestion of changing the related works review (on page 2 to 3 in the Related works subsection) in paragraphs instead of bullet style. Then, have summarized the description in table 1 (on page 4 in the Related works subsection).	On page 2 to 4, in Related Works subsection
2	Background of machine learning algorithms: RF and SVM should be provided and described in the manuscript.	The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [30] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for	On page 2, 2nd column, paragraph 1, in Introduction

		classification [31] [32], as well as SVM algorithm has a widely known technique used for classification [32]. It is why this study uses SVM and RF to classify treatment status.	subsection
3	The review of other machine learning algorithms also should be provided, and the author should conclude or justify why RF and SVM are a good combination for this study.	Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used together in research, including research to classify and/or predict the treatment status of Covid-19 patients.	On page 2, 2nd column, paragraph 1, in Introduction subsection
4	What kind of data was used in this study? Which source? Please cite and justify.	The author has justified the kind of data used in this study (in the methodology subsection) according to the reviewers' suggestions as follows: The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The big data source used in this study is primary data from patient medical records/documents.	On page 5, 1st column, in methodology subsection
5	The authors claimed that this study involved big data. But in section IV Result and Discussion, the data used in this study is 117. The data is too small, thus contradicting the previous statement in the abstract and introduction.	The authors have added a description in the methodology subsection as follows: The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 as well as duplicate and incomplete Covid-19 patient data are removed, so that only data is left as a dataset for data mining processes.	On page 5, 1st paragraph, in the methodology subsection
6	In section 4.4.1, a detailed description of the SVM data mining method is not provided. The authors only provide general stap but did not detail out the process.	The authors have added a description as follows: In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1.	On page 7, 2nd column, in Result and Discussion subsection
7	Section 4.4.2, a detailed description of the RF data mining method is not provided. The authors only provide general stap but did not detail out the process.	The authors have added a description as follows: RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion: $Gini = 1 - \sum_{i=1}^{c} pi^2$ (4) $Entropy(S) = \sum_{i=1}^{c} -pi \times log_2(pi)$ (5) Description: S = Set of cases pi = the proportion of case i to the Set of cases	On page 7, 2nd column, in Result and Discussion subsection
8	The accuracy, Precision and Recall factor is based on what? It is unclear how you measure these factors.	The authors have added a description as follows: Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall	On page 8, column 1, paragraph 5, in Result and Discussion

		positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.	subsection
9	In conclusion, this manuscript is basically comparing two methods of machine learning; RF and SVM. The title did not reflect this. The title should be revised. Example Comparative analysis of …….. The current title could be misleading.	Authors: The manuscript title has been revised to Comparative analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients. Thanks you	

Hopefully, what we have done fulfills the wishes of the reviewers.
Thank you very much. God bless you.

Sincerely yours
Authors

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that, or that's why this study aims to propose a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. INTRODUCTION

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18][19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20][21][22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [23]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. That is, machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32] [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used together in research, including research to classify and/or predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. Whereas, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review research.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 shows a comparison between some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the last research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The big data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Figure 1 [43]. Figure 2 shows the process carried out at each stage of CRISP-DM.

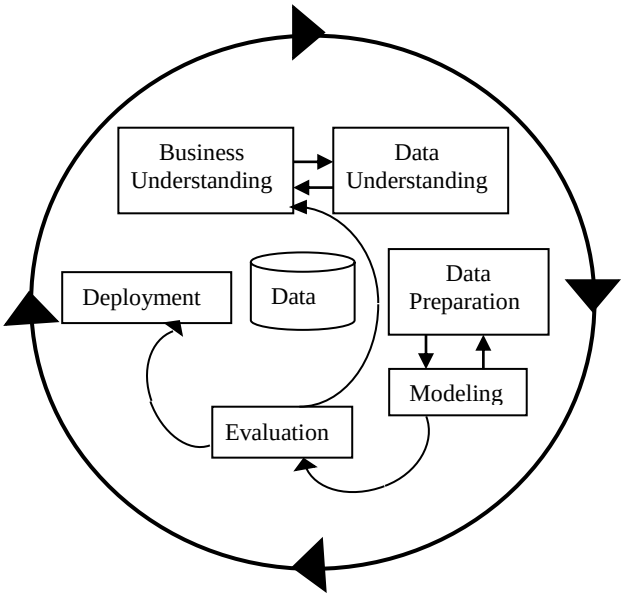

Figure 1. The CRISP-DM Process

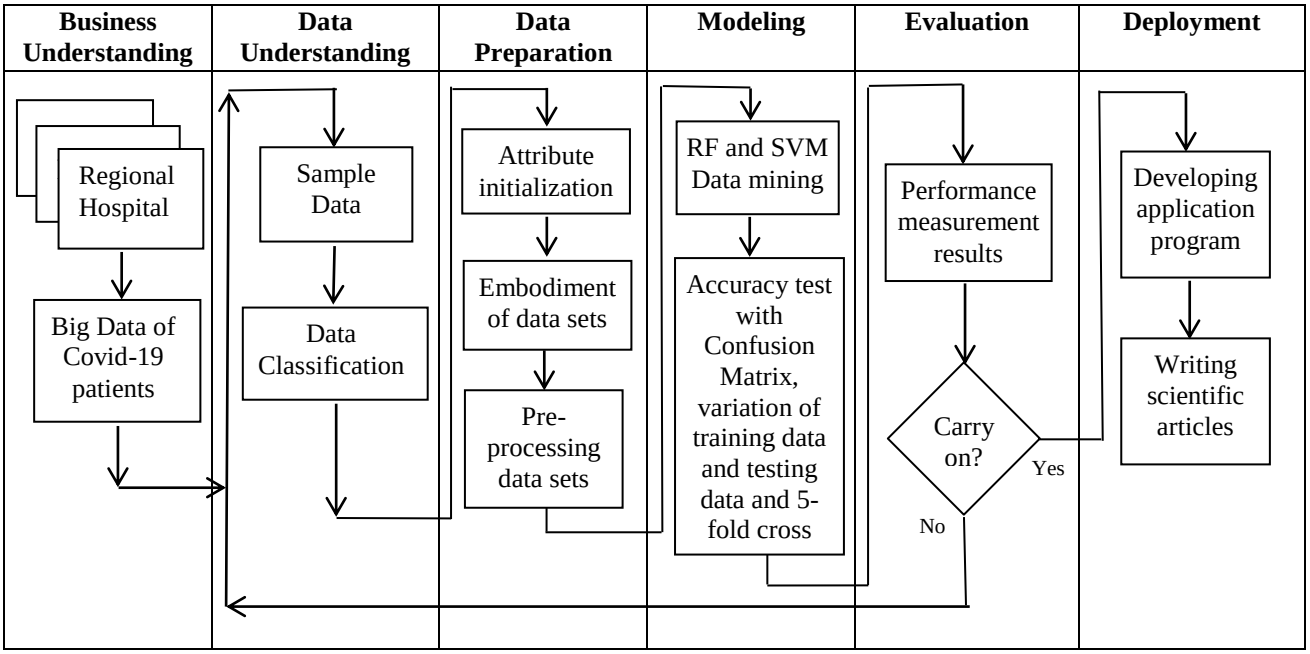

Figure 2. Data Mining Process of Covid-19 Patient Big Data

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [44], which has their respective advantages in building application programs [45][46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction. The hyperparameter used in the SVM method is kernel = linear and Probability = True. Meanwhile, the hyperparameter used for the RF hyperparameter method is n_estimators = 100. The n_estimators is the number of trees in RF.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:

S = Set of cases

pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 6.

TABLE VI. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 7 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE VII. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

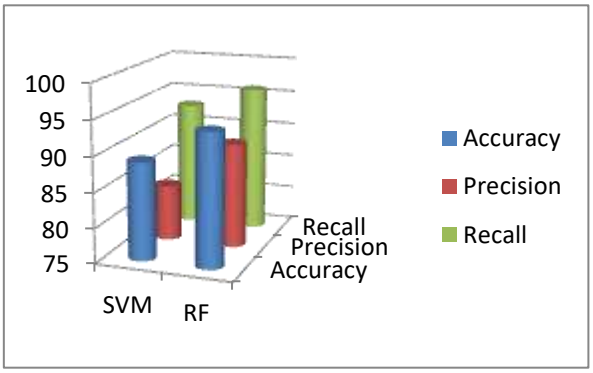

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix. 333

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 9 and Figure 4.

TABLE IX. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

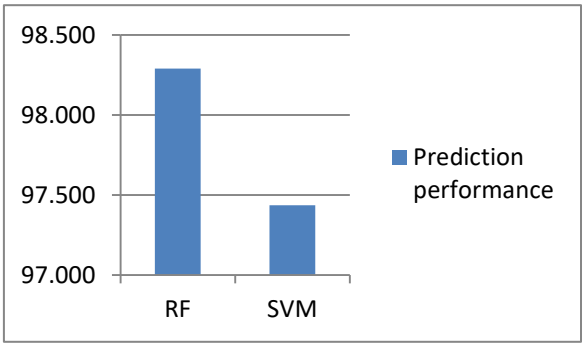

Figure 4. Predictive performance testing using K-fold cross-validation

4.6 Deployment

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of machine learning and the Internet of Things.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li,

"COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.
[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.
[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.
[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.
[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.
[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.
[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.
[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.
[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.
[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.
[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.
[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.
[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.
[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.
[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.
[20] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.
[21] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.
[22] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.
[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.
[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.
[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.
[26] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.
[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.
[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.
[29] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.
[30] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.
[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.
[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.
[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.
[34] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.
[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.
[36] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.
[37] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.
[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.
[39] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.
[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[41] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.
[42] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[43] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[45] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT	7322	TIME SUBMITTED	16-JUL-2022 07:10PM
		PAPER ID	88471119

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that or that's why this study aims to propose data mining classification method as a machine learning model to accurately predict the treatment status of Covid-19 patients, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not over because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. That is, machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [23].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [24]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [22]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [25].

Big data demands large storage media [26]. However, big data is no longer traditionally processed [27]. Instead, today's big data processing relies on machines that can provide systematic results [28]. Big data storage is generally on a computer server with a large storage capacity. Still, software also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [29]. Cloud facilitates cost-effective big data storage and analysis [29].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [29]. Target data and preprocessed data are the processes of extracting raw data from big data [29]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [30] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [31] [32], as well as SVM algorithm has a widely known technique used for classification [32]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used together in research, including research to classify and/or predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Erkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [30]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [33]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [34]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [35]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [36]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research object and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [37]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. Whereas, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [38]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review research.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [39]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [40]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 shows a comparison between some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties (from the last research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are completely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [30]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [33]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [34]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [35]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [22]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [36]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [37]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [38]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [39]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [40]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The big data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The methodology in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is standard data mining. The process in CRISP-DM comprises a six-stage [41], see Figure 1 [42]. Figure 2 shows the data mining processes in this study.

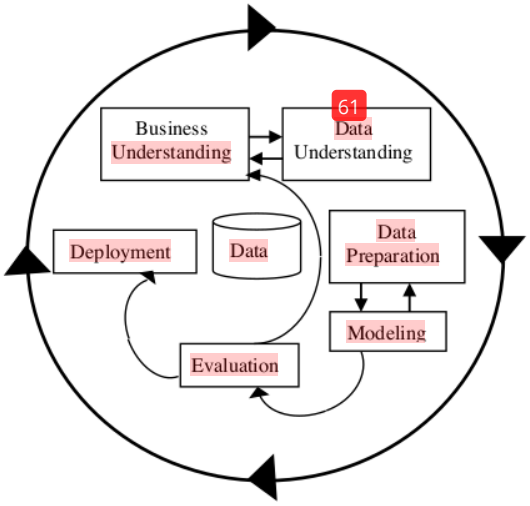

Figure 1. The CRISP-DM Process

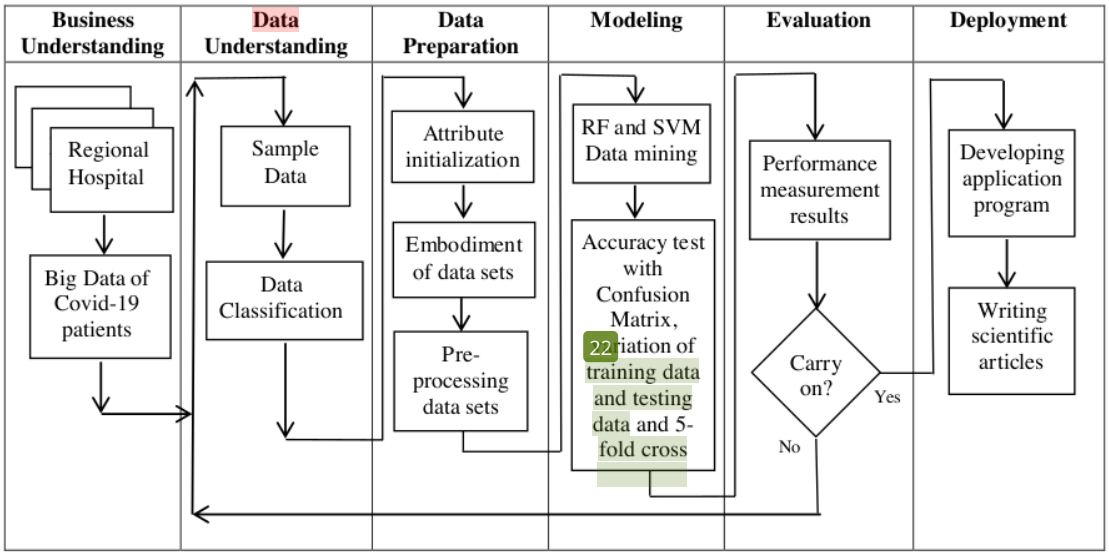

Figure 2. Data Mining Process of Covid-19 Patient Big Data

Business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted/ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension

G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	...	...	G11	G12	Class
1	Yes	No	No	No	...	...	No	No	Inpatient
2	Yes	Yes	No	Yes	...	...	No	No	Inpatient
3	Yes	No	Yes	No	...	...	No	No	Inpatient
4	Yes	No	No	No	...	...	No	No	Inpatient
5	Yes	No	No	No	...	...	No	No	Inpatient
6	Yes	No	No	No	...	...	Yes	Yes	Inpatient
7	No	No	Yes	No	...	...	No	No	Inpatient
..	...	...	...	...	...	...	...	...	...
..	...	...	...	...	...	...	...	...	...
114	No	No	No	No	...	...	No	Yes	Self-isolation
115	Yes	No	No	No	...	...	Yes	No	Inpatient
116	Yes	No	No	No	...	...	Yes	Yes	Inpatient
117	No	No	No	No	...	...	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	...	...	G11	G12	Class
1	1	-1	-1	-1	...	...	-1	-1	-1
2	1	1	-1	1	...	...	-1	-1	-1
3	1	-1	1	-1	...	...	-1	-1	-1
4	1	-1	-1	-1	...	...	-1	-1	-1
5	1	-1	-1	-1	...	...	-1	-1	-1
6	1	-1	-1	-1	...	...	1	1	-1
7	-1	-1	1	-1	...	...	-1	-1	-1
..	...	...	...	...	...	...	...	...	...
..	...	...	...	...	...	...	...	...	...
114	-1	-1	-1	-1	...	...	-1	1	1
115	1	-1	-1	-1	...	...	1	-1	-1
116	1	-1	-1	-1	...	...	1	1	-1
117	-1	-1	-1	-1	...	...	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [43], which has their respective advantages in building application programs [44][45]. The application program built in this research uses the Python programming language to facilitate patient care status prediction. The hyperparameter used in the SVM method is kernel = linear and Probability = True. Meanwhile, the hyperparameter used for the RF hyperparameter method is n_estimators = 100. The n_estimators is the number of trees in RF.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \times log_2(pi) \quad (5)$$

Description:

S = Set of cases

pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion

matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 6.

TABLE VI. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 *Evaluation*

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 7 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

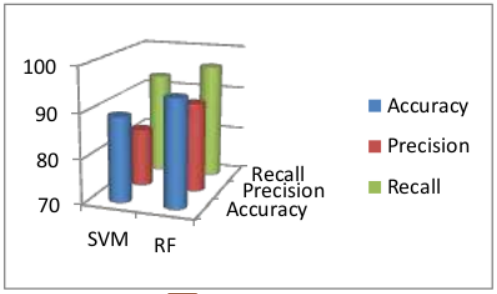

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method

is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix. 333

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 9 and Figure 4.

TABLE IX. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

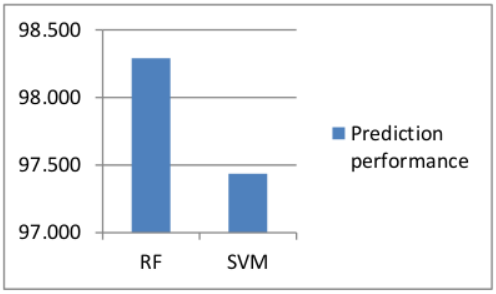

Figure 4. Predictive performance testing using K-fold cross validation

4.6 *Deployment*

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and Precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of machine learning and the Internet of Things.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. ch, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Monitor Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 3, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. 23, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 1, pp. 111–118, 2021.

[23] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[24] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[25] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[26] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[27] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[28] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[29] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.

[30] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. December, pp. 1–9, 2020.

