

JTIK ID 3428

by Jtik 3428

Submission date: 08-Apr-2020 08:34AM (UTC+0700)

Submission ID: 1292374633

File name: 3428-12792-1-RV.doc (1.58M)

Word count: 3033

Character count: 18472

PENINGKATAN KINERJA METODE SVM MENGGUNAKAN METODE KNN IMPUTASI DAN *K-MEANS-SMOTE* UNTUK KLASIFIKASI KELULUSAN MAHASISWA UNIVERSITAS BUMIGORA

*Penulis Korespondensi

1

(Naskah masuk: dd mmm yyyy, diterima untuk diterbitkan: dd mmm yyyy)

Abstrak

Salah satu permasalahan utama universitas bumigora adalah rasio antara mahasiswa yang masuk dengan mahasiswa lulus tepat waktu tidak seimbang, sehingga akan mengakibatkan penurunan penilaian akreditasi dikemudian hari. Penelitian ini menggunakan Teknik data mining untuk mengatasi permasalahan tersebut. Tujuan penelitian ini adalah meningkatkan kinerja metode SVM untuk klasifikasi kelulusan mahasiswa universitas bumigora menggunakan metode KNN Imputasi dan *K-Means-Smote*. Penelitian ini terdiri dari beberapa tahapan yaitu pengumpulan data kelulusan mahasiswa, pra-pengolahan seperti penanganan *missing values* menggunakan metode KNNI, penanganan ketidakseimbangan kelas menggunakan *K-Means-Smote*, klasifikasi menggunakan metode SVM, dan tahapan terakhir adalah pengujian kinerja SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Berdasarkan hasil pengujian yang telah dilakukan, integrasi metode KNNI, *K-Means-Smote*, dan SVM mendapatkan akurasi 92%, sensitivitas 95%, spesifisitas 90.38%, dan F1 91.26%. Penggunaan metode KNNI dan *K-Means-Smote* dapat meningkatkan kinerja metode SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*.

Kata kunci: kelulusan mahasiswa, KNN imputasi, *k-means-smote*, SVM

IMPROVEMENT PERFORMANCE OF SVM METHOD USING KNN IMPUTATION AND *K-MEANS-SMOTE* METHOD FOR GRADUATION CLASSIFICATION OF BUMIGORA UNIVERSITY STUDENTS

Abstract

One of the main problems of bumigora university is that the ratio between incoming students and students graduating on time is unbalanced, so that it will result in a decrease in the accreditation assessment in the future. This study uses data mining techniques to overcome these problems. The purpose of this study is to improve the performance of SVM method for the graduation classification of bumigora university students using KNN Imputation and *K-Means-Smote* methods. This research consists of several stages, namely the collection of student graduation data, pre-processing such as handling missing values using KNNI method, handling class imbalances using *K-Means-Smote*, classification the SVM method, and the last stage is testing SVM performance based on accuracy, sensitivity, specificity, and *f-measure*. Based on the results of test that have been carried out, the integration of the KNNI, *K-Means-Smote*, and SVM method get an accuracy of 92%, sensitivity 95%, specificity 90.38%, and *f-measure* 91.26%. The use of KNNI and *K-Means-Smote* method can improve the performance of the SVM method based on accuracy, sensitivity, specificity, and *f-measure*.

Keywords: students graduation, *knn imputation*, *k-means-smote*, SVM

1. PENDAHULUAN

Universitas bumigora salah satu universitas swasta yang terdapat di Nusa Tenggara Barat. Salah satu permasalahan utama yang dihadapi adalah terjadi peningkatan mahasiswa baru yang ingin

belajar di universitas tersebut setiap tahunnya, akan tetapi tidak diimbangi peningkatan mahasiswa yang lulus tepat waktu. Permasalahan tersebut mengakibatkan rasio antara yang masuk dengan yang keluar tidak seimbang, sehingga akan

mengakibatkan penurunan penilaian akreditasi dikemudian hari.

Salah satu cara untuk menangani permasalahan diatas, dapat menggunakan teknik data mining untuk prediksi dini kelulusan mahasiswa yang tepat waktu atau tidak, sehingga para *decision maker* dapat mengambil kebijakan dengan cepat dan tepat. Data mining merupakan proses untuk mengolah data menjadi informasi sehingga mendapatkan pola atau *pattern* dari sekumpulan data yang besar (Han, Kamber and Pei, 2012).

