

CONFERENCE PROCEEDINGS



Proceedings

**2021 International Conference on
Software Engineering & Computer
Systems and 4th International Conference
on Computational Science and Information
Management**

ICSECS-ICOCSIM 2021

2021 International Conference on Software Engineering & Computer Systems and 4th International Conference on Computational Science and Information Management (ICSECS- ICOCSIM) ICSECS-ICOCSIM 2021

Table of Contents

Welcome Message from the General Chair **xxi**
Committees **xxii**

1A: IoT, Big Data Analytics, Artificial Intelligence

Infection Disease Classification in Indonesian Language using Machine Learning 1
Mohamad Raychan Abdul Rosyid (Universitas Gadjah Mada, Indonesia) and Afiahayati Afiahayati (Universitas Gadjah Mada, Indonesia)

Internet of Things (Iot) Based Door Lock Security System 6
Tommy Cha Hweay Rhunn (Universiti Malaysia Pahang, Malaysia), Anis Farihan Mat Raffei (Universiti Malaysia Pahang, Malaysia), and Nur Shamsiah Abdul Rahman (Universiti Malaysia Pahang, Malaysia)

A Proposed Requirements Prioritization Model Based on Cost-Value Approach with Collaboration Perspective 10
Tan Amelia (Universiti Malaysia Pahang, Malaysia) and Rozlina Binti Mohamed (Universiti Malaysia Pahang, Malaysia)

SonoNet: Despeckling of Medical Ultrasound Scans using Convolutional Neural Network Architecture 16
Perna Singh (n/a, New Zealand)

Existing Semantic Ontology and Its Challenges for Enhancing Interoperability in IoT Environment 22
Moseed Mohammed (University Malaysia Pahang, Malaysia), Awanis Romli (University Malaysia Pahang, Malaysia), and Rozlina Mohamed (University Malaysia Pahang, Malaysia)

COVID-19 Analysis and Predictions Evaluation for KSA using Machine Learning	261
<i>Rawia Elarabi (Jazan University, KSA), Fatimah Alqahtani (Jazan University, KSA), Awatef Balobaid (Jazan University, KSA), Halah Zain (Jazan University, KSA), and Najla Babiker (Jazan University, KSA)</i>	

2D: Software Engineering, Knowledge Engineering

Comparative Study of QoS Parameters for Different Web Services	267
<i>Md. Abul Kalam Azad (United International University, Bangladesh), Salekul Islam (United International University, Bangladesh), A.K.M. Muzahidul Islam (United International University, Bangladesh), and Md. Saddam Hossain Mukta (United International University, Bangladesh)</i>	
Improving Messenger Accessibility for Elderly Users using User Centered Design (UCD) Methods (Study Case: WhatsApp)	273
<i>Wafa Zahida (Telkom University, Indonesia), Veronikha Effendy (Telkom University, Indonesia), and Aristyo Hadikusuma (Telkom University, Indonesia)</i>	
Fruit Ordering System Through Fruity Healthy Mobile Application	279
<i>Suraya Abu Bakar (Universiti Malaysia Pahang, Malaysia) and Liew Pei Ling (Universiti Malaysia Pahang, Malaysia)</i>	
A Review on Distance Measure Formula for Enhancing Match Detection Process of Generic Code Clone Detection Model in Java Application	285
<i>Noormaizzattul Akmaliza Abdullah (Universiti Malaysia Pahang, Malaysia), Mohd Azwan Mohamad Hamza (Universiti Malaysia Pahang, Malaysia), and Al-Fahim Mubarak Ali (Universiti Malaysia Pahang, Malaysia)</i>	
Effects of Eye Health Among Youngster While Playing Computer Game	291
<i>Nurul Saidatul Akmal Binti Ab Razak (Universiti Malaysia Pahang, Malaysia) and Rahmah Mokhtar (Universiti Malaysia Pahang, Malaysia)</i>	
Enhancing High-Quality User Stories with AQUUSA: An Overview Study of Data Cleaning Process.....	295
<i>Siti Nur Fathin Najwa Binti Mustaffa (Universiti Malaysia Pahang, Malaysia), Jamaludin Bin Sallim (Universiti Malaysia Pahang, Malaysia), and Rozlina Binti Mohamed (Universiti Malaysia Pahang, Malaysia)</i>	

