Performance Analysis of KNN and C.45 Algorithms with Kappa Measure Evaluation

Dina Mariana Siboro, Suyanto Suyanto and Amalia Amalia

June 9, 2024

Performance Analysis of KNN and C.45 Algorithms With Kappa Measure Evaluation

line 1: 1st Dina Mariana Siboro
line 2: *Magister Of Computer Science*
line 3: *University Of North Sumatra*
line 4: Medan, Indonesia
line 5: dina.siboro@students.usu.ac.id

line 1: 2nd Suyanto
line 2: *Lecturer Of Magister Of Computer Science*
line 3: *University Of North Sumatra*
line 4: Medan, Indonesia
line 5: suyanto@usu.ac.id

line 1: 3rd Amalia
line 2: *Lecturer Of Magister Of Computer Science*
line 3: *University Of North Sumatra*
line 4: Medan, Indonesia
line 5: amalia@usu.ac.id

Abstract - This study aims to develop an effective predictive model for forecasting students' graduation status at the Health Polytechnic of the Ministry of Health in Medan. Using the K-Nearest Neighbors (K-NN) method and the C4.5 algorithm, along with applying the SMOTE technique, addresses class imbalance in the dataset and enhances the accuracy of predicting students' graduation. Data were obtained from the Health Polytechnic of the Ministry of Health in Medan, including lists of students who graduated and those who did not from 2018 to 2021. The dataset comprises 3997 records with 15 features, including student ID numbers, activity statuses, Grade Point Averages (GPAs), credit unit counts, and graduation statuses. Class imbalance handling was performed using SMOTE. The KNN and C4.5 models were evaluated with three K values (1, 3, and 5) and three optimal node splitting methods. Model evaluation using the Kappa Measure indicated a high level of agreement between predicted and actual outcomes. Data normalization enhanced the consistency of analysis results, while SMOTE improved the quality of analysis and classification outcomes. Both the KNN and C4.5 models effectively predicted graduation statuses, with KNN showing improved performance with increasing K values. Cohen's Kappa values confirmed high agreement across all models. The study's conclusion underscores the effectiveness of data normalization, SMOTE, and predictive models (KNN and C4.5) in forecasting students' graduation statuses, offering valuable insights to support decision-making in higher education institutions.

Keywords— ***Prediksi kelulusan mahasiswa, K-Nearest Neighbors (K-NN), C.45, SMOTE, Kappa Measure***

I. INTRODUCTION

As the main entity in the education process, universities play a crucial role as drivers of social change, with a focus on developing the potential and skills of individuals, which in turn creates valuable contributors to the progress of the nation and state (Sedyati, 2022). Continuous evaluation of academic performance is necessary in universities to maintain a balance between the number of new student admissions and graduations each year, to avoid significant imbalances between the two numbers (Gunawan et al., 2021). In research on predicting student graduation rates, datasets with imbalance between the target classes, Pass and Fail, are often found, leading to biases in the training and testing processes of models (Amri et al., 2023; Rismayati et al., 2022; Zubair, 2022).

From the background presented, this study aims to develop an effective predictive method for predicting student class attendance status at Kemenkes Poltekkes Medan. By utilizing a combination of the K-Nearest Neighbors (K-NN) method and the C4.5 algorithm, and applying the SMOTE technique, this research seeks to address the main challenges in dealing with class imbalance in datasets. It is hoped that this research will provide a significant contribution to improving the accuracy of predicting student graduation rates and supporting better decision-making in higher education institutions. This research can also expand the understanding of the use of evaluation metrics such as kappa measure in cases of datasets with class imbalance. This can help researchers and practitioners choose the right evaluation metrics to evaluate model performance in the face of class imbalance commonly encountered in machine learning practice.

In research on predicting student graduation rates, datasets with imbalance between the Pass and Fail target classes are often found, leading to biases in the training and testing processes of models (Amri et al., 2023; Rismayati et al., 2022; Zubair, 2022). To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) method can be used as a commonly applied solution (Sutoyo & Fadlurrahman, 2020). SMOTE works by synthesizing new samples in the minority class based on existing samples, thereby creating a more balanced class distribution (Arisandi, 2023). Kappa measure is one of the important evaluation metrics because it provides a deep understanding of how well the model classifies data, especially in the context of class imbalance (Yilmaz & Demirhan, 2023). Although kappa measure has often been used in research on datasets with balanced distributions (Buntoro et al., 2021; Cano & Krawczyk, 2020; Kolesnyk & Khairova, 2022), it is rarely applied to imbalanced datasets, thus the authors are interested in investigating the suitability of using this evaluation in datasets with imbalanced target classes.