Penelitian klasifikasi kelulusan mahasiswa menggunakan teknik data mining sudah banyak dilakukan. Penelitian (Maesya and Hendiyanti, 2019) menggunakan metode *CART* untuk prediksi kelulusan mahasiswa dengan akurasi sebesar 77.5%. Penelitian (Arifin and Hadiana, 2019) menggunakan metode *C4.5* untuk prediksi kegagalan studi mahasiswa dengan akurasi sebesar 82.52%. Penelitian (Aesy and Wardoyo, 2019) menggunakan metode *Case Based Reasoning* (CBR) untuk prediksi kelulusan mahasiswa dengan akurasi 76%. Penelitian (Bisri and Rachmatika, 2019) menggunakan gabungan metode *Gradient Boosted Tree* (GBT), *SMOTE*, dan *bagging* untuk klasifikasi kelulusan mahasiswa dengan akurasi sebesar 80.57%. Penelitian (Fadli, Zulfa at 23 Ramadhani, 2018) membandingkan metode *C4.5*, *Support Vector Machine* (SVM) dan *Artificial Neural Network* (ANN) untuk klasifikasi kelulusan mahasiswa. Berdasarkan hasil penelitiannya, metode SVM memiliki akurasi terbaik dibandingkan metode ANN dan *C4.5* sebesar 90.55%.

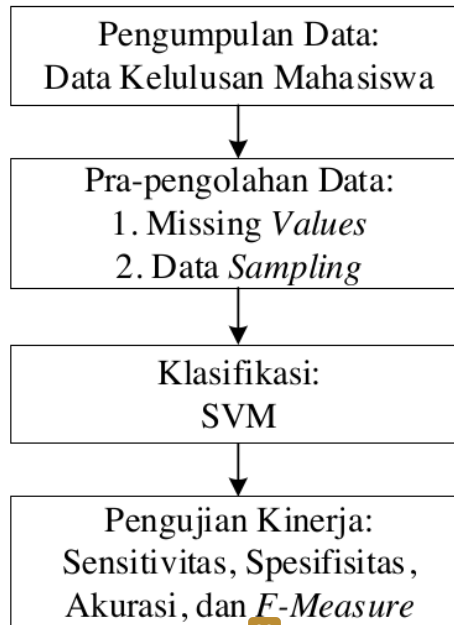
4 Berdasarkan kajian literatur diatas terdapat perbedaan penelitian yang dilakukan dengan penelitian sebelumnya adalah penelitian ini menggunakan integrasi metode SVM, KKN Imputasi (KNNI), dan K-Means-Smote. Metode KNNI digunakan untuk menangani permasalahan *missing values* pada data kelulusan yang digunakan, K-Means-Smote digunakan untuk menangani ketidakseimbangan kelas pada data kelulusan yang digunakan, dan metode SVM sebagai metode klasifikasi.

Penggabungan metode KNNI dan *K-Means-Smote* bertujuan untuk meningkatkan kinerja metode SVM berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*.

4

2. METODE PENELITIAN

Alur penelitian ini mengacu pada Gambar 1.



Gambar 1. Alur Penelitian

2.1. Pengumpulan Data

Pengumpulan data dilakukan di universitas bumigora. Data yang diambil adalah data kelulusan mahasiswa angkatan 2009-2012 berjumlah 302 data dan memiliki 8 atribut. Atribut data kelulusan yang digunakan penelitian ini ditunjukkan pada Tabel 1.

Tabel 1. Atribut Data Kelulusan Mahasiswa Universitas Bumigora

No	Atribut	Keterangan	Tipe Data
1.	Jenis Kelamin	Jenis Kelamin	Nominal (Laki, Perempuan)
2.	IPS 1	Indeks Prestasi Semester 1	Numerik
3.	IPS 2	Indeks Prestasi Semester 2	Numerik
4.	IPS 3	Indeks Prestasi Semester 3	Numerik
5.	IPS 4	Indeks Prestasi Semester 4	Numerik
6.	IPS 5	Indeks Prestasi Semester 5	Numerik
7.	IPK	Indeks Prestasi Kumulatif	Numerik
8.	Status Kelulusan	Kelas (Variabel Target)	Nominal (Tepat Waktu, Tidak Tepat Waktu)

2.2. Pra-Pengolahan Data

Pra-pengolahan data diperlukan untuk menjamin kualitas data yang digunakan pada proses klasifikasi agar metode SVM mendapatkan kinerja yang optimal. Pra-pengolahan yang digunakan pada penelitian ini adalah penanganan *missing values* menggunakan metode KNNI dan data *sampling* menggunakan metode *K-Means-Smote*. Data kelulusan mahasiswa yang digunakan memiliki 7 atribut *input* dan 1 atribut *target*, dan jumlah datanya sebanyak 302 *instance*.