[31] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[32] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[33] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[34] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[35] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[36] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[37] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[38] S. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[39] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[40] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[41] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[42] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[43] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[44] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[45] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

30%
SIMILARITY INDEX

PRIMARY SOURCES

1	thesai.org Internet	123 words — 2%
2	inass.org Internet	111 words — 1%
3	Anthony Anggrawan, Dwi Kurnianingsih, Christofer Satria. "Sistem Aplikasi Cerdas Klasterisasi Penerima Bantuan Covid-19", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Crossref	81 words — 1%
4	journal.universitasbumigora.ac.id Internet	61 words — 1%
5	www.jatit.org Internet	61 words — 1%
6	ijsrcseit.com Internet	53 words — 1%
7	www.mdpi.com Internet	52 words — 1%
8	Daniel Berrar. "Cross-Validation", Elsevier BV, 2018 Crossref	49 words — 1%

9	Haodong Li. "An Overview on Remote Sensing Image Classification Methods with a Focus on Support Vector Machine", 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), 2021 Crossref	45 words — 1%
10	openaccess.altinbas.edu.tr Internet	45 words — 1%
11	www.hindawi.com Internet	44 words — 1%
12	Fermi Dwi Wicaksono, Yusri Bin Arshad, Haeryip Sihombing. "Norm-dist Monte-Carlo integrative method for the improvement of fuzzy analytic hierarchy process", Heliyon, 2020 Crossref	42 words — 1%
13	www.ijeat.org Internet	42 words — 1%
14	Petrisia Widyasari Sudarmadji, Prisca Deviani Pakan, Rocky Yefrenes Dillak. "Diabetic Retinopathy Stages Classification using Improved Deep Learning", 2020 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS), 2020 Crossref	40 words — 1%
15	Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021 Crossref	34 words — < 1%
16	Farshid Rahimibashar, Ladan Sedighi, Alireza Shahriary, Zeljko Reiner et al. "Is there any	33 words — < 1%

association between plasma lipid profile and severity of COVID-19?", Clinical Nutrition ESPEN, 2022
Crossref

17	philpapers.org Internet	33 words	< 1%
18	www.e-journal.unair.ac.id Internet	33 words	< 1%
19	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	32 words	< 1%
20	www.researchsquare.com Internet	32 words	< 1%
21	www.aimspress.com Internet	31 words	< 1%
22	Nurfadila Utami, Mustakim. "Data Distribution Modelling in Supervised Learning Algorithm is for The classification of Prospective Recipient Candidate", 2021 4th International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021 Crossref	28 words	< 1%
23	saucis.sakarya.edu.tr Internet	28 words	< 1%
24	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet	27 words	< 1%
25	Repositorioacademico.Upc.Edu.Pe Internet	27 words	< 1%

26 Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A novel approach to predict COVID-19 using support vector machine", Elsevier BV, 2021 — 26 words — < 1%
Crossref

27 Tawsifur Rahman, Amith Khandakar, Md Enamul Hoque, Nabil Ibtehaz et al. "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access, 2021 — 26 words — < 1%
Crossref

28 egonw.github.io — 26 words — < 1%
Internet

29 Bhabani Shankar Prasad Mishra, Om Pandey, Satchidananda Dehuri, Sung-Bae Cho. "Unsupervised Functional Link Artificial Neural Networks for Cluster Analysis", IEEE Access, 2020 — 25 words — < 1%
Crossref

30 Joshua Guzman, Gwen Arianne Recoco, Al Wahid Pandi, Jerico M. Padrones, Jonathan Jared Ignacio. "Evaluating workplace safety in the oil and gas industry during the COVID-19 pandemic using occupational health and safety Vulnerability Measure and partial least square Structural Equation Modelling", Cleaner Engineering and Technology, 2022 — 25 words — < 1%
Crossref

31 Tora Fahrudin, Nelsi Wisna, Patrick Adolf Telnoni, Dedy Rahman Wijaya. "Sales Forecasting Web Application in Small and Medium Enterprise", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022 — 25 words — < 1%
Crossref

32	Vimal Kamleshkumar Bhatt. "Assessing the Significance and Impact of Artificial Intelligence and Machine Learning in Placement of Advertisements", 2021 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), 2021 Crossref	25 words — < 1%
33	smartcities.ieee.org Internet	25 words — < 1%
34	tj.uettaxila.edu.pk Internet	25 words — < 1%
35	seaninstitute.org Internet	24 words — < 1%
36	www.gtu.edu.tr Internet	24 words — < 1%
37	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018 Crossref	23 words — < 1%
38	jurnal.ugm.ac.id Internet	22 words — < 1%
39	www.ijppr.humanjournals.com Internet	21 words — < 1%
40	Qingyun Ren, Bingyin Zhou, Liang Tian, Wei Guo. "Detection of COVID-19 With CT Images Using Hybrid Complex Shearlet Scattering Networks", IEEE Journal of Biomedical and Health Informatics, 2022 Crossref	20 words — < 1%

41 Catherine P. Adans-Dester, Stacy Bamberg, Francesco P. Bertacchi, Brian Caulfield et al. "Can mHealth Technology Help Mitigate the Effects of the COVID-19 Pandemic?", IEEE Open Journal of Engineering in Medicine and Biology, 2020
Crossref
19 words — < 1%

42 Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita. "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review", Computational and Structural Biotechnology Journal, 2021
Crossref
19 words — < 1%

43 Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022
Crossref
19 words — < 1%

44 bura.brunel.ac.uk
Internet
19 words — < 1%

45 dl.acm.org
Internet
19 words — < 1%

46 eprints.utm.my
Internet
19 words — < 1%

47 ijece.iaescore.com
Internet
19 words — < 1%

48 Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020
Crossref
18 words — < 1%

49	journalofbigdata.springeropen.com Internet	17 words — < 1%
50	www.researchgate.net Internet	17 words — < 1%
51	annals-csis.org Internet	16 words — < 1%
52	files.covid19treatmentguidelines.nih.gov Internet	16 words — < 1%
53	journals.gaftim.com Internet	16 words — < 1%
54	journals.plos.org Internet	16 words — < 1%
55	www.etasr.com Internet	16 words — < 1%
56	www.iseg.ulisboa.pt Internet	15 words — < 1%
57	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	14 words — < 1%
58	Sajana Maharjan, Faisal Mueen Qamer, Mir Matin, Govinda Joshi, Sanjeev Bhuchar. "Integrating Modelling and Expert Knowledge for Evaluating Current and Future Scenario of Large Cardamom Crop in Eastern Nepal", Agronomy, 2019 Crossref	14 words — < 1%
	www.ncbi.nlm.nih.gov	

59	Internet	14 words — < 1%
60	Roberto Magherini, Elisa Mussi, Michaela Servi, Yary Volpe. "Emotion recognition in the times of COVID19: Coping with face masks", Intelligent Systems with Applications, 2022 Crossref	13 words — < 1%
61	nadre.ethernet.edu.et Internet	13 words — < 1%
62	Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19", Physiological Genomics, 2020 Crossref	12 words — < 1%
63	David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 Crossref	12 words — < 1%
64	I Ayuningtias, I Jaya, M Iqbal. "Identification of yellowfin tuna (Thunnus albacares), mackerel tuna (Euthynnus affinis), and skipjack tuna (Katsuwonus pelamis) using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 Crossref	12 words — < 1%
65	Spiridon Vergis, Konstantinos Bezas, Themis P. Exarchos. "Chapter 3 Decision Support System for Breast Cancer Detection Using Biomarker Indicators", Springer Science and Business Media LLC, 2021 Crossref	12 words — < 1%

66 Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022
Crossref
12 words — < 1%

67 ejournal.uin-suska.ac.id
Internet
12 words — < 1%

68 riull.ull.es
Internet
12 words — < 1%

69 Cryssa A. E. Piter, Setiwan Hadi, Intan Nurma Yulita. "Multi-Label Classification for Scientific Conference Activities Information Text Using Extreme Gradient Boost (XGBoost) Method", 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021
Crossref
11 words — < 1%

70 Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021
Crossref
11 words — < 1%

71 Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020
Crossref
11 words — < 1%

72 arxiv.org
Internet
11 words — < 1%

73 docplayer.net

Internet

11 words — < 1%

74 core.ac.uk
Internet

10 words — < 1%

75 dokumen.pub
Internet

10 words — < 1%

76 elifesciences.org
Internet

10 words — < 1%

77 tojqi.net
Internet

10 words — < 1%

78 www.semanticscholar.org
Internet

10 words — < 1%

79 Anita Desiani, Adinda Ayu Lestari, M Al-Ariq, Ali Amran, Yuli Andriani. "Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification", Pattimura International Journal of Mathematics (PIJMath), 2022
Crossref

9 words — < 1%

80 Dennis Peace Ezeobi, William Wasswa, Angella Musimenta, Stella Kyoyagala. "A Neonatal Sepsis Prediction Algorithm Using Electronic Medical Record (EMR) Data from Mbarara Regional Referral Hospital (MRRH)", Research Square Platform LLC, 2022
Crossref Posted Content

9 words — < 1%

81 Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. ""F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability",

9 words — < 1%

2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021
Crossref

82	eprints.ucm.es Internet	9 words — < 1%
83	www.bmj.com Internet	9 words — < 1%
84	www.medrxiv.org Internet	9 words — < 1%
85	www.science.gov Internet	9 words — < 1%
86	M. Fatih Aslan, Kadir Sabanci, Ewa Ropelewska. "A CNN-Based Solution for Breast Cancer Detection With Blood Analysis Data: Numeric to Image", 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021 Crossref	8 words — < 1%
87	Rachmadita Andreswari, Dita Aprilia Rahmani, Ria Rahmawati, Achmad Akbar H. "Comparative Analysis of K-Means and K-Nearest Neighbor Algorithm for Telecom Fraud Detection", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022 Crossref	8 words — < 1%
88	Sushil Kumar Mishra, Nitish Kumar Ojha. "A Novel Approach for Face Mask Detection using Tensorflow for Covid-19", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021 Crossref	8 words — < 1%

89	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
90	annalsofrscb.ro Internet	8 words — < 1%
91	eprints.umm.ac.id Internet	8 words — < 1%
92	etd.aau.edu.et Internet	8 words — < 1%
93	ijai.iaescore.com Internet	8 words — < 1%
94	jurnal.iaii.or.id Internet	8 words — < 1%
95	preprints.aijr.org Internet	8 words — < 1%
96	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	7 words — < 1%
97	Annisa Nurul Puteri, Dewiani, Zulkifli Tahir. "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods", 2019 International Conference	6 words — < 1%

on Information and Communications Technology (ICOIACT), 2019
Crossref

98 Li, Dan, Yinghai Ke, Huili Gong, and Xiaojuan Li. "Object-Based Urban Tree Species Classification Using Bi-Temporal WorldView-2 and WorldView-3 Images", Remote Sensing, 2015. — 6 words — < 1%
Crossref

99 Melissa A. Hausburg, Jason S. Williams, Kaysie L. Banton, Charles W. Mains, Michael Roshon, David Bar-Or. "C1 esterase inhibitor-mediated immunosuppression in COVID-19: friend or foe?", Clinical Immunology Communications, 2022 — 6 words — < 1%
Crossref

100 Muhammad Daniyal, Kassim Tawiah, Sara Muhammadullah, Kwaku Opoku-Ameyaw. "Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data", Journal of Healthcare Engineering, 2022 — 6 words — < 1%
Crossref

101 Seongdo Kim, Do Hyoung Shin. "Forecasting short-term air passenger demand using big data from search engine queries", Automation in Construction, 2016 — 6 words — < 1%
Crossref

102 Shilei Lu, Yang Liu, Yuwei Li, Ran Wang. "Multidimensional performance-based evaluation method of high-performance cold source in green building", Energy and Buildings, 2021 — 6 words — < 1%
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT	7322	TIME SUBMITTED	16-JUL-2022 07:10PM
		PAPER ID	88471119

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that or that's why this study aims to propose a data mining classification method as a machine learning model to accurately predict the treatment status of Covid-19 patients, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [19][20][21]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [22]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. That is, machine learning performs tasks like a medical specialist in determining the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [23].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [24]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [22]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [25].

Big data demands large storage media [26]. However, big data is no longer traditionally processed [27]. Instead, today's big data processing relies on machines that can provide systematic results [28]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [29]. Cloud facilitates cost-effective big data storage and analysis [29].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [29]. Target data and preprocessed data are the processes of extracting raw data from big data [29]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [30] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [31] [32], as well as SVM algorithm has a widely known technique used for classification [32]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used together in research, including research to classify and/or predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Erkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [30]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [33]. The

difference between previous research and the research in this article is that the previous resea only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [34]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [35]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et a (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [22]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [36]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research object and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [37]. This previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. Whereas, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [38]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review research.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [39]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [40]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 shows a comparison between some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent previous related work by some researchers, the research carried out in this article has novelties (from the last research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are completely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [30]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [33]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [34]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [35]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [22]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [36]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [37]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [38]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [39]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [40]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data used is data on Covid-19 patients from a regional hospital in Mataram, Indonesia. The big data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The methodology in this study uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is standard data mining. The process in CRISP-DM comprises a six-stage [41], see Figure 1 [42]. Figure 2 shows the data mining processes in this study.

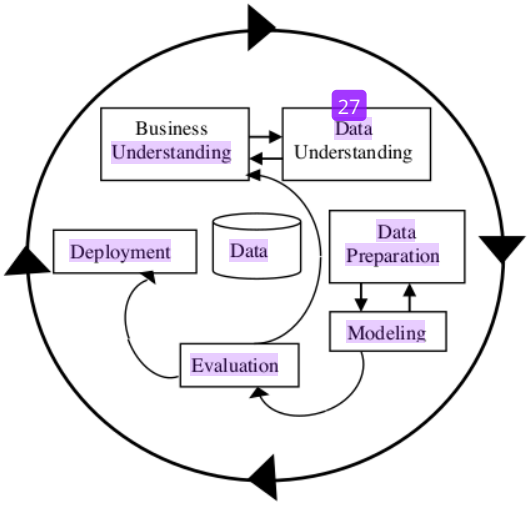

Figure 1. The CRISP-DM Process

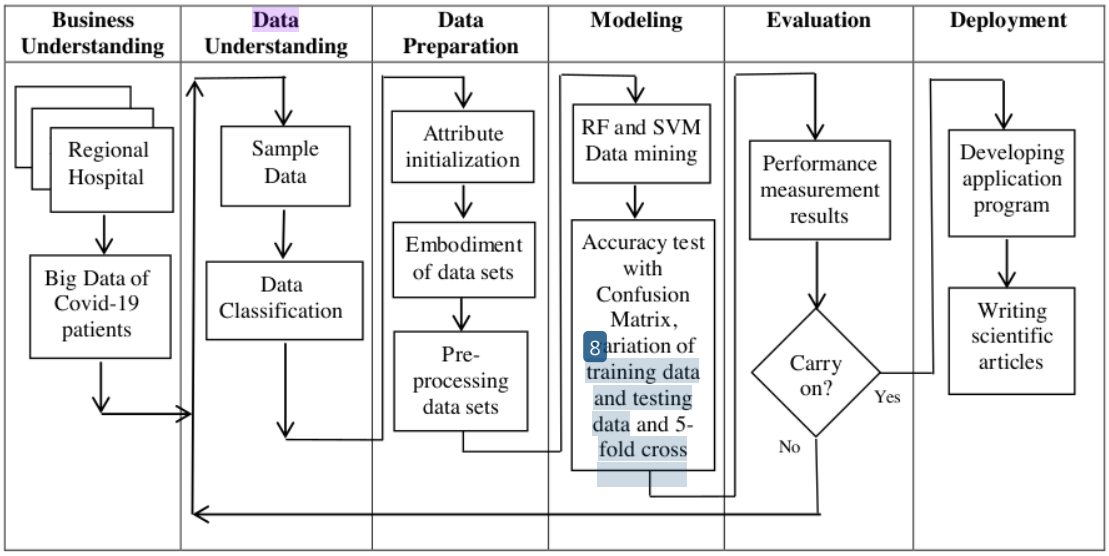

Figure 2. Data Mining Process of Covid-19 Patient Big Data

Business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted/ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension

G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, Known various programming language [43], which has their respective advantages in building application programs [44][45]. The application program built in this research uses the Python programming language to facilitate patient care status prediction. The hyperparameter used in the SVM method is kernel = linear and Probability = True. Meanwhile, the hyperparameter used for the RF hyperparameter method is n_estimators = 100. The n_estimators is the number of trees in RF.

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree. The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \times log_2(pi) \quad (5)$$

Description:

S = Set of cases

pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion

matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 6.

TABLE VI. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set the remaining subset is the training set, and so on until the 5th fold.

4.5 *Evaluation*

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 7 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE 8 SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

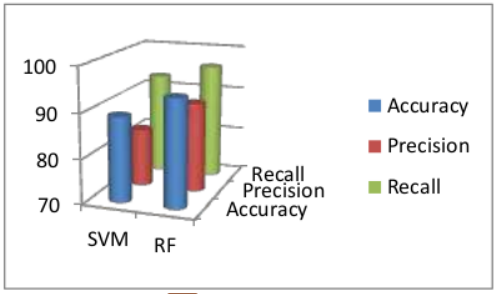

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 8.