2E: IoT, Artificial Intelligence, Image Processing, E- Learning

Light Deep Learning Model Architecture for Chest X-ray Based Covid-19 Detection	301
<i>Putra Sumari (Universiti Sains Malaysia, Malaysia), Saqib Jamal Syed (Universiti Sains Malaysia, Malaysia), and Liang Han Sheng (Universiti Sains Malaysia, Malaysia)</i>	
Towards an Optimized Dragonfly Algorithm using Hill Climbing Local Search to Tackle the Low Exploitation Problem	306
<i>Bibi Aamirah Shafaa Emambocus (Sunway University, Malaysia) and Muhammed Basheer Jasser (Sunway University, Malaysia)</i>	

The Abstract of Thesis Classifier by using Naive Bayes Method	312
<i>Hairani Hairani (Universitas Bumigora, Indonesia), Anthony Anggrawan (Universitas Bumigora, Indonesia), Ahmad Islahul Wathan (Universitas Bumigora, Indonesia), Kurniadin Abd Latif (Universitas Bumigora, Indonesia), Khairan Marzuki (Universitas Bumigora, Indonesia), and Muhammad Zulfikri (Universitas Bumigora, Indonesia)</i>	
IoT Based Sport Healthcare Monitoring System	316
<i>Muhammad Naim Mohd Samsuddin (Universiti Malaysia Pahang, Malaysia), Anis Farihan Mat Raffei (Universiti Malaysia Pahang, Malaysia), and Nur Shamsiah Abdul Rahman (Universiti Malaysia Pahang, Malaysia)</i>	
Review on Target Tracking Method in Multimedia Wireless Sensor Networks	320
<i>Afritha Amelia (Universitas Sumatera Utara, Indonesia), Muhammad Zarlis (Universitas Sumatera Utara, Indonesia), Suherman Suherman (Universitas Sumatera Utara, Indonesia), and Syahril Efendi (Universitas Sumatera Utara, Indonesia)</i>	
Prediction using A Neural Network Algorithm Approach (A Review)	325
<i>T.H.F Harumy (Universitas Sumatera Utara, Indonesia), M. Zarlis (Universitas Sumatera Utara, Indonesia), S. Effendi (Universitas Sumatera Utara, Indonesia), and M.S Lidya (Universitas Sumatera Utara, Indonesia)</i>	
Features of Single Value Coordinate System (SVCS) for Earthquake Forecasting using Single Layer Hierarchical Graph Neuron (SLHGN)	331
<i>Benny Benyamin Nasution (Politeknik Negeri Medan, Indonesia)</i>	
Neural Network as a Preferred Method for Microarray Data Classification	337
<i>Putri Tsatsabila Ramadhani (University of North Sumatera, Indonesia) and Benny Benyamin Nasution (University of North Sumatera, Indonesia)</i>	

2F: Information Management, E-Learning

Sentinel: The Development of a Web and Mobile Application for the Development and Testing of an E-Service Learning Interprofessional Telehealth Community Based Rehabilitation Program Among Hypertensive Clients	341
<i>Rafael Benedict E. Bacungan (University of Santo Tomas, Philippines), Kurt Martin C. Choi (University of Santo Tomas, Philippines), Jansen Patrick A. Chua (University of Santo Tomas, Philippines), Jericho P. Dupo (University of Santo Tomas, Philippines), and Noel Estrella (University of Santo Tomas, Philippines)</i>	
A Mathematical Educational Game Application for Primary School Slow Learner	348
<i>Nur Syakirah Binti Kamarulzaman (Universiti Malaysia Pahang, Malaysia), Danakorn Nincarean Eh Phon (Universiti Malaysia Pahang, Malaysia), and Mohd Syazwan Baharuddin (Universiti Kebangsaan Malaysia, Malaysia)</i>	
Serious Games and Preventive Self-Care for Diabetes: A Conceptual Framework	354
<i>Siti Normaziah Ihsan (Universiti Malaysia Pahang, Malaysia), Tuty Asmawaty Abd Kadir (Universiti Malaysia Pahang, Malaysia), and Abdul Rafiq Abdullah (Universiti Malaysia Pahang, Malaysia)</i>	