2.2.1. Missing Values

Missing values merupakan sebuah nilai atau informasi yang hilang dari data dan dapat menurunkan akurasi metode klasifikasi(Hartini, 2017). Data kelulusan yang digunakan sebanyak 302 instance, terdapat missing values sebanyak 94 instance. Penelitian ini menggunakan metode KNNI untuk menangani data missing values, karena memiliki kinerja yang baik (Hartini, 2017). Metode KNNI merupakan sebuah metode yang mengimplementasikan konsep kerja metode nearest neighbor. Metode KNNI menangani permasalahan missing values dengan cara mengisi data yang terdapat missing values dengan nilai tetangga terdekat berdasarkan k diberikan. Data yang terdapat missing values ditunjukkan pada Tabel 2, dan data missing values yang sudah ditangani menggunakan metode KNNI ditunjukkan pada Tabel 3.

Tahapan-tahapan metode KNNI (Hartini, 2017) (Minakshi, Vohra and Gimpy, 2014):

1. Tentukan nilai K, Penelitian ini menggunakan K = 5.
2. Menghitung jarak antar instance nilai missing dan instance data training. Formula digunakan untuk menghitung jarak Euclidean ditunjukkan pada persamaan (1).

$$D(x,y) = \sum_{i=1}^n \sqrt{x_i^2 - y_i^2} \quad (1)$$

3. Setelah menghitung jarak Euclidean dipilih nilai data yang memiliki jarak terdekat. Jika menggunakan nilai K=5, maka memilih 5 nilai dengan jarak terdekat. Hitung rata-rata dari nilai K=5 yang terdekat.
4. Menghitung nilai rata-rata berdasarkan K yang dipilih. Formula yang digunakan untuk menghitung rata-ratanya ditunjukkan pada persamaan (2).

$$M = 1/n \sum_{i=1}^n m_i \quad (2)$$

5. Mengisi nilai missing dengan hasil dari proses 4.

Tabel 2. Nilai Missing Value

No.	JK	IP Semester					IPK	Status
		1	2	3	4	5		
1.	L	2.55	2.52	3	2.93	2.99	3.02	TTW
2.	P	3.16	3.1	3	3.29	3.33	3.35	TW
3.	L	2.01	2.12	2	2.4	2.51	2.64	TTW
4.	L	1.62	1.7	2	2.01	2.14	2.41	TTW
5.	L	2.94	2.86	3	3.17	3.19	3.2	TTW
6.	L	2.31	2.31	2	Nan	Nan	Nan	TTW
7.	L	2.63	2.73	3	2.91	2.97	3.03	TW
8.	P	2.26	2.26	2	Nan	Nan	Nan	TTW
9.	L	2.34	2.36	3	2.59	2.69	2.76	TTW
10.	L	3.35	3.47	3	3.46	3.46	Nan	TW
11.	L	2.07	2.22	2	2.4	2.32	2.32	TTW
12.	P	3.42	2.85	3	3.13	3.5	3.15	TW
13.	L	3.17	2.9	3	3.39	3.41	3.22	TW
14.	L	Nan	2.32	2	2.5	2.5	2.42	TTW
15.	P	3.42	2.85	3	3.13	3.5	3.15	TW

Tabel 3. Hasil Penanganan Nilai Missing Value Menggunakan KNN Imputasi

No.	JK	IP Semester					IPK	Status
		1	2	3	4	5		
1.	L	2.55	2.52	3	2.93	2.99	3.02	TTW
2.	P	3.16	3.1	3	3.29	3.33	3.35	TW
3.	L	2.01	2.12	2	2.4	2.51	2.64	TTW
4.	L	1.62	1.7	2	2.01	2.14	2.41	TTW
5.	L	2.94	2.86	3	3.17	3.19	3.2	TTW
6.	L	2.31	2.31	2	2.38	2.43	2.5	TTW
7.	L	2.63	2.73	3	2.91	2.97	3.03	TW
8.	P	2.26	2.26	2	2.38	2.43	2.5	TTW
9.	L	2.34	2.36	3	2.59	2.69	2.76	TTW
10.	L	3.35	3.47	3	3.46	3.46	3.21	TW
11.	L	2.07	2.22	2	2.4	2.32	2.32	TTW
12.	P	3.42	2.85	3	3.13	3.5	3.15	TW
13.	L	3.17	2.9	3	3.39	3.41	3.22	TW
14.	L	2.07	2.32	2	2.5	2.5	2.42	TTW
15.	P	3.42	2.85	3	3.13	3.5	3.15	TW