TABLE VIII. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method

is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix. 333

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 9 and Figure 4.

TABLE IX. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

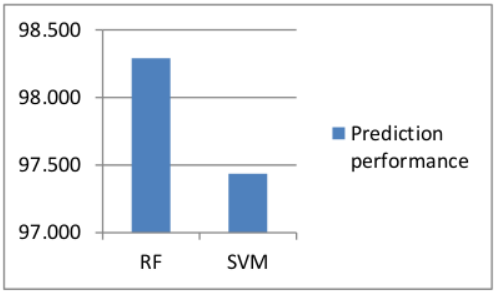

Figure 4. Predictive performance testing using K-fold cross validation

4.6 *Deployment*

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and Precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of machine learning and the Internet of Things.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis: a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[19] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[20] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[21] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[22] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[23] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[24] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[25] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[26] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[27] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[28] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[29] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.

[30] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[31] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[32] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[33] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[34] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[35] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[36] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[37] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[38] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[39] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[40] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[41] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[42] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[43] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[44] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[45] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

17%
SIMILARITY INDEX

PRIMARY SOURCES

1	inass.org Internet	82 words — 1%
2	www.mdpi.com Internet	52 words — 1%
3	Daniel Berrar. "Cross-Validation", Elsevier BV, 2018 Crossref	49 words — 1%
4	www.jatit.org Internet	48 words — 1%
5	www.ncbi.nlm.nih.gov Internet	40 words — 1%
6	thesai.org Internet	34 words — 1%
7	Norah Alballa, Isra Al-Turaiki. "Machine Learning Approaches in COVID-19 Diagnosis, Mortality, and Severity Risk Prediction: A Review", Informatics in Medicine Unlocked, 2021 Crossref	32 words — 1%
8	Nurfadila Utami, Mustakim. "Data Distribution Modelling in Supervised Learning Algorithm is for The classification of Prospective Recipient Candidate", 2021 4th	28 words — < 1%

International Seminar on Research of Information Technology and Intelligent Systems (ISRITI), 2021
Crossref

9 Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A novel approach to predict COVID-19 using support vector machine", Elsevier BV, 2021 — 26 words — < 1%
Crossref

10 journal.universitasbumigora.ac.id — 26 words — < 1%
Internet

11 files.covid19treatmentguidelines.nih.gov — 24 words — < 1%
Internet

12 Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018 — 23 words — < 1%
Crossref

13 jurnal.ugm.ac.id — 22 words — < 1%
Internet

14 www.hindawi.com — 20 words — < 1%
Internet

15 Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 — 19 words — < 1%
Crossref

16 bura.brunel.ac.uk — 19 words — < 1%
Internet

17	eprints.utm.my Internet	19 words — < 1%
18	ijece.iaescore.com Internet	19 words — < 1%
19	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%
20	www.researchgate.net Internet	17 words — < 1%
21	annals-csis.org Internet	16 words — < 1%
22	journals.gaftim.com Internet	16 words — < 1%
23	www.ijstr.org Internet	15 words — < 1%
24	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	14 words — < 1%
25	Sajana Maharjan, Faisal Mueen Qamer, Mir Matin, Govinda Joshi, Sanjeev Bhuchar. "Integrating Modelling and Expert Knowledge for Evaluating Current and Future Scenario of Large Cardamom Crop in Eastern Nepal", Agronomy, 2019 Crossref	14 words — < 1%

26 Roberto Magherini, Elisa Mussi, Michaela Servi, Yary Volpe. "Emotion recognition in the times of COVID19: Coping with face masks", Intelligent Systems with Applications, 2022 — 13 words — < 1%
Crossref

27 nadre.ethernet.edu.et — 13 words — < 1%
Internet

28 Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19", Physiological Genomics, 2020 — 12 words — < 1%
Crossref

29 David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 — 12 words — < 1%
Crossref

30 I Ayuningtias, I Jaya, M Iqbal. "Identification of yellowfin tuna (Thunnus albacares), mackerel tuna (Euthynnus affinis), and skipjack tuna (Katsuwonus pelamis) using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 — 12 words — < 1%
Crossref

31 Spiridon Vergis, Konstantinos Bezas, Themis P. Exarchos. "Chapter 3 Decision Support System for Breast Cancer Detection Using Biomarker Indicators", Springer Science and Business Media LLC, 2021 — 12 words — < 1%
Crossref

32 Yi Zhang, Shuwang Yang, Hang Zhang. "Research on Urban Traffic Industrial Management under Big Data: Taking Traffic Congestion as an Example", Journal of Advanced Transportation, 2022 — 12 words — < 1%

Crossref

33	riull.ull.es Internet	12 words — < 1%
34	Cryssa A. E. Piter, Setiwan Hadi, Intan Nurma Yulita. "Multi-Label Classification for Scientific Conference Activities Information Text Using Extreme Gradient Boost (XGBoost) Method", 2021 International Conference on Artificial Intelligence and Big Data Analytics, 2021 Crossref	11 words — < 1%
35	Dmytro Chumachenko, Tetyana Chumachenko, Ievgen Meniailov, Olena Muradyan, Grigoriy Zholtkevych. "Forecasting of COVID-19 Epidemic Process by Lasso Regression", 2021 IEEE International Conference on Information and Telecommunication Technologies and Radio Electronics (UkrMiCo), 2021 Crossref	11 words — < 1%
36	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	11 words — < 1%
37	arxiv.org Internet	11 words — < 1%
38	docplayer.net Internet	11 words — < 1%
39	core.ac.uk Internet	10 words — < 1%
40	dokumen.pub Internet	10 words — < 1%

41	elifesciences.org Internet	10 words — < 1%
42	www.semanticscholar.org Internet	10 words — < 1%
43	Anita Desiani, Adinda Ayu Lestari, M Al-Ariq, Ali Amran, Yuli Andriani. "Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification", Pattimura International Journal of Mathematics (PIJMath), 2022 Crossref	9 words — < 1%
44	Dennis Peace Ezeobi, William Wasswa, Angella Musimenta, Stella Kyoyagala. "A Neonatal Sepsis Prediction Algorithm Using Electronic Medical Record (EMR) Data from Mbarara Regional Referral Hospital (MRRH)", Research Square Platform LLC, 2022 Crossref Posted Content	9 words — < 1%
45	Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. ""F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability", 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021 Crossref	9 words — < 1%
46	eprints.ucm.es Internet	9 words — < 1%
47	www.bmj.com Internet	9 words — < 1%
48	www.medrxiv.org Internet	9 words — < 1%

49	www.science.gov Internet	9 words — < 1%
50	M. Fatih Aslan, Kadir Sabanci, Ewa Ropelewska. "A CNN-Based Solution for Breast Cancer Detection With Blood Analysis Data: Numeric to Image", 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021 Crossref	8 words — < 1%
51	Rachmadita Andreswari, Dita Aprilia Rahmani, Ria Rahmawati, Achmad Akbar H. "Comparative Analysis of K-Means and K-Nearest Neighbor Algorithm for Telecom Fraud Detection", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022 Crossref	8 words — < 1%
52	Sushil Kumar Mishra, Nitish Kumar Ojha. "A Novel Approach for Face Mask Detection using Tensorflow for Covid-19", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021 Crossref	8 words — < 1%
53	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
54	annalsofrscb.ro Internet	8 words — < 1%
55	eprints.umm.ac.id Internet	8 words — < 1%
56	etd.aau.edu.et Internet	8 words — < 1%

57	journalofbigdata.springeropen.com Internet	8 words — < 1%
58	jurnal.iaii.or.id Internet	8 words — < 1%
59	openaccess.altinbas.edu.tr Internet	8 words — < 1%
60	www.researchsquare.com Internet	8 words — < 1%
61	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	7 words — < 1%
62	Annisa Nurul Puteri, Dewiani, Zulkifli Tahir. "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019 Crossref	6 words — < 1%
63	Melissa A. Hausburg, Jason S. Williams, Kaysie L. Banton, Charles W. Mains, Michael Roshon, David Bar-Or. "C1 esterase inhibitor-mediated immunosuppression in COVID-19: friend or foe?", Clinical Immunology Communications, 2022 Crossref	6 words — < 1%
64	Muhammad Daniyal, Kassim Tawiah, Sara Muhammadullah, Kwaku Opoku-Ameyaw.	6 words — < 1%

"Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data", Journal of Healthcare Engineering, 2022
Crossref

65 Seongdo Kim, Do Hyoung Shin. "Forecasting short-term air passenger demand using big data from search engine queries", Automation in Construction, 2016 — 6 words — < 1%
Crossref

66 Shilei Lu, Yang Liu, Yuwei Li, Ran Wang. "Multidimensional performance-based evaluation method of high-performance cold source in green building", Energy and Buildings, 2021 — 6 words — < 1%
Crossref

67 Suresh Vatakuti, Jeroen L. A. Pennings, Emilia Gore, Peter Olinga, Geny M. M. Groothuis. "Classification of Cholestatic and Necrotic Hepatotoxicants Using Transcriptomics on Human Precision-Cut Liver Slices", Chemical Research in Toxicology, 2016 — 6 words — < 1%
Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

Correction Note for Second Round

Manuscript ID: JAIT-5790
Title: Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Dear Reviewers
We have made improvements to our manuscript following the instructions and suggestions of the reviewers. About which parts we have revised and added writing, it is as explained below:

<table>
<tr><th></th><th>Comments to Authors</th><th>The Improvements/Corrections
(the correction is marked with a yellow highlight)</th><th>In-Page</th></tr>
<tr><td>1</td><td>I suggested the authors to report the hyperparameter details on the ML models. But they are partially addressed. There are many hyperparameters available for SVM and RF (like loss, optimizer, max_iterations, etc.,), not just two. This is very crucial for research in ML. I request the authors to report the details in a table.</td><td>TABLE VI. USE OF HYPERPARAMETER ON SVM AND RF METHOD
<table><tr><th>Classifier Method</th><th>Hyperparameter</th><th>Value</th></tr><tr><td>SVM</td><td>C
Kernel
Degree
Gamma
Coef
Tol

Max_iter</td><td>1
Rbf
3
Scale
0
0,000
1
-1</td></tr><tr><td>RF</td><td>n_estimators
Criterion
Max_depth
Min_samples_split
Min_samples_leaf</td><td>100
Gini
None
2
1</td></tr></table></td><td>In-Page 7, 2nd column</td></tr>
<tr><td>2</td><td>Confusion matrix is still not plotted. Figure 3 does not represent a confusion matrix. A confusion matrix is essential to know the classification performance of the model for classes.</td><td>TABLE VIII. CONFUSION MATRIX OF SVM
<table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>Class</th><th>Self-isolation</th><th>Inpatient</th></tr><tr><td rowspan="2">Actual</td><td>Self-isolation</td><td>4</td><td>0</td></tr><tr><td>Inpatient</td><td>2</td><td>12</td></tr></table>TABLE IX. CONFUSION MATRIX OF RF
<table><tr><th colspan="2" rowspan="2"></th><th colspan="2">Prediction</th></tr><tr><th>Class</th><th>Self-isolation</th><th>Inpatient</th></tr><tr><td rowspan="2">Actual</td><td>Self-isolation</td><td>4</td><td>0</td></tr><tr><td>Inpatient</td><td>1</td><td>13</td></tr></table></td><td>In-Page 8, 2nd column</td></tr>
<tr><td>3</td><td>Still some language errors are prevalent</td><td>The English language of the manuscript has undergone extensive English editing</td><td></td></tr>
</table>

Hopefully, what we have done fulfills the wishes of the reviewers. Thank you very much.

Sincerely yours
Authors

Plagiarism Check _ Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT	7491	TIME SUBMITTED	03-AUG-2022 02:46PM
		PAPER ID	89134525

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning predict with high accuracy. For that, or that's why this study aims to propose data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both in patient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18][19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20][21][22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [23]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. Machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the treating status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, so also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32] [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used in research, including research to classify and predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Erkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. At the same time, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 compares some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the prior research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the novelty's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is on Covid-19 patients from a regional hospital in Mataram, Indonesia. The significant data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Figure 1 [43]. Figure 2 shows the process carried out at each stage of CRISP-DM.

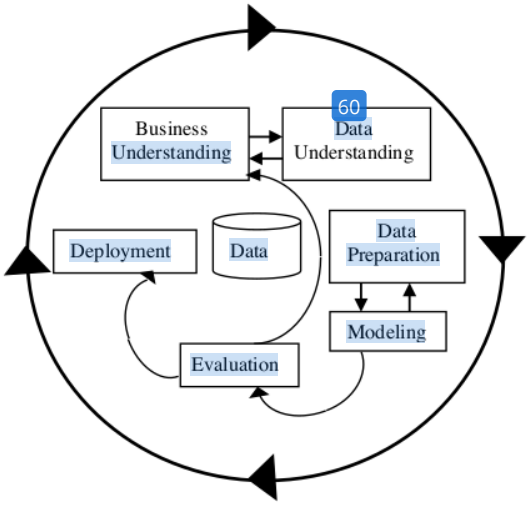

Figure 1. The CRISP-DM Process

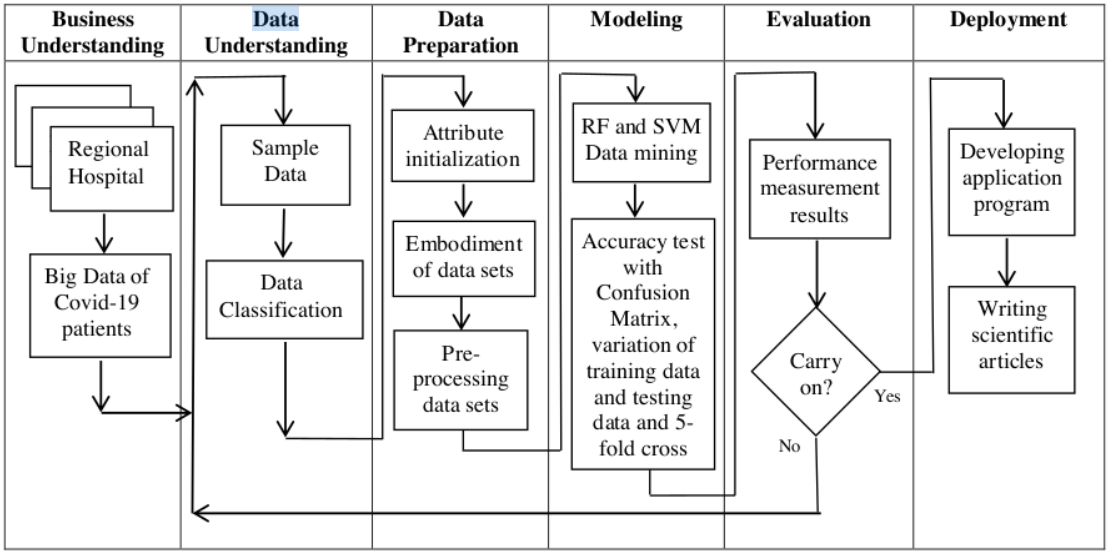

Figure 2. Data Mining Process of Covid-19 Patient Big Data

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs or symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, known various programming language [44], which has their respective advantages in building application programs [45][46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction (as show in Table 6).

TABLE VI. USE OF HYPERPARAMETER ON SVM AND RF METHOD

Classifier Method	Hyperparameter	Value
SVM		1
	Kernel	Rbf
	Degree	3
	Gamma	Scale
	Coef	0
	Tol	0,000
		1
	Max_iter	-1
Random Forest	n_estimators	100
	Criterion	Gini
	Max_depth	None
	Min_samples_split	2
	Min_samples_leaf	1

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree.

The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:
S = Set of cases
pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 7. The results of the predictions of the SVM and RF methods are shown in Tables 8 and 9.

TABLE VII. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

TABLE VIII. CONFUSION MATRIX OF SVM

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	2	12

TABLE IX. CONFUSION MATRIX OF RF

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	1	13

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 10 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE X. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

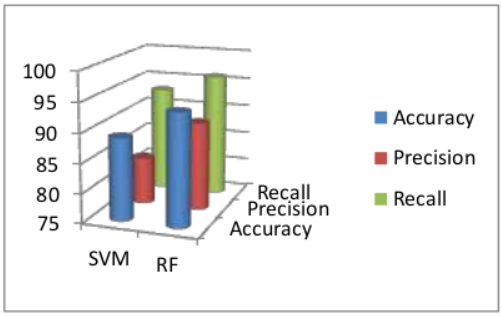

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy, precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 11.

TABLE XI. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 12 and Figure 4.

TABLE XII. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

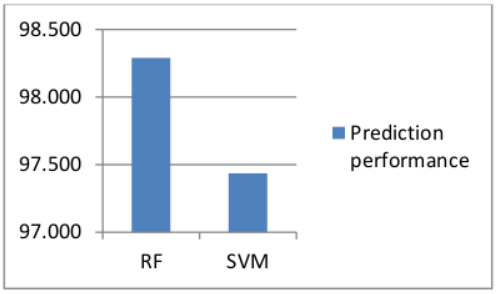

Figure 4. Predictive performance testing using K-fold cross-validation

4.6 Deployment

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of

machine learning and the Internet of Things.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

References

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2833–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a Laboratory Study of Expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.

[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 11, pp. 319–326, 2021.

[20] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, May, pp. 1–17, 2020.