Committee Members

GENERAL CHAIR

DR. JAMALUDIN SALLIM

CHAIR

DR. FAUZIAH ZAINUDDIN (ICSECS)

DR. MOHD AZWAN HAMZA (ICoCSIM)

SECRETARY

DR. MOHD ZAMRI OSMAN

DR. NOOR AZIDA SAHABUDIN

DR. NUR HAFIEZA BINTI ISMAIL

TREASURER

DR. NUR SHAMSI AH ABDUL RAHMAN

PROCEEDING COMMITTEE

ASSOC. PROF. DR. ARAFATUR RAHMAN

DR. SYIFAK IZHAR HISHAM

MEMBERS

DR. ROZLINA MOHAMED

DR. YUSNITA MUHAMAD NOOR

DR. MOHD IZHAM MOHD JAYA

TS. DR. MRITHA RAMALINGAM

TS. DR. KOHBALAN A/L MOORTHY

DR. ZAHIAN BINTI ISMAIL

TECHNICAL COMMITTEE

DR. ZAFRIL RIZAL AZMI

TS. DR. TAHA HUSSEIN ALAALDEEN RASSEM

MEMBERS

DR. LIEW SIAU CHIN

TS. DR. MOHD ARFIAN ISMAIL

DR. NOR SYAHIDATUL NADIAH ISMAIL

DR. SALWANA MOHAMAD @ ASMARA

DR. ABDULLAH NASSER

INTERNATIONAL AFFAIRS

ASSOC. PROF. DR. ABDUL RAHMAN ALSEWARI (ICSECS)

TS. DR. FERDA ERNAWAN (ICSECS)

DR. AHMAD FAKHRI BIN AB. NASIR (ICoCSIM)

INDUSTRIAL & ALUMNI

DR. ZALILI MUSA
DR. SYAFIQ FAUZI KAMARULZAMAN

SPONSORSHIP COMMITTEE

DR. AHMAD FIRDAUS ZAINAL ABIDIN
DR. AZLEE ZABIDI

MEMBERS

DR. RAHMAH MOKHTAR
DR. SURAYA ABU BAKAR
ASSOC. PROF. DR MOHAMED ARIFF AMEEDDEEN
DR NGAHZAIFA AB GHANI
TS DR AWANIS ROMLI
ARIFIN SALLEH

HOSPITALITY, LOGISTIC & TRANSPORTATION COMMITTEE

DR. LUHUR BAYUAJI
DR. ANIS FARIHAN MAT RAFFEI
AMINATUL NOR MOHAMED SAID

MEMBERS

ROSLINA MOHD SIDEK

MEDIA & PUBLICITY COMMITTEE

DR. DANAKORN NINCAREAN

MEMBERS

NURLINDA KADRI
IZYAN HUSNA IBRAHIM

PROMOTION COMMITTEE

DR. NUR SHAZWANI KAMARUDIN
DR. NOR SARADATUL AKMAR

EVENT MANAGEMENT

DR. MOHD FAIZAL AB RAZAK
NORANIZA SAMAT

SECRETARIAT

ISMALINA MOHD ISAH
NORAINI MD ZAMRI
NOOR ASHIKIN RAMLY
NOOR AFTALINA OMAR
DARWINA RASTAM TAN
SYARIFAH AZLIN SYED YUSOP
SURIANA ABIDIN

TECHNICAL SUPPORT

CHE YAHAYA YAAKOB
DR. MUHAMMAD 'ARIF BIN MOHAMAD
TS. DR. HOH WEI SIANG
IMRAN EDZEREIQ
SYAHRULANUAR NGAH
MAHMUD ABDUL SAMAD
AMIRUL HUSNI ABDUL GHAFAR
MOHD AMERUL SHUIB
MOHD NAIM GATI @MOHD GATI
MUHAMMAD TAUFIK MOHAMAD REFFIN
MOHD FAIZUL GHAFAR
ABDUL RAHMAN ABDUL KARIM
WAN MD NAHARRUDDIN WAN ZULKIFLI
AHMAD MUSTAQIM MOHAMAD GHANI
KHAIRIL CHAIRY
MASTURA SARKON
RUZAINAH ABDULLAH
KHAIRUN NISSAK ABDULLAH