2.2.2. Data Sampling

Data kelulusan yang digunakan penelitian ini terdapat ketidakseimbangan kelas antara kelas tepat waktu dan tidak tepat waktu. Penelitian ini menggunakan metode K-Means-Smote untuk menangani ketidakseimbangan kelasnya. K-Means-Smote merupakan hasil modifikasi dari metode sythetic minority oversampling technique (SMOTE) yang dikembangkan oleh (Douzas, Bacao and Last, 2018). Metode Smote menambahkan kelas minoritas dengan cara membangkitkan data sintesis berdasarkan k-nearest neighbor) antar kelas minoritas(Chawla, Bowyer and Hall, 2002), sedangkan pada Metode K-Means-Smote terlebih dahulu melakukan pengelompokan pada kelas mayoritas dan minoritas. Jika suatu kelompok terdapat nilai rasio ketidakseimbangan > 1, maka kelas minoritas ditambahkan dengan metode SMOTE (Douzas, Bacao and Last, 2018). Data kelulusan yang tidak seimbang dan juga data seimbang menggunakan metode K-Means-Smote ditunjukkan pada Tabel 4. Adapun tahapan-tahapan penanganan ketidakseimbangan kelas menggunakan K-Means-Smote ditunjukkan pada referensi (Douzas, Bacao and Last, 2018; Hairani, Saputro and Fadli, 2020).

Tabel 4. Perbedaan Kelas Data Kelulusan Mahasiswa

Metode Sampling	Jumlah Instance	
	Tepat Waktu	Tidak Tepat Waktu
Orginial	78	224
K-Means-Smote	224	224

2.3. Metode Klasifikasi

Metode Support Vector Machine (SVM) merupakan sebuah metode klasifikasi supervised learning. Konsep kerja dari metode SVM adalah mencari atau menemukan hyperplane yang optimal dengan memaksimalkan jarak atau margin antar kelas data dengan menggunakan persamaan (3) (Zaki, Meira Jr and Meira, 2014).

$$h(x) = w^T x + b \quad (3)$$

$$= w_1 x_1 + w_2 x_2 + \dots + w_d x_d + b$$

Persamaan (3) dapat didetailkan menjadi beberapa persamaan yang ditunjukkan pada persamaan (4) dan (5).

$$[(w^T \cdot x) + b] \geq 1 \text{ untuk } y_i = +1 \quad (4)$$

$$[(w^T \cdot x) + b] \leq -1 \text{ untuk } y_i = -1 \quad (5)$$

2.4. Pengujian Kinerja

Pengujian kinerja menggunakan tabel 29 konfusi. Tabel konfusi digunakan menghitung jumlah data yang diklasifikasikan secara benar dan salah yang ditunjukkan pada Tabel 5. Pengujian menggunakan metode *split* data dengan 80% data *training* dan 20% data *testing*

Tabel 5. Matrik Konfusi

Prediksi	Aktual	
	Tepat Waktu	Tidak Tepat Waktu
Tepat Waktu	TP	FP
Tidak Tepat Waktu	FN	TN

Formula digunakan untuk menghitung akurasi (3), sensitivitas (4), spesifisitas (5), *precision* (6), dan *f-measure* (Luque, et al., 2019).

$$\text{Akurasi} = \frac{TP + TN}{TP + FN + TN + FP} \quad (3)$$

$$\text{Sensitivitas (Recall)} = \frac{TP}{TP + FN} \quad (4)$$

$$\text{Spesifisitas} = \frac{TN}{TN + FP} \quad (5)$$

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

$$F_1 \text{ score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right) \quad (7)$$

Keterangan :

True Positif (TP) : kelas tepat waktu yang diprediksi tepat waktu.

False Positif (FP) : kelas tidak tepat waktu, tetapi diprediksi tepat waktu.

True Negatif (TN) : kelas tidak tepat waktu, tetapi diprediksi tidak tepat waktu.