[21] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[22] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[26] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[29] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[30] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.

[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[34] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[36] R. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[37] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[39] H. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship

Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[41] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–2022.

[42] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[43] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[45] A. Anggrawan, C. K. Nugraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Plagiarism Check _ Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

30%
SIMILARITY INDEX

PRIMARY SOURCES

1	thesai.org Internet	132 words — 2%
2	inass.org Internet	111 words — 1%
3	www.jatit.org Internet	102 words — 1%
4	Anthony Anggrawan, Dwi Kurnianingsih, Christofer Satria. "Sistem Aplikasi Cerdas Klasterisasi Penerima Bantuan Covid-19", MATRIK : Jurnal Manajemen, Teknik Informatika dan Rekayasa Komputer, 2022 Crossref	81 words — 1%
5	journal.universitasbumigora.ac.id Internet	61 words — 1%
6	ijsrcseit.com Internet	53 words — 1%
7	www.mdpi.com Internet	52 words — 1%
8	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%

9	Haodong Li. "An Overview on Remote Sensing Image Classification Methods with a Focus on Support Vector Machine", 2021 International Conference on Signal Processing and Machine Learning (CONF-SPML), 2021 Crossref	45 words — 1%
10	openaccess.altinbas.edu.tr Internet	45 words — 1%
11	www.bmj.com Internet	44 words — 1%
12	Fermi Dwi Wicaksono, Yusri Bin Arshad, Haeryip Sihombing. "Norm-dist Monte-Carlo integrative method for the improvement of fuzzy analytic hierarchy process", Heliyon, 2020 Crossref	42 words — 1%
13	Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021 Crossref	42 words — 1%
14	S V Deshmukh, A Roy. "An Empirical Exploration of Artificial Intelligence in Medical Domain for Prediction and Analysis of Diabetic Retinopathy: Review", Journal of Physics: Conference Series, 2021 Crossref	40 words — 1%
15	egonw.github.io Internet	34 words — < 1%
16	www.ijeat.org Internet	34 words — < 1%

17	Farshid Rahimibashar, Ladan Sedighi, Alireza Shahriary, Zeljko Reiner et al. "Is there any association between plasma lipid profile and severity of COVID-19?", Clinical Nutrition ESPEN, 2022 Crossref	33 words — < 1%
18	philpapers.org Internet	33 words — < 1%
19	www.e-journal.unair.ac.id Internet	33 words — < 1%
20	www.hindawi.com Internet	33 words — < 1%
21	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	32 words — < 1%
22	www.aimspress.com Internet	31 words — < 1%
23	saucis.sakarya.edu.tr Internet	28 words — < 1%
24	2dc40e33-085f-40e0-8172-9a1f898c1942.filesusr.com Internet	27 words — < 1%
25	Geng-Xin Xu, Chen Liu, Jun Liu, Zhongxiang Ding et al. "Cross-Site Severity Assessment of COVID-19 from CT Images via Domain Adaptation", IEEE Transactions on Medical Imaging, 2021 Crossref	27 words — < 1%

26	www.researchgate.net Internet	27 words — < 1%
27	Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A novel approach to predict COVID-19 using support vector machine", Elsevier BV, 2021 Crossref	26 words — < 1%
28	Tawsifur Rahman, Amith Khandakar, Md Enamul Hoque, Nabil Ibtehaz et al. "Development and Validation of an Early Scoring System for Prediction of Disease Severity in COVID-19 Using Complete Blood Count Parameters", IEEE Access, 2021 Crossref	26 words — < 1%
29	Bhabani Shankar Prasad Mishra, Om Pandey, Satchidananda Dehuri, Sung-Bae Cho. "Unsupervised Functional Link Artificial Neural Networks for Cluster Analysis", IEEE Access, 2020 Crossref	25 words — < 1%
30	Joshua Guzman, Gwen Arianne Recoco, Al Wahid Pandi, Jerico M. Padrones, Jonathan Jared Ignacio. "Evaluating workplace safety in the oil and gas industry during the COVID-19 pandemic using occupational health and safety Vulnerability Measure and partial least square Structural Equation Modelling", Cleaner Engineering and Technology, 2022 Crossref	25 words — < 1%
31	Tora Fahrudin, Nelsi Wisna, Patrick Adolf Telnoni, Dedy Rahman Wijaya. "Sales Forecasting Web Application in Small and Medium Enterprise", 2021 International Seminar on Machine Learning, Optimization, and Data Science (ISMODE), 2022 Crossref	25 words — < 1%

32	Vimal Kamleshkumar Bhatt. "Assessing the Significance and Impact of Artificial Intelligence and Machine Learning in Placement of Advertisements", 2021 IEEE International Conference on Technology Management, Operations and Decisions (ICTMOD), 2021 Crossref	25 words — < 1%
33	smartcities.ieee.org Internet	25 words — < 1%
34	Dan Song, Ruofeng Tong, Jian Chang, Tongtong Wang, Jiang Du, Min Tang, Jian J. Zhang. "Chapter 7 Clothes Size Prediction from Dressed-Human Silhouettes", Springer Science and Business Media LLC, 2017 Crossref	24 words — < 1%
35	seaninstitute.org Internet	24 words — < 1%
36	www.gtu.edu.tr Internet	24 words — < 1%
37	Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018 Crossref	23 words — < 1%
38	jurnal.ugm.ac.id Internet	22 words — < 1%
39	www.ijppr.humanjournals.com Internet	21 words — < 1%
40	www.ncbi.nlm.nih.gov Internet	21 words — < 1%

41	Qingyun Ren, Bingyin Zhou, Liang Tian, Wei Guo. "Detection of COVID-19 With CT Images Using Hybrid Complex Shearlet Scattering Networks", IEEE Journal of Biomedical and Health Informatics, 2022 Crossref	20 words —	< 1%
42	Xiang Fang, Yuchong Hu, Pan Zhou, Dapeng Oliver Wu. "V^{3}H: View Variation and View Heredity for Incomplete Multiview Clustering", IEEE Transactions on Artificial Intelligence, 2020 Crossref	20 words —	< 1%
43	Eleni S. Adamidi, Konstantinos Mitsis, Konstantina S. Nikita. "Artificial intelligence in clinical care amidst COVID-19 pandemic: A systematic review", Computational and Structural Biotechnology Journal, 2021 Crossref	19 words —	< 1%
44	I Gede Susrama Mas Diyasa, Agus Prayogi, Intan Yuniar Purbasari, Ariyono Setiawan, Sugiarto, Prismahardi Aji Riantoko. "Data Classification of Patient Characteristics Based on Nutritional Treatment Using the K-Nearest Neighbors Algorithm", 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), 2021 Crossref	19 words —	< 1%
45	Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022 Crossref	19 words —	< 1%
46	bura.brunel.ac.uk Internet	19 words —	< 1%
47	dl.acm.org Internet	19 words —	< 1%

48	eprints.utm.my Internet	19 words — < 1%
49	harvest.usask.ca Internet	19 words — < 1%
50	I Ayuningtias, I Jaya, M Iqbal. "Identification of yellowfin tuna (Thunnus albacares), mackerel tuna (Euthynnus affinis), and skipjack tuna (Katsuwonus pelamis) using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 Crossref	18 words — < 1%
51	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%
52	files.covid19treatmentguidelines.nih.gov Internet	17 words — < 1%
53	journals.gaftim.com Internet	16 words — < 1%
54	journals.plos.org Internet	16 words — < 1%
55	www.etasr.com Internet	16 words — < 1%
56	www.iseg.ulisboa.pt Internet	16 words — < 1%
57	www.hippocampus.si Internet	

15 words — < 1%

58 Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 — 14 words — < 1%
Crossref

59 Sajana Maharjan, Faisal Mueen Qamer, Mir Matin, Govinda Joshi, Sanjeev Bhuchar. "Integrating Modelling and Expert Knowledge for Evaluating Current and Future Scenario of Large Cardamom Crop in Eastern Nepal", Agronomy, 2019 — 14 words — < 1%
Crossref

60 nadre.ethernet.edu.et — 13 words — < 1%
Internet

61 Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19", Physiological Genomics, 2020 — 12 words — < 1%
Crossref

62 Guangxue Wan, Xuemei Li, Kedong Yin, Yufeng Zhao, Benshuo Yang. "Application of a novel time-delay grey model based on mixed-frequency data to forecast the energy consumption in China", Energy Reports, 2022 — 12 words — < 1%
Crossref

63 Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 — 12 words — < 1%
Crossref

64 Spiridon Vergis, Konstantinos Bezas, Themis P. Exarchos. "Chapter 3 Decision Support System for Breast Cancer Detection Using Biomarker Indicators", Springer Science and Business Media LLC, 2021
Crossref
12 words — < 1%

65 ejournal.uin-suska.ac.id
Internet
12 words — < 1%

66 Hongyi Duanmu, Thomas Ren, Haifang Li, Neil Mehta, Adam J Singer, Jeffrey M Levsky, Michael L Lipton, Tim Q Duong. "Deep learning of longitudinal chest X-ray and clinical variables predicts duration on ventilator and mortality in COVID-19 patients", Research Square Platform LLC, 2022
Crossref Posted Content
11 words — < 1%

67 Yoga Pristyanto, Anggit Ferdita Nugraha, Irfan Pratama, Akhmad Dahlan. "Ensemble Model Approach For Imbalanced Class Handling on Dataset", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020
Crossref
11 words — < 1%

68 arxiv.org
Internet
11 words — < 1%

69 docplayer.net
Internet
11 words — < 1%

70 www.covprecise.org
Internet
11 words — < 1%

71 Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-
10 words — < 1%

Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021
Crossref

72 Qiao Yang, Jixi Li, Zhijia Zhang, Xiaocheng Wu et al. "A Machine Learning Model to Predict Death Outcome in Severe COVID-19 Patients", Research Square Platform LLC, 2020 — 10 words — < 1%
Crossref Posted Content

73 dokumen.pub — 10 words — < 1%
Internet

74 link.springer.com — 10 words — < 1%
Internet

75 www.semanticscholar.org — 10 words — < 1%
Internet

76 Anita Desiani, Adinda Ayu Lestari, M Al-Ariq, Ali Amran, Yuli Andriani. "Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification", Pattimura International Journal of Mathematics (PIJMath), 2022 — 9 words — < 1%
Crossref

77 Dennis Peace Ezeobi, William Wasswa, Angella Musimenta, Stella Kyoyagala. "A Neonatal Sepsis Prediction Algorithm Using Electronic Medical Record (EMR) Data from Mbarara Regional Referral Hospital (MRRH)", Research Square Platform LLC, 2022 — 9 words — < 1%
Crossref Posted Content

78 Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. ""F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability", — 9 words — < 1%

2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021
Crossref

79	journals.sbmu.ac.ir Internet	9 words — < 1%
80	jurnal.iaii.or.id Internet	9 words — < 1%
81	www.medrxiv.org Internet	9 words — < 1%
82	www.science.gov Internet	9 words — < 1%
83	"Cloud Computing and Services Science", Springer Science and Business Media LLC, 2013 Crossref	8 words — < 1%
84	M. Fatih Aslan, Kadir Sabanci, Ewa Ropelewska. "A CNN-Based Solution for Breast Cancer Detection With Blood Analysis Data: Numeric to Image", 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021 Crossref	8 words — < 1%
85	Rachmadita Andreswari, Dita Aprilia Rahmani, Ria Rahmawati, Achmad Akbar H. "Comparative Analysis of K-Means and K-Nearest Neighbor Algorithm for Telecom Fraud Detection", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022 Crossref	8 words — < 1%
86	Sushil Kumar Mishra, Nitish Kumar Ojha. "A Novel Approach for Face Mask Detection using	8 words — < 1%

Tensorflow for Covid-19", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021
Crossref

87 Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 — 8 words — < 1%
Crossref

88 annalsofrscb.ro — 8 words — < 1%
Internet

89 eprints.umm.ac.id — 8 words — < 1%
Internet

90 etd.aau.edu.et — 8 words — < 1%
Internet

91 ijai.iaescore.com — 8 words — < 1%
Internet

92 ijisae.org — 8 words — < 1%
Internet

93 journalofbigdata.springeropen.com — 8 words — < 1%
Internet

94 plus.mcmaster.ca — 8 words — < 1%
Internet

95 www.researchsquare.com — 8 words — < 1%
Internet

96 Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of — 7 words — < 1%

Drugs Used", International Journal of Advanced Computer Science and Applications, 2021
Crossref

97 Annisa Nurul Puteri, Dewiani, Zulkifli Tahir. "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019 — 6 words — < 1%
Crossref

98 David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016 — 6 words — < 1%
Crossref

99 Li, Dan, Yinghai Ke, Huili Gong, and Xiaojuan Li. "Object-Based Urban Tree Species Classification Using Bi-Temporal WorldView-2 and WorldView-3 Images", Remote Sensing, 2015. — 6 words — < 1%
Crossref

100 Mohammed. "Feature Discretization for Individuality Representation in Twins Handwritten Identification", Journal of Computer Science, 2011 — 6 words — < 1%
Crossref

101 Muhammad Daniyal, Kassim Tawiah, Sara Muhammadullah, Kwaku Opoku-Ameyaw. "Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data", Journal of Healthcare Engineering, 2022 — 6 words — < 1%
Crossref

102 Shilei Lu, Yang Liu, Yuwei Li, Ran Wang. "Multidimensional performance-based evaluation method of high-performance cold source in green building", — 6 words — < 1%

Energy and Buildings, 2021

Crossref

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF

Plagiarism Check _ Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

By Anthony Anggrawan

WORD COUNT	7491	TIME SUBMITTED	03-AUG-2022 02:46PM
		PAPER ID	89134525

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning predict with high accuracy. For that, or that's why this study aims to propose a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both in patient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18][19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20][21][22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [23]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. Machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the treating status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32] [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used in research, including research to classify and predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Hakan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objects and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. At the same time, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 compares some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the prior research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is on Covid-19 patients from a regional hospital in Mataram, Indonesia. The significant data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Figure 1 [43]. Figure 2 shows the process carried out at each stage of CRISP-DM.

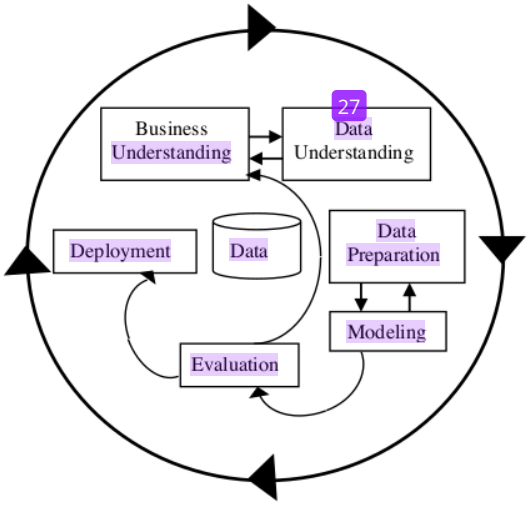

Figure 1. The CRISP-DM Process

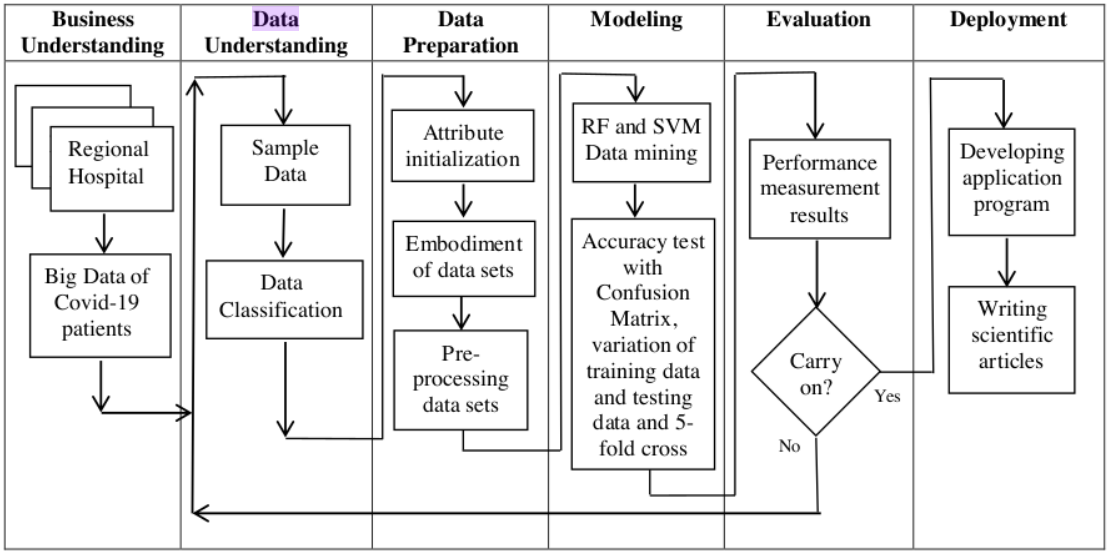

Figure 2. Data Mining Process of Covid-19 Patient Big Data

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria of patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. Data set of the signs and symptoms and the treatment status of Covid-19 patients

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs or symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, known various programming language [44], which has their respective advantages in building application programs [45][46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction (as show in Table 6).