PROGRAM & PROTOCOL COMMITTEE

DR. JUNAIDA SULAIMAN
TS. DR SURYANTI AWANG
DR. WAN ISNI SOFIAH WAN DIN
DR. ABDUL SAHLI FAKHARUDIN
DR. ZURIANI MUSTAFFA
DR NOORLIN MOHD ALI
TS. DR NOORHUZAIMI@KARIMAH
DR. BARIAH YUSOF
DR. NABILAH FILZAH BINTI MOHD RADZUAN
DR. SITI SUHAILA BINTI ABDUL HAMID

The Abstract of Thesis Classifier by Using Naive Bayes Method

Hairani Hairani
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
hairani@universitasbumigora.ac.id

Anthony Anggrawan
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
anthony.anggrawan@universitasbumigora.ac.id

Ahmad Islahul Wathan
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
wathanruna@gmail.com

Kurniadin Abd Latif
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
kurniadin@universitasbumigora.ac.id

Khairan Marzuki
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
khairan@universitasbumigora.ac.id

Muhammad Zulfikri
Faculty of Engineering and Design
Universitas Bumigora
Mataram, Indonesia
mzulfikri@universitasbumigora.ac.id

Abstract— The thesis is a requirement for graduation from Bumigora university. The final year student's problem is determining the research topic because the undergraduate thesis collection of Computer Science is not grouped or classified based on student competencies. The purpose of this study was to compare the performance of the naïve Bayes method with TF-IDF weighting and without TF-IDF weighting for the classification of thesis topics based on the abstract. The stages of this research are data collection, text pre-processing, term weighting with TF-IDF and without TF-IDF, Naive Bayes method implementation, and result evaluation. Based on the results of the tests that have been done, the naïve Bayes method with TF-IDF has an accuracy of 81.74%, a precision of 86.1%, and a sensitivity of 80.15%. While the naïve Bayes method without TF-IDF weighting produces 88.69% accuracy, 89.76% precision, and 90.49% sensitivity. Thus, the naïve Bayes method without TF-IDF weighting has better performance than TF-IDF weighting for the classification of thesis topics based on the abstract.

Keywords—naïve bayes, TF-IDF weighting, abstract classification

I. INTRODUCTION

The thesis is one of the graduation requirements for undergraduate Computer Science students at Bumigora University. Students can start working on their thesis if the research topic has been approved through a synopsis exam. So far, students have difficulties in determining the proposed thesis topic. One of the difficulties is because the existing collection of an undergraduate thesis in Computer Science is not grouped or classified based on student competencies. Automatic thesis grouping or classification of topics is one solution that can make it easier for students to find references to research titles based on their competence. The competencies of students in the S1 Computer Science program at Bumigora university are computer networks, multimedia, and software engineering (RPL).

One of the solutions offered by this paper is to use the concept of text mining. Previous research used various methods for text mining-based thesis document analysis such as the k-means method [1]–[4], K-Nearest Neighbor [5]–[7], Cosine Similarity [8], [9], Decision Tree and Naïve Bayes [10], SVM and Naïve Bayes [11]. Research [10] compared Decision Trees, Naïve Bayes, and k-NN methods to predict thesis graduation. Based on the results of his research, the k-NN method has the best accuracy compared to the decision tree and naïve Bayes methods at 80.39%. Research [4] used

the k-means method for grouping thesis titles. Before grouping, the first weighting of words is carried out using the TF - IDF method. Research [9] uses the cosine similarity method for the classification of thesis documents. Before grouping, the first weighting of words is carried out using the TF - IDF method.

Based on previous research, there is a difference made with this research, namely the research carried out a classification of thesis topics based on the abstract using the naïve Bayes method and also using the k-fold cross-validation test method. The aim is to compare the performance of the naïve Bayes method with TF-IDF weighting and without TF-IDF weighting for the classification of thesis topics based on the abstract. The performance used in this study is accuracy, precision, and sensitivity.