False Negatif (FN) : kelas tepat waktu, tetapi diprediksi tidak tepat waktu.

3. HASIL DAN PEMBAHASAN

Penelitian ini dimulai dari tahapan pengumpulan data, pra-penelitian data, klasifikasi, dan pengujian kinerja. Data yang digunakan penelitian ini adalah data kelulusan mahasiswa universitas bumigora angkatan 2009-2012. Pra-pengolahan penelitian ini menggunakan metode KNNI untuk penanganan *missing values*, *K-Means-*

Smote digunakan untuk menangani permasalahan ketidakseimbangan kelas pada data kelulusan yang digunakan, dan metode SVM digunakan sebagai metode klasifikasi. Pengujian kinerja menggunakan tabel konfusi dengan 80% data *training* dan 20% data *testing* berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Berdasarkan pengujian metode SVM yang telah dilakukan menggunakan *split* data dengan 80% data *training* dan 20% data *testing*, didapatkan hasil berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure* yang ditunjukkan pada Tabel 6 dan Tabel 7.

Tabel 6. Matrik Konfusi Metode SVM

Prediksi	Aktual			
	Original		K-Means-Smote	
	Tepat Waktu	Tidak Tepat Waktu	Tepat Waktu	Tidak Tepat Waktu
Tepat Waktu	9	7	36	5
Tidak Tepat Waktu	5	40	2	47

$$\text{Akurasi} = \frac{36 + 47}{36 + 2 + 47 + 5} = \frac{83}{90} = 92\%$$

$$\text{Sensitivitas (Recall)} = \frac{36}{36 + 2} = \frac{36}{38} = 95\%$$

$$\text{Spesifisitas} = \frac{47}{47 + 5} = \frac{47}{52} = 90,38\%$$

$$\text{Precision} = \frac{36}{36 + 5} = \frac{36}{41} = 87,80\%$$

$$F - \text{measure} = 2 \left(\frac{87,80 * 95}{87,80 + 95} \right) = 91,26\%$$

Tabel 7. Hasil Akurasi Metode SVM

Kinerja	SVM	SVM + KNNI + K-Means-Smote (Metode usulan)
Akurasi	0.80	92%
Sensitivitas	0.64	95%
Spesifisitas	0.85	90.38%
<i>F-measure</i>	0.58	91.26%

Berdasarkan Tabel 7 menunjukkan peningkatan kinerja metode SVM dengan KNNI dan K-Means-Smote berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Hal ini terjadi, karena penambahan data kelas tepat waktu menggunakan K-Means-Smote sehingga terjadi peningkatan kerjanya berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure*. Dengan demikian, penggunaan metode KNNI dan K-Means-Smote dapat meningkatkan kinerja metode SVM (Hartini, 2017)(Caparriño, Sison and Medina, 2019)(Jordanov, Petrov and Petrozziello, 2018) (Douzas, Bacao and Last, 2018; Hairani, Saputro and Fadli, 2020). Kontribusi penelitian ini dengan penelitian sebelumnya ditunjukkan pada Tabel 8, Tabel 9, dan Tabel 10.

Tabel 8. Hasil Akurasi Metode SVM

Metode Klasifikasi	Akurasi
(Purnamasari, Rini and Sukemi, 2019) C4.5	69.79%
(Sutoyo and Almaarif, 2021) Naive Bayes	73.73%
(Nikmatun and Waspada, 2019) KNN	75.95%
(Supianto, Julisar Dwitama and Hafis, 2018) C4.5	77.01%
(Bisri and Rachmatika, 2019) GBT-Smote	80.57%
(Putri, Andreswari and Hasibuan, 2019) C4.5	82.24%
(Fadli, Zulfa and Ramadhani, 2018) SVM	90.55%
SVM - KNNI - K-Means-Smote (Metode diusulkan)	92%

Tabel 9. Hasil Sensitivitas Metode SVM

Metode Klasifikasi	Sensitivitas
(Purnamasari, Rini and Sukemi, 2019) C4.5	73.58%
(Sutoyo and Almaarif, 2021) Naive Bayes	73.6%
(Nikmatun and Waspada, 2019) KNN	77.65%
(Supianto, Julisar Dwitama and Hafis, 2018) C4.5	-
(Bisri and Rachmatika, 2019) GBT-Smote	85.68%
(Putri, Andreswari and Hasibuan, 2019) C4.5	92.05%
(Fadli, Zulfa and Ramadhani, 2018) SVM	84.75%
SVM + KNNI + K-Means-Smote (Metode diusulkan)	95%