TABLE VI. USE OF HYPERPARAMETER ON SVM AND RF METHOD

Classifier Method	Hyperparameter	Value
SVM		1
	Kernel	Rbf
	Degree	3
	Gamma	Scale
	Coef	0
	Tol	0,000
		1
	Max_iter	-1
Random Forest	n_estimators	100
	Criterion	Gini
	Max_depth	None
	Min_samples_split	2
	Min_samples_leaf	1

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree.

The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:
S = Set of cases
pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 7. The results of the predictions of the SVM and RF methods are shown in Tables 8 and 9.

TABLE VII. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

TABLE VIII. CONFUSION MATRIX OF SVM

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	2	12

TABLE IX. CONFUSION MATRIX OF RF

		Prediction	
	Class	Self-isolation	Inpatient
Actual	Self-isolation	4	0
	Inpatient	1	13

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 10 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE X. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

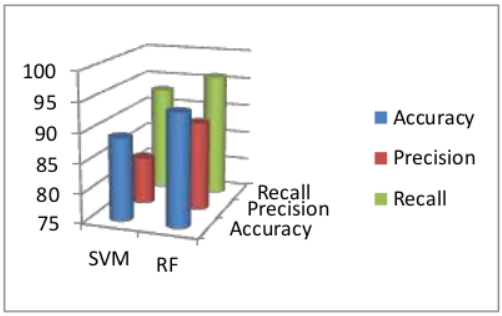

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 11.

TABLE XI. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method in other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 12 and Figure 4.

TABLE XII. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

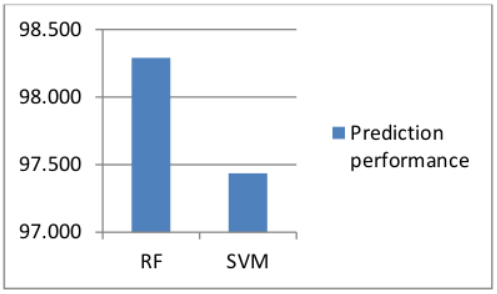

Figure 4. Predictive performance testing using K-fold cross-validation

4.6 *Deployment*

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of

machine learning and the Internet of Things.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

References

[1] C. Rothe *et al*., "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al*., "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al*., "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al*., "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al*., "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.

[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[20] H. Zhang *et al*., "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19): A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[21] X. Jiang *et al*., "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[22] J. Liu *et al*., "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[26] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[29] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[30] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.

[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[34] C. Iwendi *et al*., "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[36] B. Hao *et al*., "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[37] H. Zhao *et al*., "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[39] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship

Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.

[41] V. Demichev *et al*., "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.

[42] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.

[43] P. Chapman *et al*., *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.

[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.

[45] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.

[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Plagiarism Check _ Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

ORIGINALITY REPORT

17%
SIMILARITY INDEX

PRIMARY SOURCES

1	inass.org Internet	82 words — 1%
2	www.ncbi.nlm.nih.gov Internet	56 words — 1%
3	www.mdpi.com Internet	52 words — 1%
4	Daniel Berrar. "Cross-Validation", Elsevier BV, 2019 Crossref	49 words — 1%
5	www.jatit.org Internet	48 words — 1%
6	thesai.org Internet	44 words — 1%
7	Norah Alballa, Isra Al-Turaiki. "Machine learning approaches in COVID-19 diagnosis, mortality, and severity risk prediction: A review", Informatics in Medicine Unlocked, 2021 Crossref	32 words — 1%
8	Soham Guhathakurata, Souvik Kundu, Arpita Chakraborty, Jyoti Sekhar Banerjee. "A novel	26 words — < 1%

approach to predict COVID-19 using support vector machine", Elsevier BV, 2021
Crossref

9 journal.universitasbumigora.ac.id
Internet
26 words — < 1%

10 Dan Song, Ruofeng Tong, Jian Chang, Tongtong Wang, Jiang Du, Min Tang, Jian J. Zhang. "Chapter 7 Clothes Size Prediction from Dressed-Human Silhouettes", Springer Science and Business Media LLC, 2017
Crossref
24 words — < 1%

11 Handoyo Widi Nugroho, Teguh Bharata Adji, Noor Akhmad Setiawan. "Performance Improvement of C4.5 Algorithm using Difference Values Nodes in Decision Tree", 2018 6th International Conference on Cyber and IT Service Management (CITSM), 2018
Crossref
23 words — < 1%

12 jurnal.ugm.ac.id
Internet
22 words — < 1%

13 I Gede Susrama Mas Diyasa, Agus Prayogi, Intan Yuniar Purbasari, Ariyono Setiawan, Sugiarto, Prismahardi Aji Riantoko. "Data Classification of Patient Characteristics Based on Nutritional Treatment Using the K-Nearest Neighbors Algorithm", 2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS), 2021
Crossref
19 words — < 1%

14 Jinhui Xu, Deli Zhang, Lin Feng, Yang Liu, Qingchun Diao. " The efficacy and safety of topical combined with superficial ‐ ray therapy ( ) in treating periungual warts ", Dermatologic Therapy, 2022
Crossref
19 words — < 1%

15	bura.brunel.ac.uk Internet	19 words — < 1%
16	eprints.utm.my Internet	19 words — < 1%
17	I Ayuningtias, I Jaya, M Iqbal. "Identification of yellowfin tuna (Thunnus albacares), mackerel tuna (Euthynnus affinis), and skipjack tuna (Katsuwonus pelamis) using deep learning", IOP Conference Series: Earth and Environmental Science, 2021 Crossref	18 words — < 1%
18	Shoffi Izza Sabilla, Riyanarto Sarno, Kuwat Triyana, Kenshi Hayashi. "Deep learning in a sensor array system based on the distribution of volatile compounds from meat cuts using GC–MS analysis", Sensing and Bio-Sensing Research, 2020 Crossref	18 words — < 1%
19	files.covid19treatmentguidelines.nih.gov Internet	17 words — < 1%
20	www.bmj.com Internet	17 words — < 1%
21	www.researchgate.net Internet	17 words — < 1%
22	journals.gaftim.com Internet	16 words — < 1%
23	www.ijstr.org Internet	16 words — < 1%
24	www.hippocampus.si Internet	15 words — < 1%

25	Mohammad-H. Tayarani N.. "Applications of artificial intelligence in battling against covid-19: A literature review", Chaos, Solitons & Fractals, 2020 Crossref	14 words — < 1%
26	Sajana Maharjan, Faisal Mueen Qamer, Mir Matin, Govinda Joshi, Sanjeev Bhuchar. "Integrating Modelling and Expert Knowledge for Evaluating Current and Future Scenario of Large Cardamom Crop in Eastern Nepal", Agronomy, 2019 Crossref	14 words — < 1%
27	nadre.ethernet.edu.et Internet	13 words — < 1%
28	Ahmad Alimadadi, Sachin Aryal, Ishan Manandhar, Patricia B. Munroe, Bina Joe, Xi Cheng. "Artificial intelligence and machine learning to fight COVID-19", Physiological Genomics, 2020 Crossref	12 words — < 1%
29	Guangxue Wan, Xuemei Li, Kedong Yin, Yufeng Zhao, Benshuo Yang. "Application of a novel time-delay grey model based on mixed-frequency data to forecast the energy consumption in China", Energy Reports, 2022 Crossref	12 words — < 1%
30	Guokang Zhu, Jia Li, Zi Meng, Yi Yu et al. "Learning from Large-Scale Wearable Device Data for Predicting Epidemics Trend of COVID-19", Discrete Dynamics in Nature and Society, 2020 Crossref	12 words — < 1%
31	Spiridon Vergis, Konstantinos Bezas, Themis P. Exarchos. "Chapter 3 Decision Support System	12 words — < 1%

for Breast Cancer Detection Using Biomarker Indicators", Springer Science and Business Media LLC, 2021
Crossref

32 Hongyi Duanmu, Thomas Ren, Haifang Li, Neil Mehta, Adam J Singer, Jeffrey M Levsky, Michael L Lipton, Tim Q Duong. "Deep learning of longitudinal chest X-ray and clinical variables predicts duration on ventilator and mortality in COVID-19 patients", Research Square Platform LLC, 2022
Crossref Posted Content
11 words — < 1%

33 Yoga Pristyanto, Anggit Ferdita Nugraha, Irfan Pratama, Akhmad Dahlan. "Ensemble Model Approach For Imbalanced Class Handling on Dataset", 2020 3rd International Conference on Information and Communications Technology (ICOIACT), 2020
Crossref
11 words — < 1%

34 arxiv.org
Internet
11 words — < 1%

35 docplayer.net
Internet
11 words — < 1%

36 www.covprecise.org
Internet
11 words — < 1%

37 Putu Desiana Wulaning Ayu, Sri Hartati, Aina Musdholifah, Detty S. Nurdiati. "Chapter 20 Amniotic Fluids Classification Using Combination of Rules-Based and Random Forest Algorithm", Springer Science and Business Media LLC, 2021
Crossref
10 words — < 1%

38 Qiao Yang, Jixi Li, Zhijia Zhang, Xiaocheng Wu et al. "A machine learning model to predict death
10 words — < 1%

outcome in severe COVID-19 patients", Research Square Platform LLC, 2020
Crossref Posted Content

39 dokumen.pub
Internet
10 words — < 1%

40 link.springer.com
Internet
10 words — < 1%

41 www.semanticscholar.org
Internet
10 words — < 1%

42 Anita Desiani, Adinda Ayu Lestari, M Al-Ariq, Ali Amran, Yuli Andriani. "Comparison of Support Vector Machine and K-Nearest Neighbors in Breast Cancer Classification", Pattimura International Journal of Mathematics (PIJMath), 2022
Crossref
9 words — < 1%

43 Dennis Peace Ezeobi, William Wasswa, Angella Musimenta, Stella Kyoyagala. "A Neonatal Sepsis Prediction Algorithm Using Electronic Medical Record (EMR) Data from Mbarara Regional Referral Hospital (MRRH)", Research Square Platform LLC, 2022
Crossref Posted Content
9 words — < 1%

44 Prismahardi Aji Riyantoko, Sugiarto, I Gede Susrama Mas Diyasa, Kraugusteeliana. ""F.Q.A.M" Feyn-QLattice Automation Modelling: Python Module of Machine Learning for Data Classification in Water Potability", 2021 International Conference on Informatics, Multimedia, Cyber and Information System (ICIMCIS, 2021
Crossref
9 words — < 1%

45 journals.sbmu.ac.ir
Internet
9 words — < 1%

46	jurnal.iaii.or.id Internet	9 words — < 1%
47	www.medrxiv.org Internet	9 words — < 1%
48	"Cloud Computing and Services Science", Springer Science and Business Media LLC, 2013 Crossref	8 words — < 1%
49	M. Fatih Aslan, Kadir Sabanci, Ewa Ropelewska. "A CNN-Based Solution for Breast Cancer Detection With Blood Analysis Data: Numeric to Image", 2021 29th Signal Processing and Communications Applications Conference (SIU), 2021 Crossref	8 words — < 1%
50	Maayan Frid-Adar, Rula Amer, Ophir Gozes, Jannette Nassar, Hayit Greenspan. "COVID-19 in CXR: From Detection and Severity Scoring to Patient Disease Monitoring", IEEE Journal of Biomedical and Health Informatics, 2021 Crossref	8 words — < 1%
51	Rachmadita Andreswari, Dita Aprilia Rahmani, Ria Rahmawati, Achmad Akbar H. "Comparative Analysis of K-Means and K-Nearest Neighbor Algorithm for Telecom Fraud Detection", 2022 2nd International Conference on Information Technology and Education (ICIT&E), 2022 Crossref	8 words — < 1%
52	Sushil Kumar Mishra, Nitish Kumar Ojha. "A Novel Approach for Face Mask Detection using Tensorflow for Covid-19", 2021 Third International Conference on Inventive Research in Computing Applications (ICIRCA), 2021 Crossref	8 words — < 1%

53	Weiqiu Jin, Shuqing Dong, Chengqing Yu, Qingquan Luo. "A data-driven hybrid ensemble AI model for COVID-19 infection forecast using multiple neural networks and reinforced learning", Computers in Biology and Medicine, 2022 Crossref	8 words — < 1%
54	annalsofrscb.ro Internet	8 words — < 1%
55	dfzljdn9uc3pi.cloudfront.net Internet	8 words — < 1%
56	eprints.umm.ac.id Internet	8 words — < 1%
57	etd.aau.edu.et Internet	8 words — < 1%
58	journalofbigdata.springeropen.com Internet	8 words — < 1%
59	openaccess.altinbas.edu.tr Internet	8 words — < 1%
60	plus.mcmaster.ca Internet	8 words — < 1%
61	www.researchsquare.com Internet	8 words — < 1%
62	Anthony Anggrawan, Christofer Satria, Che Ku Nuraini, Lusiana -, Ni Gusti Ayu Dasriani, Mayadi -. "Machine Learning for Diagnosing Drug Users and Types of Drugs Used", International Journal of Advanced Computer Science and Applications, 2021 Crossref	7 words — < 1%

63 Annisa Nurul Puteri, Dewiani, Zulkifli Tahir. "Comparison of Potential Telemarketing Customers Predictions with a Data Mining Approach using the MLPNN and RBFNN Methods", 2019 International Conference on Information and Communications Technology (ICOIACT), 2019
Crossref
6 words — < 1%

64 David J. P. Williams. "The great ESCAPE - a clinical pharmacologist's journey in stroke research", British Journal of Clinical Pharmacology, 2016
Crossref
6 words — < 1%

65 Mohammed. "Feature Discretization for Individuality Representation in Twins Handwritten Identification", Journal of Computer Science, 2011
Crossref
6 words — < 1%

66 Muhammad Daniyal, Kassim Tawiah, Sara Muhammadullah, Kwaku Opoku-Ameyaw. "Comparison of Conventional Modeling Techniques with the Neural Network Autoregressive Model (NNAR): Application to COVID-19 Data", Journal of Healthcare Engineering, 2022
Crossref
6 words — < 1%

67 Shilei Lu, Yang Liu, Yuwei Li, Ran Wang. "Multidimensional performance-based evaluation method of high-performance cold source in green building", Energy and Buildings, 2021
Crossref
6 words — < 1%

68 Suresh Vatakuti, Jeroen L. A. Pennings, Emilia Gore, Peter Olinga, Geny M. M. Groothuis. "Classification of Cholestatic and Necrotic Hepatotoxicants Using Transcriptomics on Human Precision-Cut Liver Slices", Chemical Research in Toxicology, 2016
Crossref
6 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	ON	EXCLUDE MATCHES	OFF

Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients

Anthony Anggrawan, Mayadi, Christofer Satria, Bambang Krismono Triwijoyo and Ria Rismayati
Universitas Bumigora, Mataram, Indonesia
Email: {anthony.anggrawan, mayadi.yadot, chris, bkrismono, riris}@universitasbumigora.ac.id

***Abstract*—Covid-19 has become a global pandemic that causes many deaths, so medical treatment for Covid-19 patients gets special attention, whether hospitalized or self-isolated. However, the problem in medical action is not easy, and the most frequent mistakes are due to inaccuracies in medical decision-making. Meanwhile, machine learning can predict with high accuracy. For that, or that's why this study aims to propose a data mining classification method as a machine learning model to predict the treatment status of Covid-19 patients accurately, whether hospitalized or self-isolated. The data mining method used in this research is the Random Forest (RF) and Support Vector Machine (SVM) algorithm with Confusion Matrix and k-fold Cross Validation testing. The finding indicated that the machine learning model has an accuracy of up to 94% with the RF algorithm and up to 92% with the SVM algorithm in predicting the Covid-19 patient's treatment status. It means that the machine learning model using the RF algorithm has more accurate accuracy than the SVM algorithm in predicting or recommending the treatment status of Covid-19 patients. The implication is that RF machine learning can help/replace the role of medical experts in predicting the patient's care status.**

***Index Terms*—Data Mining, Random Forest, Support Vector Machine, Prediction, Covid-19, Machine Learning**

I. Introduction

The Covid-19 disease is currently a world pandemic [1][2][3]. It is the cause of the global health crisis [4][5][6], which is not only because of its high-speed transmission [5][7], but more than 100 million people have died infected worldwide, and more than two million people have died from it [5]. Covid-19 is a highly contagious viral disease that requires special care and follow-up predictive analytics for better treatment of the disease [8]. However, the Covid-19 pandemic poses a significant challenge to providing health care and services for patients [7]. So it is not surprising that researchers use many research methods to control the Covid-19 pandemic, including the research methods that have received the most attention: prediction, statistical, and epidemiological [6]. Generally, medical actions taken for Covid-19 patients are isolated [9], namely hospitalization or self-isolation. However, these hospitalized Covid-19 patients are receiving intensive medical care from doctors.

In essence, the care status for Covid-19 patients is self-isolation for patients with non-severe illness status and hospitalization for patients who are seriously ill and at critical risk or cause death. Hospitals or medical doctors take various ways to reduce the number of deaths of Covid-19 patients, including by regulating the status of services in hospitals. Expert doctors recommend that patients self-isolate or should be hospitalized after analyzing the patient's medical data. Determining the level of care for Covid-19 patients is a form of medical treatment or treatment for Covid-19 patients to get proper treatment or care.