II. RESEARCH METHOD

The stages used in this study are shown in Figure 1.

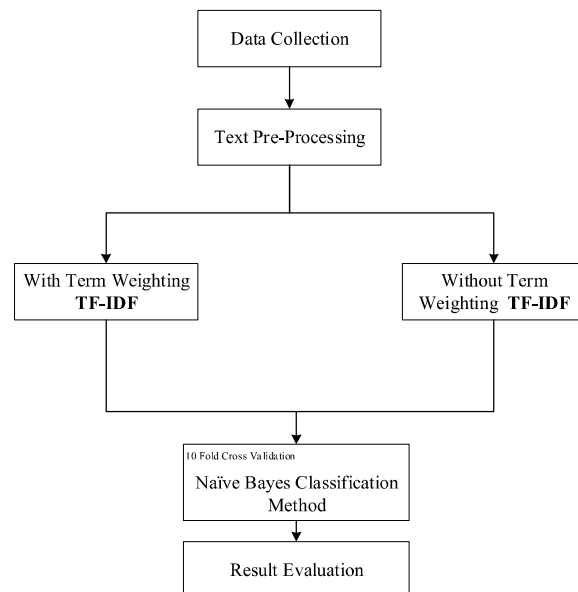


Figure 1. Research Methodology

A. Data Collection

The data used in this study are the thesis abstracts of 2020 graduate computer science undergraduate students obtained

from www.repository.universitاسbumigora.ac.id. The data collected were 115 thesis abstract data, consisting of 36 topics of computer networks, 23 multimedia, and 55 software engineering (RPL).

B. Text Pre-Processing

Text pre-processing used to get quality data. The classification was carried out using the naïve Bayes method. The text pre-processing technique used in this study consists of case folding, tokenization, stop word removal, and stemming [12]. Case folding is used to change text to lowercase. Tokenization is used to separate text into tokens. Stopword removal is used to remove unnecessary words such as conjunctions. Stemming is used to change all words that have affixes into basic words.

C. Term Weighting TF-IDF

The term weighting process is used to give a weight value to each word. The term weighting method used in this study is the Term Frequency - Inverse Document Frequency (TF-IDF). The TF-IDF method combines two concepts, namely TF and IDF. TF looks for the occurrence value of terms in related documents, the more occurrences of terms in the related document, the better. Meanwhile, the IDF concept is inversely proportional to the TF method, the less frequently the terms appear in all documents the better. TF - IDF method is calculated using equation (1) [13].

$$W_{ij} = tf_{ij} \times idf_j = tf_{ij} \times \log\left(\frac{N}{df_j}\right) \quad (1)$$

W_{ij} is the weight of term j to document i. tf_{ij} is the number of occurrences of term j in the document d. N is the number of documents, and df_j is the number of occurrences of term j throughout the document.

D. Naïve Bayes Method Classification

At this stage, the classification is carried out using the naïve Bayes method. The basic concept of the naïve Bayes method is a probability-based classification method that assumes independence from the dependent variable and is also a conditional model based on the Bayes theorem [14][15]. The Naïve Bayes method which is calculated based on equation (2).

$$P(c | \text{term document } d) = P(c) \times P(t_1|c) \times P(t_2|c) \times P(t_n|c) \quad (2)$$

$P(c)$ is the prior probability of class c. $P(c | \text{term document } d)$ is the probability of the appearance of a term in document d including class c. $P(t_n | c)$ is the probability of occurrence of term n known to class c.

The process of calculating the prior probability for class c uses equation (3).

$$P(c) = \frac{N_c}{N} \quad (3)$$

N_c is the number of class c in all documents, while N is the total number of documents. The calculation of the probability of occurrence of term n is calculated using equation (4) involving the laplacian technique.

$$P(t_n | c) = \frac{\text{count}(t_n, c) + 1}{\text{count}(c) + |v|} \quad (4)$$

$\text{count}(t_n, c)$ is the number of terms t_n appearing in the training data with class c. $\text{count}(c)$ is the number of terms in the class training data c. weighting is used to give weight to the value of each word. is the number of terms in the training data. V is the number of terms in the training data.