Tabel 10. Hasil Spesifisitas Metode SVM

Metode Klasifikasi	Spesifisitas
(Purnamasari, Rini and Sukemi, 2019) C4.5	65.12%
(Sutoyo and Almaarif, 2021) Naive Bayes	70.48%
(Nikmatun and Waspada, 2019) KNN	-
(Supianto, Julisar Dwitama and Hafis, 2018) C4.5	-
(Bisri and Rachmatika, 2019) GBT-Smote	69.49%
(Putri, Andreswari and Hasibuan, 2019) C4.5	36.84%
(Fadli, Zulfa and Ramadhani, 2018) SVM	90%
SVM + KNNI + K-Means-Smote (Metode diusulkan)	90.38%

Tabel 10. Hasil F-measure Metode SVM

Metode Klasifikasi	F-measure
(Purnamasari, Rini and Sukemi, 2019) C4.5	68.49%
(Sutoyo and Almaarif, 2021) Naive Bayes	73.6%
(Nikmatun and Waspada, 2019) KNN	-
(Supianto, Julisar Dwitama and Hafis, 2018) C4.5	-
(Bisri and Rachmatika, 2019) GBT-Smote	85.79%
(Putri, Andreswari and Hasibuan, 2019) C4.5	89.50%
(Fadli, Zulfa and Ramadhani, 2018)	85.47%

SVM	
SVM + KNNI + K-Means-Smote (Metode diusulkan)	91.26%

Kinerja metode yang diusulkan secara keseluruhan memiliki kinerja paling baik dibandingkan metode yang digunakan oleh peneliti sebelumnya. Metode yang diusulkan memperoleh akurasi 92%, sensitivitas 95%, spesifisitas 90.38%, dan *f-measure* 91.26% menggunakan *split* data 80:20. Dengan demikian, penggunaan metode KNNI dan K-Means-Smote dapat meningkatkan kinerja metode SVM (Hartini, 2017)(Capariño, Sison and Medina, 2019)(Jordanov, Petrov and Petrozziello, 2018) (Douzas, Bacao and Last, 2018; Hairani, Saputro and Fadli, 2020).

4. KESIMPULAN

Metode yang diusulkan (KNNI - K-Means-Smote - SVM) menghasilkan kinerja terbaik dibandingkan penelitian yang lain berdasarkan akurasi, sensitivitas, spesifisitas, dan *f-measure* masing-masing sebesar 92%, 95%, 90, dan 92%. Terjadi peningkatan sebesar 12% untuk akurasi, 31% untuk sensitivitas, 5% untuk spesifisitas, dan 34% untuk *f-measure* karena penambahan data kelas tepat waktu menggunakan *K-Means-Smote*. Dengan demikian, metode yang diusulkan dapat digunakan untuk klasifikasi kelulusan mahasiswa.

DAFTAR PUSTAKA

- AESYI, U. S. and WARDOYO, R. (2019) 'Prediction of Length of Study of Student Applicants Using Case Based Reasoning', *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 13(1), p. 11. doi: 10.22146/ijccs.28076.
- ARI, D. and HADIANA, A. (2019) 'Computer-aided Techniques for Predicting the Failure of Student Studies Using the Decision Tree method', *IOP Conference Series: Materials Science and Engineering*, 662(2), pp. 1-9. doi: 10.1088/1757-899X/662/2/022112.
- BISRI, A. and RACHMATIKA, R. (2019) 'Integrasi Gradient Boosted Trees dengan SMOTE dan Bagging untuk Deteksi Kelulusan Mahasiswa', *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, 8(4), p. 309. doi: 10.22146/jnteti.v8i4.529.
- CAPARINO, E. T., SISON, A. M. and MEDINA, R. P. (2019) 'Application of the modified imputation method to missing data to increase classification performance', in *2019 IEEE 4th International Conference on Computer and Communication Systems, ICCCS 2019*. IEEE, pp. 134-139. doi: 10.1109/CCOMS.2019.8821632.
- CHAWLA, N., OWYER, K. W. and HALL, L. O. (2002) 'SMOTE: Synthetic Minority Over-