Errors in decision-making often occur because decision-makers consider several criteria as the basis for decision-making [10]. So it is not surprising that previous researchers emphasized that the errors most often occur due to inaccuracies in decision making [11], and decision making is a difficult task because of the impact of the decisions made [11]. Likewise, in recommending whether a Covid-19 patient should be hospitalized or self-isolated, several criteria from the disease symptoms and the results of medical tests are the basis for considering whether a patient should be hospitalized or self-isolated. In essence, it is difficult to accurately determine the treatment status of Covid-19 patients, both inpatient and self-isolation.

Meanwhile, Machine Learning is a rapidly growing part of computer science today [12]. Although in most scientific studies, machine learning is popular, it is still very limited in health studies [13]. Machine learning helps mining data to predict mining results accurately [14]. Machine learning is a helpful technique for finding correlations based on cases to predict [15]. With the availability of big data, it is possible to develop various solutions using machine learning [16][17]; moreover, with advances in information and communication technology [18][19], it is straightforward to collect the required big data. Among the solutions using machine learning, one of which is predictive modeling [20][21][22]. Furthermore, machine learning can uncover hidden patterns in big data, distinguish patterns better and more accurately [13], and provide high-accuracy prediction results [23]. For this reason (or

Manuscript received xxx xx, 20xx; revised xxxx x, 20xx.

why), this study's objective is to propose a machine learning system model for decision-making solutions (predictions) for the treatment status of Covid-19 patients, both inpatient and self-isolation, using data mining methods.

The implication is that the proposed machine learning system model can help and even replace the role of medical experts (specialist doctors) in making medical decisions for Covid-19 patients, whether hospitalization or self-isolation. Machine learning performs tasks like a medical specialist in deciding the results of the diagnosis of the nursing status of Covid-19 patients based on medical data of Covid-19 patients. Furthermore, machine learning can work tirelessly, time and place, and has intelligence like an expert, so it is not surprising that previous research confirms that intelligent machines can make superior decisions to experts because humans have a human error factor [24].

Machine learning can predict classification to predict class membership and regression to show numerical values [12]. While data mining is part of machine learning that can make system models have artificial intelligence. Artificial intelligence is a breakthrough in today's technology that has been widely used in prediction [25]. The embodiment of artificial intelligence in machine learning with data mining methods is an iterative process of training and repeated testing of data sets (big data) on the system model. In short, machine learning has an artificial intelligence role in predicting new data with high accuracy [23]. After all, predicting individuals with symptoms of being infected with Covid-19 mandates machine learning (application-based) and contributes to effectively isolating Covid-19 patients [26].

Big data demands large storage media [27]. However, big data is no longer traditionally processed [28]. Instead, today's big data processing relies on machines that can provide systematic results [29]. Big data storage is generally on a computer server with a large storage capacity. Still, some also make it happen by renting online cloud data storage services such as Amazon Simple Storage Service (Amazon S3) and Google Cloud [30]. Cloud facilitates cost-effective big data storage and analysis [30].

Big data processing techniques in data mining include several stages: target data, preprocessed data, data mining, and evaluation/analysis of mining results [30]. Target data and preprocessed data are the processes of extracting raw data from big data [30]. Target data is to select the required data (sample data) and classify data. The preprocessed data is to prepare data sets for data mining, including cleaning up incomplete data, duplicate data, and converting string data into numeric coded data. Finally, data mining and evaluation results extract hidden information by applying data mining methods suitable for the objectives and analyzing them.

Many Data Mining methods include K-Means, Naïve Bayes, KNN (k-Nearest Neighbor), ID3 (Iterative Dichotomiser 3), C4.5, Cart, RF, SVM, and others. There are two types or methods of machine learning, namely supervised machine learning and unsupervised machine learning. It is referred to as a supervised learning method when the subject's membership is known, and training is carried out to classify new data into its category. On the other hand, it is referred to as an unsupervised learning method when the subject's membership is unknown, and the closest distance search is to categorize the groups. The Data Mining methods used in this research are RF and SVM algorithms. RF and SVM are prevalent machine learning algorithms used in various scientific studies [31] and constitute data classification techniques with supervised learning methods. The RF machine learning algorithm has been widely applied for classification [32] [33], as well as SVM algorithm has a widely known technique used for classification [33]. It is why this study uses SVM and RF to classify treatment status. Given that the SVM and RF machine learning algorithms are both popularly used by many researchers, the SVM and RF machine learning algorithms are the most appropriate combination used in research, including research to classify and predict the treatment status of Covid-19 patients.

However, it is essential to know the accuracy of predicting the care status of Covid-19 patients from the system model proposed in this study and whether the patient should be hospitalized or self-isolated. Therefore, this study also further tested the percentage of machine learning efficacy or accuracy in predicting the treatment status of Covid-19 patients. The accuracy of predicting the treatment status of Covid-19 patients is tested on both RF and SVM machine learning methods.

The organization of the following writing of this manuscript is as follows: The second subsection discusses several of the related works of previous researchers and their relevance to the work in this research article. The third subsection describes Research Methodology, which discusses methods used in research in recommending patient care status. Meanwhile, the fourth subsection discusses the results of the study. Finally, it ends with a subsection that concludes the study's findings, the novelty of the research results, and advice for further research.

II. Related Works

This subsection provides an overview of some related works from the latest scientific articles compared with the work in this research article.

Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) performed the classification of satellite remote sensing images using machine learning algorithms with RF, SVM, and decision tree classifier (DT) [31]. This previous research is different from the research in the article on the research objectives and the object under study. In the meantime, Celestine Iwendi et al. (2020) proposed the Random Forest model to predict the disease severity of Covid-19 patients [34]. The

difference between previous research and the research in this article is that the previous research only used one method, namely Random Forest. In contrast, the research in this article used two methods, namely Random Forest and SVM. The difference also lies in the prediction criteria and class; previous research predicts the severity of the illness of Covid-19 patients, while the research in this article predicts the treatment status of Covid-19 patients.

Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) proposed a machine learning model to classify breast cancer by applying RF and SVM methods [35]. Previous research and the article in this research are both using RF and SVM methods. However, the previous research has research objectives that are not the same as the research in this article. The prior study classified breast cancer as patients with breast cancer. In contrast, the research in this article predicts the treatment status of Covid-19 patients.

Based on patient clinical data, using statistical methods, Boran Hao et al. (2020) developed a model to predict pneumonia severity in Covid-19 patients using the Natural Language Processing tool [36]. However, this previous research differs in the research purpose and way compared to the research in this article. Meanwhile, Anthony Anggrawan et al. (2021) implemented machine learning to diagnose drug users and types of drug-using Forward Chaining and Certainty Factor methods [23]. Meanwhile, the research in this article develops machine learning to predict the patient's treatment status, whether inpatient or self-isolation, based on symptoms or patient medical data using RF and SVM.

Hongwei Zhao et al. (2021) built a model to predict the number of cases of Covid-19 patients in the future using the Poisson and Gamma distribution [37]. Similarities between articles in this study and the previous one proposed a model with a machine learning approach. However, this previous research differs in the research purpose and method compared to the research in this article. In the meantime, Bassam Mahboub et al. (2021) developed a model to predict the length of hospital stay with the decision tree (DT) method [8]. This article's research differs from previous research; the difference lies in the research objectives and techniques used. If prior research predicts the length of stay for Covid-19 patients, the research in this article predicts whether Covid-19 patients should be hospitalized or self-isolated. This article's research does not use the DT method but uses the RF and SVM methods.

Soham Guhathakurata et al. (2021) predicted whether a person is infected with Covid-19 or not using SVM [38]. However, this previous study differed in its objectives from this article's research. The previous research predicts patients suffering from Covid-19 or not utilizing the SVM data mining method. In contrast, the research in this article indicates the patient's care status using RF and SVM data mining methods. At the same time, Ankit Mehrotraa and Reeti Agarwal (2021) reviewed the usefulness of the Data Mining method for the Covid-19 pandemic [39]. This previous research is a literature review study that concludes that the Data Mining method plays an essential role in health care, diagnosing diseases, and recommending cures. However, it is different from this article's research because it is an experimental study, not a literature review.

Pratiyush Guleria et al. (2022) proposed a machine learning model to predict the death rate of Covid-19 patients [14]. However, previous research has different objectives and data mining methods compared to this article's research. The difference is that previous studies examined the infection rate of Covid-19 patients to predict the cure/death rate of Covid-19 patients using the SVM, Decision Trees, and Naïve Bayes data mining methods. In contrast, the research in this article predicts the care status of Covid-19 patients using RF and SVM data mining methods [14].

Anthony Anggrawan, Mayadi, Christofer Satria, and Lalu Ganda Rady Putra (2022) developed a machine learning model for scholarship recipients' recommendations by using Analytical Hierarchy Process (AHP) and the Multi-Objective Optimization Method by Ratio Analysis (Moora) methods [40]. However, the previous research differs in the purpose and way compared to this article's research. In contrast, Vadim Demichev et al. (2022) offered a model to optimize the treatment or intensive care of seriously ill Covid-19 patients with plasma proteomics [41]. This previous research is different from the research in the article on the research objectives, research method, and the object under study.

Table 1 compares some of the most recent previous related work with the work carried out in this study. By referring to the elaboration of the most recent last related work by some researchers, the research carried out in this article has novelties (from the prior research gap) that previous researchers have not studied. In essence, the gap in earlier research is that no one has researched machine learning models to predict the inpatient status or self-isolation of Covid-19 patients by involving RF and SVM algorithms. In addition, the 12 criteria used to indicate the treatment status of Covid-19 patients are entirely different from previous similar studies (as shown in Table 1 in the Criteria/Attributes column). So, the study's originality lies in proposing a machine learning model to predict the nursing status of Covid-19 patients, whether inpatient or self-isolation, which previous researchers have never done. Besides that, the novelty is also in the method used, not just one data mining method in predicting the treatment status of Covid-19 patients, but using two data mining methods. So this study can show differences in the accuracy of the RF and SVM methods in predicting the treatment status of Covid-19 patients.

TABLE I. COMPARISON OF THIS ARTICLE'S WORK WITH SOME PREVIOUS RELATED WORKS

Research by	Research methods			Criteria/Attributes		Research Object	Accuracy Test
	RF	SVM	ML	Number	Name		
Askin Kavzoglu, Furkan Bilucan, and Alihan Teke (2020) [31]	Yes	Yes	Yes	10	Coastal Aerosol, Blue, Green, Red, Vegetation Red Edge, NIR, Narrow NIR, Water vapor, SWIR-Cirrus, SWIR	Satellite remote sensing images	Yes
Celestine Iwendi et al. (2020) [34]	Yes	No	Yes	6	Symptom1. symptom2, symptom3, symptom4, symptom5, symptom6	Illness severity of Covid-19 patients	No
Chelvian Aroef, Yuda Rivan, and Zuherman Rustam (2020) [35]	Yes	Yes	Yes	9	Age, Body Mass Index (BMI), Glucose, Insulin, Homa, Leptin, Adiponectin, Resistin, MCP 1	Breast cancer	Yes
Boran Hao et al. (2020) [36]	No	No	No	10	Radiology Opacities, Respiratory Rate, Age, Fever Male, Albumin, Anion Gap, SpO2, LDH, Calcium	The severity of pneumonia in Covid-19 patients	Yes
Anthony Anggrawan et al. (2021) [23]	No	No	Yes	27	Out of breath, Anxious, Nausea, Diarrhea, Convulsions, Easily angry, Depression, Sleep patterns change, Sweating, Chills, Shaking, Insomnia, Fast heart rate, Blood pressure rises, Difficult to focus, Difficult to rest, Weight loss, Dry mouth, Blurred vision, Changed skin color, Constipation, Stomachache, Drowsiness, Itching, Difficulty urinating, Mood swings, Dizziness	Drug users and types of drug-using	Yes
Hongwei Zhao et al. (2021) [37]	No	No	No	0	-	Number of cases of Covid-19 patients	No
Bassam Mahboub et al. (2021) [8]	No	No	Yes	5	Age, Gender, Nationality, Blood group, BMI	The treatment period for Covid-19 patient	Yes
Soham Guhathakurata et al. (2021) [38]	No	Yes	Yes	8	Temp, Breathing rate, Hypertension, Heartbeat rate (HBR), Acute respiratory disease syndrome (ARDS), Chest pain, Heart disease, Cough with sputum (CWS)	Predicting whether patients are infected with Covid-19 or not	No
Ankit Mehrotraa and Reeti Agarwal (2021) [39]	No	No	No	0	-	Discussing the data mining method's role in the Covid-19 pandemic	No
Pratiyush Guleria et al. (2022) [14]	No	Yes	Yes	0	-	The death rate of Covid-19 patient	Yes
Anthony Anggrawan, et al. (2022) [40]	No	No	Yes	6	Achievement index, achievement points, recommendation, organizational activity, semester level, and completeness of documents	Scholarship recipient	Yes
Vadim Demichev et al. (2022) [41]	No	No	No	0	-	Optimization of treatment for Covid-19 patients	Yes
Our research	Yes	Yes	Yes	12	Pneumonia, ARDS, CHF, AKI, CAD, Dyspnea, NSTEMI, ADHF, HHD, Febris, Anosmia, Ageusia	Care status of Covid-19 patients, whether inpatient or self-isolation	Yes

Note: ML = Machine Learning

III. METHODOLOGY

This study applies two data mining methods or machine learning algorithms: RF and SVM. The big data is on Covid-19 patients from a regional hospital in Mataram, Indonesia. The significant data source used in this study is primary data from patient medical records/documents. The attributes of the patient's disease symptoms and care status classes amount to thousands of patient medical record data. Patient datasets containing non-Covid-19 and duplicate and incomplete Covid-19 patient data are removed, so only data is left as a dataset for data mining processes. Medical record data of disease symptoms obtained from string data is then converted into numeric data. The development of the application program in this study uses the Python computer programming language.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods. The data mining process in this research uses CRISP-DM (Cross-Industry Standard Process for Data Mining). CRISP-DM is a standard data mining process. The process in CRISP-DM comprises a six-stage [42], as shown in Figure 1 [43]. Figure 2 shows the process carried out at each stage of CRISP-DM.

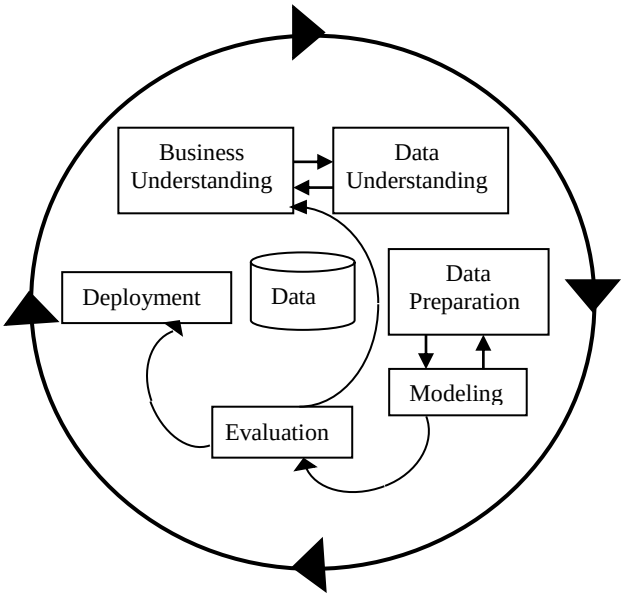

Figure 1. The CRISP-DM Process

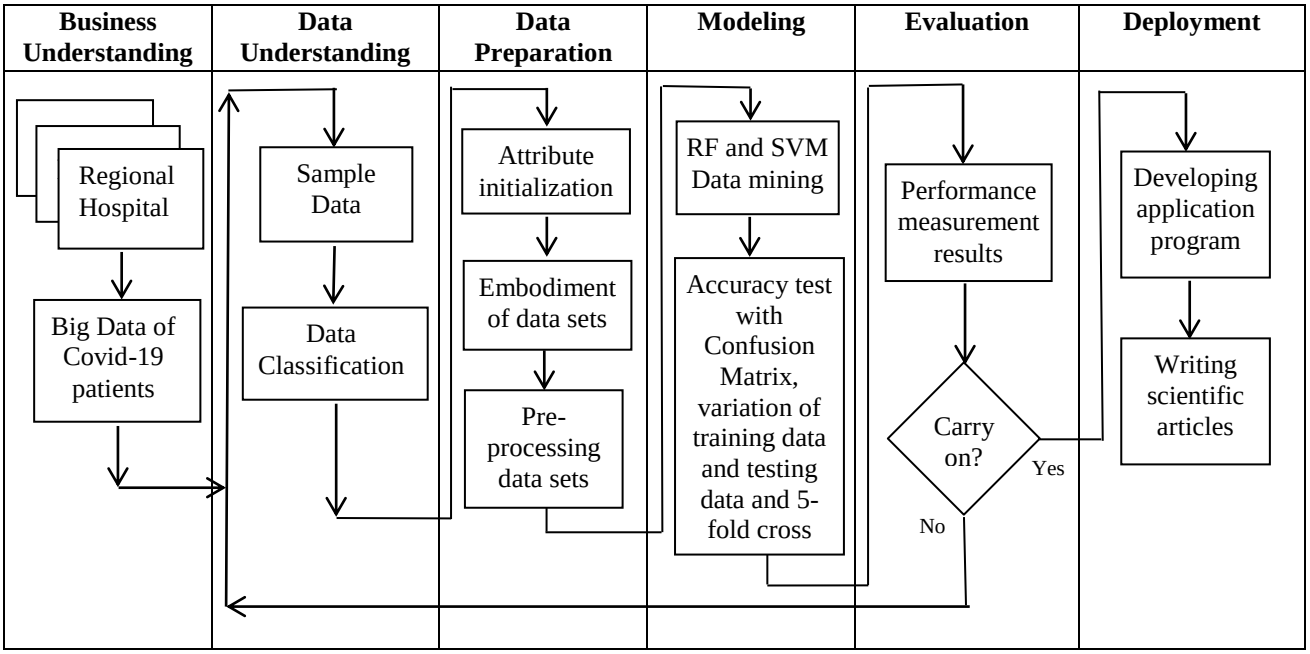

Figure 2. Data Mining Process of Covid-19 Patient Big Data

In Figure 2, business understanding is the stage of sorting out thousands of hospital patient medical data to collect the required patient data. The next stage is understanding the data collected as representative data for Covid-19 patients. This Covid-19 patient data classifies the patient's signs and symptoms and treatment status, which needs further processing at the next stage. The next stage is the data preparation stage, which essentially determines the attributes of the names of signs and symptoms of Covid-19 patients. The embodiment of the dataset containing knowledge according to treatment status refers to the signs and symptoms that the patient has (marked with Yes) or does not have (marked with No). The next thing to do is preprocess the dataset, changing the category value of the symptom attribute and the class attribute with the number 1 and the number -1. The process of extracting raw data obtained in the previous process stage is data that is further processed by data mining methods or used as learning machine learning data at the modeling stage. So that machine learning can predict. The next stage is the evaluation stage, namely, knowing the predictive reliability of the data mining or machine learning method. Then the last stage is the deployment stage to disseminate research results so that they are helpful for implementation by various parties, especially hospitals and other professionals, in the form of developing application programs and scientific articles.