Data classified by the multinomial naïve Bayes method are grouped into training and testing data first. The distribution of training and testing data in this study uses the k-fold cross-validation method by dividing the data as much as the specified k. Each fold can be used as training and testing data in turn. This research uses 10 fold data validation method.

E. Result Evaluation

At this stage, the results are evaluated based on accuracy, precision, and sensitivity using the confusion matrix table shown in Table I.

TABLE I. CONFUSION MATRIX

Actual	Predicted	
	Positive	Negative
Positive	True Positive (TP)	False Negative (FN)
Negative	False Positive (FP)	True Negative (TN)

Evaluation of results based on accuracy, precision, and sensitivity using equations (5), (6), and (7).

$$\text{Accuracy} = \frac{TP+TN}{TP+FN+TN+FP} \quad (5)$$

$$\text{Precision} = \frac{TP}{TP+FP} \quad (6)$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} \quad (7)$$

III. RESULT AND DISCUSSION

A. Data Collection

The data used in this study are the thesis abstracts of 2020 graduate computer science undergraduate students obtained from www.repository.universitاسbumigora.ac.id. The data collected were 115 thesis abstract data, consisting of 36 topics of computer networks, 23 multimedia, and 55 software engineering (RPL). The sample abstract data of this research thesis is shown in Table II.

TABLE II. THESIS ABSTRACT DATASET

No	Abstract	Topic
1.	Perkembangan teknologi informasi sangat cepat seperti Internet of Things (IoT), dimana seseorang dapat melakukan segala aktivitasnya dengan mudah dengan mengandalkan sistem Internet of Things (IoT). Seiring dengan perkembangan zaman maka semakin canggih teknologi yang dihasilkan baik digunakan sebagai hal yang positif maupun melakukan hal yang negatif, tak terkecuali pada system peternakan sehingga perlu mengembangkan teknologi untuk manajemen pakan ternak khususnya hewan ternak ayam broiler. Pengembangan sistem menggunakan sistem Internet of Things dan sistem penjadwalan otomasi dimana sistem Internet of Things (IoT) adalah sistem yang berfungsi melakukan controller pada alat elektronik. Metodologi	Jaringan

No	Abstract	Topic
	penelitian yang digunakan adalah Network Development Life Cycle (NDLC), terdiri dari; analisis, desain, prototype dan ujicoba. Pada tahap analisis memuat tentang pengumpulan data, tahap desain memuat rancangan sistem pemberian pakan ternak, prototyping memuat instalasi konfigurasi dan membangun kerangka sistem pakan ternak. Ujicoba memuat tentang pengujian sistem pemberian pakan ternak secara otomatis atau terjadwal. Kesimpulan dari penelitian ini adalah mengimplementasi Sever VPS dengan sistem nodemcu dalam pemberian pakan ternak berbasis Internet of Things (IoT) untuk efisiensi dalam pemberian pakan ternak ayam.	

B. Text Pre-Processing

Text pre-processing used to get quality data. The classification was carried out using the naïve Bayes method. The text pre-processing technique used in this study consists of case folding, tokenization, stop word removal, and stemming. The examples of text pre-processing stages are shown in Table III.

TABLE III. EXAMPLE OF TEXT PREPROCESSING

Pre-processing	Result
Data Original	Tujuan pembuatan sistem pakar diagnosis jenis penyakit THT adalah memudahkan masyarakat umum untuk mengetahui jenis penyakit THT diderita tanpa perlu datang ke dokter spesialis THT
Case Folding	tujuan pembuatan sistem pakar diagnosis jenis penyakit tht adalah memudahkan masyarakat umum untuk mengetahui jenis penyakit tht diderita tanpa perlu datang ke dokter spesialis tht
Tokenization	['tujuan', 'pembuatan', 'sistem', 'pakar', 'diagnosis', 'jenis', 'penyakit', 'tth', 'adalah', 'memudahkan', 'masyarakat', 'umum', 'untuk', 'mengetahui', 'jenis', 'penyakit', 'tth', 'diderita', 'tanpa', 'perlu', 'datang', 'ke', 'dokter', 'spesialis', 'tth']
stop word removal	['sistem', 'pakar', 'diagnosis', 'jenis', 'penyakit', 'tth', 'masyarakat', 'mengetahui', 'jenis', 'penyakit', 'tth', 'diderita', 'dokter', 'spesialis', 'tth']
stemming	['sistem', 'pakar', 'diagnosis', 'jenis', 'sakit', 'tth', 'masyarakat', 'tahu', 'jenis', 'sakit', 'tth', 'derita', 'dokter', 'spesialis', 'tth']