- sampling TEchnique', *Journal of Artificial Intelligence Research*, 16, pp. 341–378.
- 11 DOUZAS, G., BACAO, F. and LAST, F. (2018a) 'Improving imbalanced learning through a heuristic oversampling method based on k-means and SMOTE', *Informatika Sciences*. Elsevier Inc., 465, pp. 1–20. doi: 10.1016/j.ins.2018.06.056.
- 2 FADLI, A., ZIFA, M. I. and RAMADHANI, Y. (2018) 'Performance Comparison of Data Mining Classification Algorithms for Early Warning System of Students Graduation Timeliness', *Jurnal Teknologi dan Sistem Komputer*, 6(4), p. 158. doi: 10.14710/jtsiskom.6.4.2018.158-163.
- 2 HAIRANI, H., SAPUTRO, K. E. and FADLI, S. (2020) 'K-Means-Smote untuk penanganan ketidakseimbangan kelas dalam klasifikasi penyakit diabetes dengan C4.5, SVM, dan Naive Bayes', *Jurnal Teknologi dan Sistem Komputer*, 8(2). Available at: <https://jtsiskom.undip.ac.id/index.php/jtsiskom/article/view/13544>
- 17 HAN, J., KAMBER, M. and PEI, J. (2012) 'Data Mining Concepts and Techniques', in. Waltham: Morgan Kaufmann.
- 7 HARTINI, E. (2017) 'Implementation of Missing Values Handling Method for Evaluating the System/Component Maintenance Historical Data', *Jurnal Teknologi Reaktor Nuklir Tri Dasa Mega*, 19(1), p. 11. doi: 10.17146/tdm.2017.19.1.3159.
- 6 JORDANOV, I., PETROV, N. and PETROZZIELLO, A. (2018) 'Classifiers Accuracy Improvement Based on Missing Data Imputation', *Journal of Artificial Intelligence and Soft Computing Research*, 8(1), pp. 31–48. doi: 10.1515/jaiscr-2018-0022.
- 12 LUQUE, A. *et al.* (2019) 'The impact of class imbalance in classification performance metrics based on the binary confusion matrix', *Pattern Recognition*. Elsevier Ltd, 91, pp. 216–231. doi: 10.1016/j.patcog.2019.02.023.
- MAESYA, A. and HENDIYANTI, T. (2019) 'Forecasting Student Graduation with Classification and Regression Tree (CART) Algorithm', *IOP Conference Series: Materials Science and Engineering*, 621(1), pp. 1–6. doi: 10.1088/1757-899X/621/1/012005.
- 10 MINAKSHI, VOHRA, R. and GIMPY (2014) 'Missing Value Imputation in Multi Attribute Data Set', *International Journal of Computer Science and Information Technologies*, 5(4), pp. 5315–5321.
- 19 NIKMATUN, I. A. and WASPADA, I. (2019) 'Implementasi Data Mining untuk Klasifikasi Masa Studi Mahasiswa Menggunakan Algoritma K-Nearest Neighbor', *Jurnal SIMETRIS*, 10(2), pp. 421–432.
- PURNOMASARI, E., RINI, D. P. and SUKEMI (2019) 'Prediction of the Student Graduation's Level using C4.5 Decision Tree Algorithm', in *International Conference on Electrical Engineering and Computer Science (ICECOS) 2019*, pp. 192–195. doi: 10.1109/icecos47637.2019.8984493.
- 3 PUTRI, D. Y., ANDRESWARI, R. and HASIBUAN, M. A. (2019) 'Analysis of Students Graduation Target Based on Academic Data Record Using C4.5 Algorithm Case Study: Information Systems Students of Telkom University', in *2018 6th International Conference on Computer and IT Service Management, CITSM 2018*. IEEE, pp. 1–6. doi: 10.1109/CITSM.2018.8674366.
- 5 SUPIANTO, A. A., JULISAR DWITAMA, A. and HAFIS, M. (2018) 'Decision Tree Usage for Student Graduation Classification: A Comparative Case Study in Faculty of Computer Science Brawijaya University', in *3rd International Conference on Sustainable Information Engineering and Technology, SIET 2018 - Proceedings*. IEEE, pp. 308–311. doi: 10.1109/SIET.2018.8693158.
- SUTOGHO, E. and ALMAARIF, A. (2021) 'Educational Data Mining untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Naïve Bayes Classifier', *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 1(10), pp. 95–101.
- 13 ZAKI, M. J., MEIRA JR, W. and MEIRA, W. (2014) *Data mining and analysis: fundamental concepts and algorithms*. Cambridge University Press.