This research uses a confusion matrix and k-fold cross-validation to measure the classification performance of RF and SVM methods.

IV. Result and Discussion

4.1 Business Understanding

The significant data acquisition of Covid-19 patients needed for research is obtained from the hospital. The data collected from medical document data from all Covid-19 patients registered at the hospital includes the patient's name, disease symptoms, and the treatment status specified. There are thousands of data on Covid-19 patients. The patient dataset containing incomplete data and non-covid-19 was omitted or ignored. The critical information extracted at this stage is first to find the attributes or criteria of the class of treatment status (hospitalization or self-isolation); second, to find the category of each feature of the treatment status class. The existing attributes and categories represent disease symptoms and other medical data from Covid-19 patients. Based on Covid-19 patient data adopted from the hospital, there are 12 symptom criteria or patient medical data that are used as references by expert doctors in determining the status of patient care, whether to be hospitalized or self-isolated. Furthermore, big data containing several symptom criteria or patient medical data is used in training and testing the prediction model proposed in this study. Therefore the offered machine learning model has artificial intelligence in predicting.

4.2 Data Understanding

The Data Understanding stage is preparing the data set from the research. The dataset from this study is a data representation of the Covid-19 patient sample, which contains sign and symptom data and treatment status. Table 2 shows the association between signs and symptoms of disease and treatment status in the study data set.

TABLE II. DATA SET OF THE SIGNS AND SYMPTOMS AND THE TREATMENT STATUS OF COVID-19 PATIENTS

No	Disease Sign and Symptom	Treatment
1	Pneumonia, Dyspnea	Inpatient
2	Pneumonia, ARDS, AKI, Febris	Inpatient
3	Pneumonia, CHF, CAD, Dyspnea	Inpatient
4	Pneumonia, AKI	Inpatient
5	Pneumonia, CAD	Inpatient
6	Pneumonia, Dyspnea, Anosmia, Ageusia	Inpatient
7	CHF, NSTEMI	Inpatient
..	…..	….
..	…..	….
114	Febris, Anosmia, Ageusia	Self-isolation
115	Pneumonia, Anosmia	Inpatient
116	Pneumonia, Anosmia, Ageusia	Inpatient
117	HHD	Inpatient

4.3 Data Preparation

Each patient confirmed positive for Covid-19 has a different diagnosis from others, and some patients have similar diagnoses. There were 117 patients with Covid-19 who had a different diagnosis from the others. In this study, the number of signs and symptoms or the number of research criteria is 12 signs and symptoms, or the number of research criteria is 12 signs and symptoms or 12 criteria (see Table 3).

The signs and symptoms of each Covid-19 patient (G01, G02, ... G12 or Gi where i = 1, 2,3 ...12) are not all the same from one patient to another. For this reason, the attributes of each patient's data are different, and some are the same between one patient and another, as shown in Table 4. If the sign or symptom attribute is No, the patient does not have these signs or symptoms. On the other hand, if the sign or symptom attribute is Yes, the patient has these signs or symptoms.

Furthermore, the preprocessing of the data set is done by changing the Gi with xi and the Gi Yes attribute value with the number 1 while the Gi No attribute with the number -1. In addition, dataset preprocessing is also carried out on class attributes, namely changing the independent isolation class attribute category with the number 1 and the inpatient class attribute category with the number -1, as shown in Table 5.

TABLE III. DATA SET RELATED TO RESEARCH ATTRIBUTES AND DISEASE SIGNS AND SYMPTOMS

Attribute	Sign and Symptoms	Word extension
G01	Pneumonia	Pneumonia
G02	ARDS	Acute Respiratory Distress Syndrome
G03	CHF	Congestive Heart Failure
G04	AKI	Acute Kidney Injury
G05	CAD	Coronary Artery Disease
G06	Dyspnea	Dyspnea
G07	NSTEMI	Non-ST-Segment Elevation Myocardial Infarction
G08	ADHF	Acute Decompensated Heart Failure
G09	HHD	Hypertensive Heart Disease
G10	Febris	Febris
G11	Anosmia	Anosmia
G12	Ageusia	Ageusia

TABLE IV. DATA SET OF KNOWLEDGE BASED ON TREATMENT STATUS REFERRING TO THE SIGNS AND SYMPTOMS

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	Yes	No	No	No	…	…	No	No	Inpatient
2	Yes	Yes	No	Yes	…	…	No	No	Inpatient
3	Yes	No	Yes	No	…	…	No	No	Inpatient
4	Yes	No	No	No	…	…	No	No	Inpatient
5	Yes	No	No	No	…	…	No	No	Inpatient
6	Yes	No	No	No	…	…	Yes	Yes	Inpatient
7	No	No	Yes	No	…	…	No	No	Inpatient
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	No	No	No	No	…	…	No	Yes	Self-isolation
115	Yes	No	No	No	…	…	Yes	No	Inpatient
116	Yes	No	No	No	…	…	Yes	Yes	Inpatient
117	No	No	No	No	…	…	No	No	Inpatient

TABLE V. PREPROCESSING OF THE DATA SET RESULT

No	G01	G02	G03	G04	…	…	G11	G12	Class
1	1	-1	-1	-1	…	…	-1	-1	-1
2	1	1	-1	1	…	…	-1	-1	-1
3	1	-1	1	-1	…	…	-1	-1	-1
4	1	-1	-1	-1	…	…	-1	-1	-1
5	1	-1	-1	-1	…	…	-1	-1	-1
6	1	-1	-1	-1	…	…	1	1	-1
7	-1	-1	1	-1	…	…	-1	-1	-1
..	…	…	…	…	…	…	…	…	…
..	…	…	…	…	…	…	…	…	…
114	-1	-1	-1	-1	…	…	-1	1	1
115	1	-1	-1	-1	…	…	1	-1	-1
116	1	-1	-1	-1	…	…	1	1	-1
117	-1	-1	-1	-1	…	…	-1	-1	-1

4.4 Modeling

The proposed machine learning model to predict Covid-19 treatment status in this study applies the RF and SVM data mining classification methods. In addition, known various programming language [44], which has their respective advantages in building application programs [45][46]. The application program built in this research uses the Python programming language to facilitate patient care status prediction (as show in Table 6).

TABLE VI. USE OF HYPERPARAMETER ON SVM AND RF METHOD

Classifier Method	Hyperparameter	Value
SVM	C	1
	Kernel	Rbf
	Degree	3
	Gamma	Scale
	Coef	0
	Tol	0,000 1
	Max_iter	-1
RF	n_estimators	100
	Criterion	Gini
	Max_depth	None
	Min_samples_split	2
	Min_samples_leaf	1

4.4.1 SVM data mining method

The process of realizing the classification using the SVM data mining method is as follows: (a) Forming a linear equation from the training data that has gone through the preprocessing stage; (b) Finding the values of w and b by means of elimination and substitution of linear equations; (c) Finding the value of the classification decision with the function.

In SVM, there are two implementation models: mathematical programming techniques and kernel functions. This study applies kernel functions and focuses on classifying two categories of class attributes. The class attribute is a treatment for yi = +1, −1. The formula of the SVM data mining method is: (a) to form a linear equation from the training data; (b) find the value of w and b, and (c) the value of the classification decision is as follows.

$$S = ((x1, y1), \dots, (xl, yl)) \quad (1)$$

$$yi\ ((w.\ xi) + b) \geq 1, i = 1, \dots, \quad (2)$$

$$(x) = w.\ x + b \quad (3)$$

Description:

S= set; x = attribute; y = class; w = weight; b = bias

4.4.2 RF data mining method

The process of realizing the classification using the RF data mining method is as follows: (a) Generating a random subset of data; (b) Creating a decision tree (Root tree, branch tree & leaves tree) from each attribute and class; and (c) Testing each decision tree with data testing and calculating the accuracy of each decision tree.

RF uses bootstrap samples from training data to create a tree from a randomly selected subset. The chosen predictor is a candidate for splitting the decision tree.

The results of the category predictions from the treatment class based on the results of the highest voting were chosen as the final prediction results. The formula for the RF data mining method is the Gini criterion and the Entropy criterion:

$$Gini = 1 - \sum_{i=1}^{c} pi^2 \quad (4)$$

$$Entropy(S) = \sum_{i=1}^{c} -pi \text{ x } log_2(pi) \quad (5)$$

Description:
S = Set of cases
pi = the proportion of case i to the Set of cases

4.4.3 Confusion matrix

This research uses a confusion matrix to measure the performance of the classification method. The confusion matrix is a method that can be used to measure the performance of a classification method. In essence, the confusion matrix can produce information by comparing the system's classification results with the classification results that should be.

In measuring performance using the Confusion Matrix, four terms represent the results of the classification process, namely: True Positive (TP) or positive data detected correctly; False Positive (FP) or negative data detected is positive; True Negative (TN) or negative data detected correctly, and False Negative (FN) or positive data detected is negative. Meanwhile, the calculation of accuracy, prediction, and Recall in the confusion matrix can use the following equation:

$$accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (6)$$

$$Precision = \frac{TP}{TP+FP} \quad (7)$$

$$Recall = \frac{TP}{TP+FN} \quad (8)$$

Accuracy states the closeness of the measurement results to the actual value, while Precision shows how close the difference in the measurement results is on repeated measurements. On the other hand, Recall states the level of success in retrieving information. Precision and Recall are necessary because Precision denotes a measure of quality, and Recall denotes a measure of quantity.

Measurement of accuracy is based on the ratio between the correct predictions (positive and negative) with the overall data. In contrast, precision measurements are based on the percentage of true positive predictions compared to overall positive predicted outcomes. Meanwhile, the recall measurement is based on the ratio of true positive predictions compared to the general actual positive data.

The format of the confusion matrix table is as shown in Table 7. The results of the predictions of the SVM and RF methods are shown in Tables 8 and 9.

TABLE VII. CONFUSION MATRIX

Class	Classified Positive	Classified Negative
Positive	True Positive	False Negative
Negative	False Negative	True Positive

TABLE VIII. CONFUSION MATRIX OF SVM

		Prediction	
	Class	**Self-isolation**	**Inpatient**
Actual	**Self-isolation**	4	0
	Inpatient	2	12

TABLE IX. CONFUSION MATRIX OF RF

		Prediction	
	Class	**Self-isolation**	**Inpatient**
Actual	**Self-isolation**	4	0
	Inpatient	1	13

4.4.4 K-fold Cross-Validation

This study used K-fold cross-validation to measure the performance of the classification method. K-fold cross-validation helps assess the performance of data mining methods by dividing the data sample randomly and grouping the data as much as the k-fold value. In the performance testing of this study with k-fold cross-validation, the dataset is partitioned into five subsets (k = 5). It allows each subgroup to have the same number and fold, which refers to the number of resulting subsets. Dataset partitioning is done by taking random samples from the dataset. However, data that has been taken previously will not be retrieved.

In the first fold, the first subset serves as the validate set (Dval), and the remaining four subsets serve as the training set (Dtrain). In the second fold, the second subset is the validate set, the remaining subset is the training set, and so on until the 5th fold.

4.5 Evaluation

The evaluation of the proposed model in this study is to measure the performance of the resulting prediction system model. The model's performance evaluation is based on the prediction system model generated by the RF and SVM methods.

4.5.1 Evaluation of prediction model with confusion matrix

Evaluation of the prediction results of the proposed system model uses the confusion matrix technique. The evaluation result using the confusion matrix is shown in Table 10 and Figure 3. The accuracy in predicting with 85% of training data and 15% of test data shows that the RF machine learning method is more accurate and precise than the SVM machine learning method.

TABLE X. SYSTEM MODEL PERFORMANCE TESTING WITH 85% OF TRAINING DATA AND 15% OF TESTING DATA

Method	Accuracy	Precision	Recall
SVM	89%	83%	93%
RF	94%	90%	96%

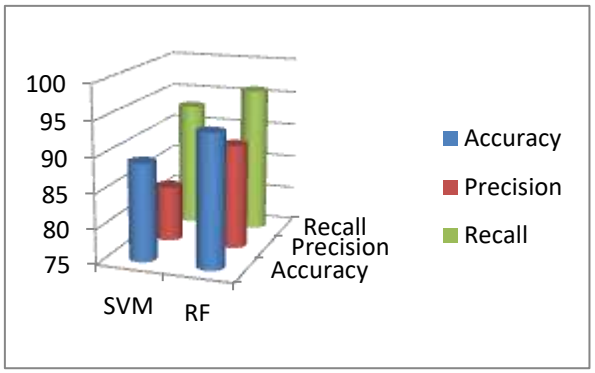

Figure 3. System Model Performance Testing with 85% of Training Data and 15% of Testing Data

Further comparison of the accuracy and precision of the prediction system model with 50% training data and 50% testing data, 60% training data and 40% testing data, 70% training data and 30% testing data, 80% training data and 20% data 90% testing and training data and 10% testing data are as shown in Table 11.

TABLE XI. PREDICTION SYSTEM MODEL PERFORMANCE TESTING WITH VARIOUS TEST DATA AND TRAINING DATA VARIATIONS

Data (in %)		Accuracy (in %)		Precision (in %)		Recall (in %)	
Training	Testing	RF	SVM	RF	SVM	RF	SVM
50	50	95	97	97	96	91	96
60	40	96	91	97	90	93	90
70	30	94	92	96	91	92	90
80	20	96	92	94	89	97	94
90	10	92	83	75	67	95	91
Average		95	91	92	87	94	92

Predicting with various test data and training data variations shows that the RF machine learning method is more accurate and precise than the SVM machine learning method. In other words, the prediction system model proposed to predict the treatment status of Covid-19 patients using the RF method is better (more accurate and precise) than the SVM machine learning method based on performance tests with a confusion matrix.

4.5.2 Evaluation of prediction model with k-fold cross-validation

The performance of the model proposed in this study uses a 5-fold cross-validation on both RF and SVM prediction models presented in Table 12 and Figure 4.

TABLE XII. PREDICTION PERFORMANCE TESTING WITH K-FOLD CROSS-VALIDATION

RF (in %)	SVM (in %)
98.290	97.436

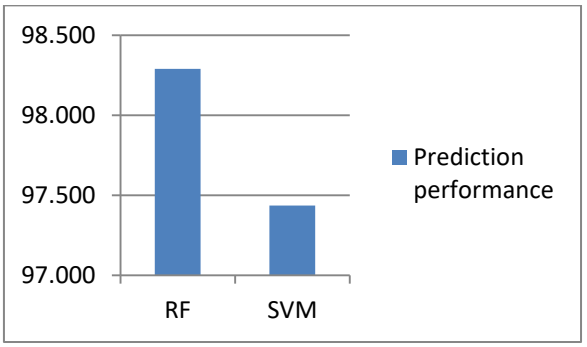

Figure 4. Predictive performance testing using K-fold cross-validation

4.6 Deployment

One of the deployments in this research is making scientific articles on machine learning system models that are produced to be published in reputable scientific papers. Thus, the results obtained can be developed and become the knowledge of many parties as a responsibility for the correctness of the effects of research carried out as professional researchers. Another form of deployment is to make reports to cooperative hospital partners where data on Covid-19 patients is obtained.

V. CONCLUSION

This study found that the prediction system model for the treatment status of Covid-19 patients using the RF machine learning method had better predictive performance than the SVM machine learning method. The test of accuracy and precision in predicting the treatment status of Covid-19 patients using the confusion matrix showed that the RF machine learning method has a prediction accuracy of 94% and a precision of 90%; In comparison, the SVM machine learning method has a prediction accuracy of 89% and a precision of 83%. Further testing of the accuracy of the system model in predicting the treatment status of Covid-19 patients using k-fold cross-validation showed that the RF machine learning method had a prediction accuracy of 98.290% and the SVM machine learning method had a prediction accuracy of 97.436%. The research result implication is that RF machine learning can help or replace the role of medical personnel in predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, with high accuracy.