C. Term Weighting TF-IDF

The term weighting process is used to give weight to the value of each word. The term or word weighting method used in this study is TF-IDF. The example of the TF-IDF calculation process using the documents in Tabel III, the stemming section, is shown in Table IV.

TABLE IV. RESULT OF WEIGHTING TERM TF-IDF

Term	tf				W= tf * (IDF+1)
	D1	D	D/df	log (IDF)+1	D1
datang	1	1	1	1	1
derita	1	1	1	1	1
diagnosis	1	1	1	1	1
dokter	1	1	1	1	1
jenis	2	1	1	1	2
masyarakat	1	1	1	1	1

pakar	1	1	1	1	1
sakit	2	1	1	1	2
sistem	1	1	1	1	1
spesialis	1	1	1	1	1
tth	3	1	1	1	3

D. Naïve Bayes Method Classification

At this stage, the classification is carried out using the naïve Bayes method by comparing the performance using TF-IDF weighting and without TF-IDF weighting using equation (2).

E. Result Evaluation

At this stage, results are evaluated based on accuracy, precision, and sensitivity using the confusion matrix table shown in Table V, VI, and VII.

TABLE V. CONFUSION MATRIX OF NAÏVE BAYES WITH TF - IDF

Actual	Predicted			Sensitivity
	Jaringan	Multimedia	RPL	
Jaringan	31	0	6	83.78%
Multimedia	0	18	9	66.67%
RPL	4	1	45	90%
Precision	88.57%	94.74%	75%	

TABLE VI. CONFUSION MATRIX OF NAÏVE BAYES WITHOUT TF - IDF

Actual	Predicted			Sensitivity
	Jaringan	Multimedia	RPL	
Jaringan	33	0	4	89.19%
Multimedia	0	26	1	96.29%
RPL	5	2	43	86%
Precision	86.84%	92.86%	89.58%	

TABLE VII. PERFORMANCE RESULT OF NAÏVE BAYES METHOD

Performance	With TF - IDF	Without TF - IDF
Accuracy	81.74%	88.69%
Precision	86.1%	89.76%
Sensitivity	80.15%	90.49%

Based on the results of the tests shown in Table VII, the naïve Bayes method with TF-IDF has an accuracy of 81.74%, a precision of 86.1%, and a sensitivity of 80.15%. While the naïve Bayes method without TF-IDF weighting produces **88.69%** accuracy, **89.76%** precision, and **90.49%** sensitivity. Thus, the naïve Bayes method without TF-IDF weighting has better performance than TF-IDF weighting for the classification of thesis topics based on the abstract.

IV. CONCLUSION

Based on the results of the tests that have been done, the naïve Bayes method with TF-IDF has an accuracy of 81.74%,

a precision of 86.1%, and a sensitivity of 80.15%. While the naïve Bayes method without TF-IDF weighting produces **88.69%** accuracy, **89.76%** precision, and **90.49%** sensitivity. Thus, the naïve Bayes method without TF-IDF weighting has better performance than TF-IDF weighting for the classification of thesis topics based on the abstract. The suggestions for further research can use feature selection methods such as chi-square to improve the performance of the naïve Bayes method.