ORIGINALITY REPORT

19%

SIMILARITY INDEX

16%

INTERNET SOURCES

11%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1	amikhass.ac.id Internet Source	2%
2	jtsiskom.undip.ac.id Internet Source	2%
3	Yu S Bets, V S Tynchenko, V V Kukartsev, V V Bukhtoyarov, R S Dremin, I S Filimonov. "Information system to support the achievements assessment of future engineers in the learning process", IOP Conference Series: Materials Science and Engineering, 2020 Publication	1%
4	docplayer.info Internet Source	1%
5	"Predicting Student Failure in University Examination using Machine Learning Algorithms", International Journal of Innovative Technology and Exploring Engineering, 2020 Publication	1%
6	oatao.univ-toulouse.fr Internet Source	1%

7	Entin Hartini, Muhammad Subekti. "An Improvement of the Decision Making Grid Model in Failure- Based Maintenance on RSG-Gas System/Components", Journal of Physics: Conference Series, 2019 Publication	1%
8	ijadis.org Internet Source	1%
9	Evi Purnamasari, Dian Palupi Rini, Sukemi. "Prediction of the Student Graduation's Level using C4.5 Decision Tree Algorithm", 2019 International Conference on Electrical Engineering and Computer Science (ICECOS), 2019 Publication	1%
10	liacs.leidenuniv.nl Internet Source	1%
11	Weihong Han, Zizhong Huang, Shudong Li, Yan Jia. "Distribution-Sensitive Unbalanced Data Oversampling Method for Medical Diagnosis", Journal of Medical Systems, 2019 Publication	1%
12	Arjun Puri, Manoj Kumar Gupta. "Comparative Analysis of Resampling Techniques under Noisy Imbalanced Datasets", 2019 International Conference on Issues and Challenges in	1%

Intelligent Computing Techniques (ICICT), 2019

Publication

-
- | | | |
|----|--|----|
| 13 | export.arxiv.org
Internet Source | 1% |
|----|--|----|
-
- | | | |
|----|---|----|
| 14 | Gede Suwardika Suwardika, I Ketut Putu Suniantara. "ANALISIS RANDOM FOREST PADA KLASIFIKASI CART KETIDAKTEPATAN WAKTU KELULUSAN MAHASISWA UNIVERSITAS TERBUKA", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2019
Publication | 1% |
|----|---|----|
-
- | | | |
|----|--|----|
| 15 | es.scribd.com
Internet Source | 1% |
|----|--|----|
-
- | | | |
|----|---|-----|
| 16 | www.sworld.education
Internet Source | <1% |
|----|---|-----|
-
- | | | |
|----|---|-----|
| 17 | espace.curtin.edu.au
Internet Source | <1% |
|----|---|-----|
-
- | | | |
|----|---|-----|
| 18 | www.cse.ust.hk
Internet Source | <1% |
|----|---|-----|
-
- | | | |
|----|---|-----|
| 19 | jurnal.umk.ac.id
Internet Source | <1% |
|----|---|-----|
-
- | | | |
|----|---|-----|
| 20 | D Arifin, A Hadiana. "Computer-based Techniques for Predicting the Failure of Student Studies Using the Decision Tree method", IOP Conference Series: Materials Science and | <1% |
|----|---|-----|

Engineering, 2019

Publication

-
- | | | |
|----|--|-----|
| 21 | run.unl.pt
Internet Source | <1% |
| 22 | pypi.org
Internet Source | <1% |
| 23 | ejournal-umht.org
Internet Source | <1% |
| 24 | www.philippe-fournier-viger.com
Internet Source | <1% |
| 25 | fr.scribd.com
Internet Source | <1% |
| 26 | www.scribd.com
Internet Source | <1% |
| 27 | Amalia Luque, Alejandro Carrasco, Alejandro Martín, Ana de las Heras. "The impact of class imbalance in classification performance metrics based on the binary confusion matrix", Pattern Recognition, 2019
Publication | <1% |
| 28 | Devi Silvia Siltonga, Saifullah Saifullah, Rafika Dewi. "Analisis Metode Naive Bayes dalam Memprediksi Tingkat Pemahaman Mahasiswa Terhadap Mata Kuliah Berdasarkan Posisi Duduk", Prosiding Seminar Nasional Riset | <1% |

Information Science (SENARIS), 2019

Publication

29

[text-id.123dok.com](#)

Internet Source

<1%

30

[id.123dok.com](#)

Internet Source

<1%

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off