The novelty of this study is to propose a system model for predicting the treatment status of Covid-19 patients, whether inpatient or self-isolation, which researchers have never studied before using two machine learning methods of RF and SVM.

Further research needs to develop a machine learning system model to predict the death or recovery status of each Covid-19 patient. Another suggestion for future study is: to conduct further research using other data mining methods to predict patient care status and the status of death or recovery from Covid-19 patients and various other diseases, to build a system that not only predicts but also performs clustering, association, and estimates of various other fields of science, including patients' care status, with a combination of

machine learning and the Internet of Things.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHOR CONTRIBUTIONS

All authors undertake work assignments to complete the research and writing of this article jointly. The level of roles and tasks of research work is the basis that places each author as the first, second, and so on as the fifth author.

REFERENCES

[1] C. Rothe *et al.*, "Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany," *N. Engl. J. Med.*, vol. 382, no. 10, pp. 970–971, 2020.

[2] Li Q *et al.*, "Early transmission dynamics in Wuhan, China, of novel coronavirus–infected pneumonia," *N. Engl. J. Med.*, vol. 382, no. 13, pp. 1199–1207, 2020.

[3] L. Zou *et al.*, "SARS-CoV-2 Viral Load in Upper Respiratory Specimens of Infected Patients," *N. Engl. J. Med.*, vol. 382, no. March, pp. 1171–1179, 2020.

[4] M. A. Quiroz-Juárez, A. Torres-Gómez, I. Hoyo-Ulloa, R. D. J. de León-Montiel, and A. B. U'Ren, "Identification of high-risk COVID-19 patients using machine learning," *PLoS One*, vol. 16, no. 9 September, pp. 1–21, 2021.

[5] C. Wang, Z. Wang, G. Wang, J. Y. N. Lau, K. Zhang, and W. Li, "COVID-19 in early 2021: current status and looking forward," *Signal Transduct. Target. Ther.*, vol. 6, no. 1, 2021.

[6] J. Nayak and B. Naik, "Intelligent system for COVID-19 prognosis : a state-of-the-art survey," *Appl. Intell.*, vol. 51, no. 5, pp. 2908–2938, 2021.

[7] C. Huang *et al.*, "Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China," *Lancet*, vol. 395, no. 10223, pp. 497–506, 2020.

[8] B. Mahboub, M. T. A. Bataineh, H. Alshraideh, R. Hamoudi, L. Salameh, and A. Shamayleh, "Prediction of COVID-19 Hospital Length of Stay and Risk of Death Using Artificial Intelligence-Based Modeling," *Front. Med.*, vol. 8, no. December 2019, pp. 1–9, 2021.

[9] T. Singhal, "A Review of Coronavirus Disease-2019 (COVID-19)," *Indian J. Pediatr.*, vol. 87, no. April, pp. 281–286, 2020.

[10] P. H. Dos Santos, S. M. Neves, D. O. Sant'Anna, C. H. de Oliveira, and H. D. Carvalho, "The analytic hierarchy process supporting decision making for sustainable development: An overview of applications," *J. Clean. Prod.*, vol. 212, pp. 119–138, 2019.

[11] D. Laureiro-Martínez and S. Brusoni, "Cognitive Flexibility and Adaptive Decision-Making: Evidence from a laboratory study of expert decision-makers," *Strateg. Manag. J.*, vol. 39, no. 4, pp. 1031–1058, 2018.

[12] D. Prasad, S. Kumar Goyal, A. Sharma, A. Bindal, and V. Singh Kushwah, "System Model for Prediction Analytics Using K-Nearest Neighbors Algorithm," *J. Comput. Theor. Nanosci.*, vol. 16, no. 10, pp. 4425–4430, 2019.

[13] Z. Zhang, "Introduction to machine learning: K-nearest neighbors," *Ann. Transl. Med.*, vol. 4, no. 11, pp. 1–7, 2016.

[14] P. Guleria, S. Ahmed, A. Alhumam, and P. N. Srinivasu, "Empirical Study on Classifiers for Earlier Prediction of COVID-19 Infection Cure and Death Rate in the Indian States," *Healthc.*, vol. 10, no. 1, 2022.

[15] A. Yosipof, R. C. Guedes, and A. T. García-Sosa, "Data mining and machine learning models for predicting drug likeness and their disease or organ category," *Front. Chem.*, vol. 6, no. May, pp. 1–11, 2018.

[16] J. Bullock, A. Luccioni, K. H. Pham, C. S. N. Lam, and M. Luengo-Oroz, "Mapping the landscape of artificial intelligence applications against COVID-19," *J. Artif. Intell. Res.*, vol. 69, pp. 807–845, 2020.

[17] L. Wynants *et al.*, "Prediction models for diagnosis and prognosis of covid-19: Systematic review and critical appraisal," *BMJ*, vol. 369, no. m1328, pp. 1–16, 2020.

[18] A. Anggrawan, A. H. Yassi, C. Satria, B. Arafah, and H. M. Makka, "Comparison of Online Learning Versus Face to Face Learning in English Grammar Learning," in *The 5th International Conference on Computing Engineering and Design (ICCED)*, 2018, pp. 1–4.

[19] A. Anggrawan, "Interaction between learning preferences and methods in face-to-face and online learning," *ICIC Express Lett.*, vol. 15, no. 4, pp. 319–326, 2021.

[20] H. Zhang *et al.*, "A Tool to Early Predict Severe Corona Virus Disease 2019 (COVID-19) : A Multicenter Study using the Risk Nomogram in Wuhan and Guangdong, China.," *Cancer*, vol. 46, no. May, pp. 1–17, 2020.

[21] X. Jiang *et al.*, "Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity," *Comput. Mater. Contin.*, vol. 63, no. 1, pp. 537–551, 2020.

[22] J. Liu *et al.*, "Neutrophil-to-lymphocyte ratio predicts critical illness patients with 2019 coronavirus disease in the early stage," *J. Transl. Med.*, vol. 18, no. 1, pp. 1–12, 2020.

[23] A. Anggrawan, C. Satria, C. K. Nuraini, Lusiana, N. G. A. Dasriani, and Mayadi, "Machine Learning for Diagnosing Drug Users and Types of Drugs Used," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 11, pp. 111–118, 2021.

[24] A. Agrawal, J. S. Gans, and A. Goldfarb, "Exploring the impact of artificial Intelligence: Prediction versus judgment," *Inf. Econ. Policy*, vol. 47, pp. 1–6, 2019.

[25] K. Shankar, A. R. W. Sait, D. Gupta, S. K. Lakshmanaprabu, A. Khanna, and H. M. Pandey, "Automated detection and classification of fundus diabetic retinopathy images using synergic deep learning model," *Pattern Recognit. Lett.*, vol. 133, pp. 210–216, 2020.

[26] C. W. Yancy, "From Mitigation to Containment of the COVID-19 Pandemic: Putting the SARS-CoV-2 Genie Back in the Bottle," *JAMA - J. Am. Med. Assoc.*, vol. 323, no. 19, pp. 1891–1892, 2020.

[27] V. C. Storey and I. Y. Song, "Big data technologies and Management: What conceptual modeling can do," *Data Knowl. Eng.*, vol. 108, no. February, pp. 50–67, 2017.

[28] A. Ribeiro, A. Silva, and A. R. da Silva, "Data Modeling and Data Analytics: A Survey from a Big Data Perspective," *J. Softw. Eng. Appl.*, vol. 08, no. 12, pp. 617–634, 2015.

[29] B. Suvarnamukhi and M. Seshashayee, "Big Data Concepts and Techniques in Data Processing," *Int. J. Comput. Sci. Eng.*, vol. 6, no. 10, pp. 712–714, 2018.

[30] Y. Li, L. Guo, and Y. Guo, "An Efficient and Performance-Aware Big Data Storage System Yang," in *Communications in Computer and Information Science*, 2018, vol. 367 CCIS, no. April, pp. 1–17.

[31] T. Kavzoglu, F. Bilucan, and A. Teke, "Comparison of support vector machines, random forest and decision tree methods for classification of sentinel - 2A image using different band combinations," *ACRS 2020 - 41st Asian Conf. Remote Sens.*, no. November, pp. 1–9, 2020.

[32] A. Sarica, A. Cerasa, and A. Quattrone, "Random forest algorithm for the classification of neuroimaging data in Alzheimer's disease: A systematic review," *Front. Aging Neurosci.*, vol. 9, no. OCT, pp. 1–12, 2017.

[33] M. Sheykhmousa, M. Mahdianpari, H. Ghanbari, F. Mohammadimanesh, P. Ghamisi, and S. Homayouni, "Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review," *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.*, vol. 13, pp. 6308–6325, 2020.

[34] C. Iwendi *et al.*, "COVID-19 patient health prediction using boosted random forest algorithm," *Front. Public Heal.*, vol. 8, no. July, pp. 1–9, 2020.

[35] C. Aroef, Y. Rivan, and Z. Rustam, "Comparing random forest and support vector machines for breast cancer classification," *Telecommun. Comput. Electron. Control*, vol. 18, no. 2, pp. 815–821, 2020.

[36] B. Hao *et al.*, "Early prediction of level-of-care requirements in patients with COVID-19," *Elife*, vol. 9, pp. 1–23, 2020.

[37] H. Zhao *et al.*, "COVID-19: Short term prediction model using daily incidence data," *PLoS One*, vol. 16, no. 4 April, pp. 1–14, 2021.

[38] S. Guhathakurata, S. Kundu, Chakraborty, A. Banerjee, and J. Sekhar, "A novel approach to predict COVID-19 using support vector machine," *Data Sci. COVID-19*, pp. 351–364, 2020.

[39] A. Mehrotra and R. Agarwal, "A Review of Use of Data Mining during COVID-19 Pandemic," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 6, pp. 4547–4552, 2021.

[40] A. Anggrawan, Mayadi, C. Satria, and L. G. R. Putra, "Scholarship

Recipients Recommendation System Using AHP and Moora Methods," *Int. J. Intell. Eng. Syst.*, vol. 15, no. 2, pp. 260–275, 2022.
[41] V. Demichev *et al.*, "A proteomic survival predictor for COVID-19 patients in intensive care," *PLOS Digit. Heal.*, vol. 1, no. 1, pp. 1–17, 2022.
[42] O. Marban, G. Mariscal, and J. Segovia, "A Data Mining & Knowledge Discovery Process Model," no. January, Vienna, Austria: Data Mining and Knowledge Discovery in Real Life Applications, 2009, p. 438.
[43] P. Chapman *et al.*, *Crisp-Dm 1.0: Step-by-step data mining guide*. SPSS, 2000.
[44] A. Anggrawan, C. Satria, Mayadi, and N. G. A. Dasriani, "Reciprocity Effect between Cognitive Style and Mixed Learning Method on Computer Programming Skill," *J. Comput. Sci.*, vol. 17, no. 9, pp. 814–824, 2021.
[45] A. Anggrawan, C. K. Nuraini, Mayadi, and C. Satria, "Interplay between Cognitive Styles and Gender of Two Hybrid Learning to Learning Achievements," *J. Theor. Appl. Inf. Technol.*, vol. 99, no. 10, pp. 2404–2413, 2021.
[46] A. Anggrawan, S. Hadi, and C. Satria, "IoT-Based Garbage Container System Using NodeMCU ESP32 Microcontroller," *J. Adv. Inf. Technol.*, vol. 13, 2022.

Anthony Anggrawan currently works as an associate professor in the Department of Computer Science as a lecturer, university Rector, and State Civil Apparatus at Bumigora University, Indonesia. In addition, he is a reviewer of articles in several reputable international scientific journals. He received his Masters in Computer Science Information Technology (M.T) from the 10 November Institute of Technology, Surabaya, Indonesia. After that, he earned his first Doctoral degree (Ph.D.) in Accounting Information Systems from Universiti Utara Malaysia. Then, he received his second Doctoral degree (Dr.) from Hasanuddin University, Makassar, Indonesia, in linguistics. Finally, he earned his third Doctorate in Educational Technology from the State University of Jakarta. His research interests include Information Technology, Data Mining, Machine Learning, Online Learning, and the Internet of Things. From 2016 until now, he has been the chairman of the Association for Higher Education in Informatics and Computers (APTIKOM) for the West Nusa Tenggara region, Indonesia.

Mayadi obtained a bachelor's degree (S.Kom) in Computer Science from Bumigora University, Indonesia, and a master's degree (M.Kom) in Informatics Engineering from Amikom University, Indonesia. Currently a lecturer in the Computer Science Study Program at Bumigora University, Indonesia, and as Vice Chancellor III. His research interests are Artificial intelligence, Machine Learning, Data Mining, Big Data, Deep learning, the Internet of Things, and Data Science.

Christofer Satria received a bachelor's degree (S.Sn) in Visual Communication Design from Petra Christian University, Surabaya, Indonesia, and a master's degree (M.Sn) in Visual Communication Design from the Indonesian Art Institute (ISI) Denpasar, Bali, Indonesia. He is currently a lecturer in the Visual Communication Design Study Program at Bumigora University, Indonesia, and the head of the laboratory in photography, animation, and video. His research interests include animation learning media, video learning media, education method, Data Mining, and experimental Design. He is currently pursuing a doctorate in the same area as his expertise.

Bambang Krismono Triwijoyo currently works as an assistant professor in the Department of Computer Science as a lecturer and head of the quality assurance institute at Bumigora University, Indonesia. In addition, he is a reviewer of articles in several reputable international scientific journals. He completed his post-graduation in Computer Science from the Sepuluh November Institute of Technology, Surabaya, Indonesia, in 2003. Doctoral degree in Computer Science from Bina Nusantara University, Jakarta, Indonesia, in 2021. His research interests are Computer Vision, Image Processing, and Machine Learning. He is also active in several professional organizations, including IEEE Computer Science, the Indonesian Informatics Experts Association (IAII), and the Indonesian Pattern Recognition Association (INAPR).

Ria Rismayati obtained a bachelor's degree (S.Kom) in Informatics Engineering from the College of Management and Computer Informatics, Mataram, Indonesia, and a master's degree (M.Kom) from AMIKOM University, Yogyakarta, Indonesia. Currently, she is a lecturer in the Computer Science study program at Bumigora University, Indonesia. Her research interests include Data Mining, Artificial Intelligence, Enterprise Architecture, and Information Systems.

Gmail

Anthony Anggrawan <anthony.anggrawan@universitasbumigora.ac.id>

[jait] Manuscript ID: JAIT-5790 - Accepted for Publication

Ms. Yoyo Chow/Section Editor <yoyo.chow@ejournal.net> Thu, Aug 4, 2022 at 11:24 AM
To: Anthony Anggrawan <anthony.anggrawan@universitasbumigora.ac.id>, Mayadi <mayadi.yadot@universitasbumigora.ac.id>, Christofer Satria <chris@universitasbumigora.ac.id>, Krismono <bkrismono@universitasbumigora.ac.id>, Ria Rismayati <riris@universitasbumigora.ac.id>

Dear Anthony Anggrawan, Mayadi, Christofer Satria, Krismono, Ria Rismayati:

We are pleased to inform you that the following paper has been officially accepted for publication in Journal of Advances in Information Technology.

Manuscript ID: JAIT-5790
Title: Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients
Submission URL: http://ojs.ejournal.net/index.php/jait/authorDashboard/submission/5790

We will now make the final preparation, and then return the edited manuscript to you for your approval.

Your paper will be charged for publishing (450 USD), and the detailed payment information can be found at the end of this email. **If the payment is ready, please send us the payment screenshot and order ID for the record.**

Once the payment is confirmed, We will make the final preparation, and then return the edited manuscript to you for your approval.

Payment link: http://confsys.iconf.org/online-payment/18130
Terms of Payment: 10 days

After the confirmation of payment, we will send it to layout and production process. We are excited to move forward with your submission. Please feel free to email us with any questions.

Ms. Yoyo Chow/Section Editor
yoyo.chow@ejournal.net

Please proceed with the payment at the following link (No handling fees)
Payment link: http://confsys.iconf.org/online-payment/18130
Terms of Payment: **10 days**
Please make sure you have VISA or Mastered Card Credit Card before clicking this link, and you should also calculate the right amount and pay.

The following information is necessary.

Full Name*:
Email*:
Event Acronym*: JAIT
Event URL*: http://www.jait.us/
Acceptance/Paper ID*: JAIT-5790
Paper Title*: Comparative Analysis of Machine Learning in Predicting the Treatment Status of Covid-19 Patients
---

--
Journal of Advances in Information Technology
Website: http://www.jait.us/
Email: jait@etpub.com

Submissions

My Queue Archives

Help

Submissions

My Assigned

Search

New Submission

5790	**Anthony Anggrawan, Mayadi, Christofer Satria, Krismono, Ria Ris...** Comparative Analysis of Machine Learning in Predicting the Treatmen...	Production
5193	**Anthony Anggrawan, Hadi Sirojul, Satria Christofer** IoT-Based Garbage Container System Using NodeMCU ESP32 Microco...	Production

2 of 2 submissions