REFERENCES

- [1] D. Adhe, C. Rachman, R. Goejantoro, and D. Tisna, "Implementation Of Text Mining For Grouping Thesis Documents Using K-Means Clustering," *J. EKSPONENSIAL*, vol. 11, no. 2, pp. 167–174, 2020.
- [2] A. A. Jalal and B. H. Ali, "Text documents clustering using data mining techniques," *Int. J. Electr. Comput. Eng.*, vol. 11, no. 1, pp. 664–670, Feb. 2021, doi: 10.11591/ijece.v11i1.pp664-670.
- [3] M. Sholehudin, M. Fauzi Ali, and S. Adinugroho, "Implementasi Metode Text Mining dan K-Means Clustering untuk Pengelompokan Dokumen Skripsi (Studi Kasus : Universitas Brawijaya)," vol. 2, no. 11, pp. 5518–5524, 2018.
- [4] L. Zahrotun, N. H. Putri, and A. Nur Khusna, "The Implementation of K-Means Clustering Method in Classifying Undergraduate Thesis Titles," in *2018 12th International Conference on Telecommunication Systems, Services, and Applications (TSSA)*, Oct. 2018, pp. 1–4, doi: 10.1109/TSSA.2018.8708817.
- [5] D. M. U. Atmaja and R. Mandala, "Analisa Judul Skripsi untuk Menentukan Peminatan Mahasiswa Menggunakan Vector Space Model dan Metode K-Nearest Neighbor," *IT Soc.*, vol. 4, no. 2, pp. 1–6, Aug. 2020, doi: 10.33021/itfs.v4i2.1182.
- [6] M. Eminağaoğlu and Y. Gökşen, "A New Similarity Measure for Document Classification and Text Mining," *KnE Soc. Sci.*, vol. 2019, pp. 353–366, Jan. 2020, doi: 10.18502/kss.v4i1.5999.
- [7] S. A. Salloum, M. Al-Emran, A. A. Monem, and K. Shaalan, "Using text mining techniques for extracting information from research articles," *Stud. Comput. Intell.*, vol. 740, no. November, pp. 373–397, 2018, doi: 10.1007/978-3-319-67056-0_18.
- [8] R. Rismanto, A. Rachmad Syulistyo, and B. P. Citra Agusta, "Research Supervisor Recommendation System Based on Topic Conformity," *Int. J. Mod. Educ. Comput. Sci.*, vol. 12, no. 1, pp. 26–34, Feb. 2020, doi: 10.5815/ijmecs.2020.01.04.
- [9] R. T. Wahyuni, D. Prastiyanto, and E. Suprptono, "Penerapan Algoritma Cosine Similarity dan Pembobotan TF-IDF pada Sistem Klasifikasi Dokumen Skripsi," *J. Tek. Elektro*, vol. 9, no. 1, pp. 18–23, 2017, [Online]. Available: <https://journal.unnes.ac.id/nju/index.php/jte/article/view/10955/6659>.
- [10] A. Solichin, "Comparison of Decision Tree, Naïve Bayes and K-Nearest Neighbors for Predicting Thesis Graduation," in *2019 6th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Sep. 2019, pp. 217–222, doi: 10.23919/EECSI48112.2019.8977081.
- [11] S. Sulova, L. Todoranova, B. Penchev, and R. Nacheva, "Using Text Mining to Classify Research Papers," in *International Multidisciplinary Scientific GeoConference Surveying Geology and Mining Ecology Management, SGEM*, Jun. 2017, vol. 17, no. 21, pp. 647–654, doi: 10.5593/sgem2017/21/S07.083.
- [12] A. Rahman and A. Doewes, "Online News Classification Using Multinomial Naive Bayes," *ITSMART J. Ilm. Teknol. dan Inf.*, vol. 6, no. 1, pp. 32–38, 2017.
- [13] A. Prabhat and V. Khullar, "Sentiment classification on big data using Naïve bayes and logistic regression," in *2017 International Conference on Computer Communication and Informatics (ICCCI)*, 2017, pp. 1–5, doi: 10.1109/ICCCI.2017.8117734.
- [14] J. Ortigosa-Hernández, J. D. Rodríguez, L. Alzate, M. Lucania, I. Inza, and J. A. Lozano, "Approaching Sentiment Analysis by using semi-supervised learning of multi-dimensional classifiers," *Neurocomputing*, vol. 92, pp. 98–115, Sep. 2012, doi: 10.1016/j.neucom.2012.01.030.
- [15] X. Glorot, A. Bordes, and Y. Bengio, "Domain Adaptation for Large-Scale Sentiment Classification: A Deep Learning Approach," in *Proceedings of the 28th International Conference on International Conference on Machine Learning*, 2011, pp. 513–520.