II. RESEARCH METHODOLOGY

The research methodology employed includes several key steps, outlined as follows:

1. Data Collection
2. Preprocessing Data
   - Data Cleaning
   - Feature Selection
   - Data Encoding
3. Handling Data Imbalance

- ➢ Data Preparation
- ➢ Application of SMOTE
- ➢ Evaluation

4. Modeling
5. Modeling Evaluation
   - ➢ Model Testing
   - ➢ Calculation of Kappa Measure

A. Data Collection

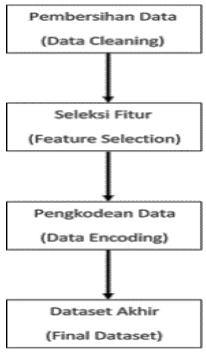
This research utilizes primary data obtained from Kemenkes Poltekkes Medan, consisting of a list of students who graduated and did not graduate during the period from 2018 to 2021. The dataset used in this study comprises 3997 records, consisting of 15 features, detailed as seen in the table below

Tabel 1. : Research Dataset Details

Fitur	Type Of Data	Elaboration
NIM	Kategorikal	Numeric, used as categories to identify individual students
Status 1	Kategorikal	Student activity status in semester 1. It can be "Active", "Inactive", "Waiting for Competency Test", or "Graduated".
IPS 1	Numerikal (Continuous)	Grade Point Average (GPA) for semester 1. It is a continuous numerical value ranging from 0 to 4.
SKS 1	Numerikal (Discrete)	Number of Credit Units in Semester 1. It is a discrete numerical value represented by an integer.
Status 2	Kategorikal	Student activity status in semester 2. It can be "Active", "Inactive", "Waiting for Competency Test", or "Graduated".
IPS 2	Numerikal (Continuous)	Grade Point Average (GPA) for semester 2. It is a continuous numerical value ranging from 0 to 4.
SKS 2	Numerikal (Discrete)	Number of Credit Units in Semester 2. It is a discrete numerical value represented by an integer.
Status 3	Kategorikal	Student activity status in semester 3. It can be "Active", "Inactive", "Waiting for Competency Test", or "Graduated".
IPS 3	Numerikal (Continuous)	Grade Point Average (GPA) for semester 3. It is a continuous numerical value ranging from 0 to 4.
SKS 3	Numerikal (Discrete)	Number of Credit Units in Semester 3. It is a discrete numerical value represented by an integer.
Status 4	Kategorikal	Student activity status in semester 4. It can be "Active", "Inactive", "Waiting for Competency Test", or "Graduated".
IPS 4	Numerikal (Continuous)	Grade Point Average (GPA) for semester 4. It is a continuous numerical value ranging from 0 to 4.
SKS 4	Numerikal (Discrete)	Number of Credit Units in Semester 4. It is a discrete numerical value represented by an integer.
IPK	Numerikal (Continuous)	Cumulative Grade Point Average (CGPA) of the student. It is a continuous numerical value ranging from 0 to 4.
Status	Kategorikal	Student graduation status at the end of their study period. It can be "Graduated" or "Not Graduated".

B. Preprocessing Data

Data preprocessing is conducted to prepare cleaner data, aiding in subsequent prediction processes. In this study, three stages of data preprocessing are performed, as depicted in the figure below.



Gambar 1. Step Of Preprocessing Data

Pada Preprocessing Data dilakukan tahapan yaitu :

1. Data Cleaning is performed to clean the dataset from incomplete data. In the initial dataset, there are several missing values in the features Status 1, IPS 1, SKS 1, Status 2, IPS 2, SKS 2, Status 3, IPS 3, SKS 3, Status 4, IPS 4, SKS 4.
2. Feature Selection: After data cleaning, feature selection is performed. Out of 15 features, 14 features are selected, namely: Status 1, IPS 1, SKS 1, Status 2, IPS 2, SKS 2, Status 3, IPS 3, SKS 3, Status 4, IPS 4, SKS 4, IPK, Status. Out of these 14 features, 13 features (features 1 to 13) are utilized as predictive features, while the 14th feature (Status) is employed as the predictive targetPengkodean
3. Data Encoding: This involves converting data from string format to numeric format.

C. Handling Data Imbalance

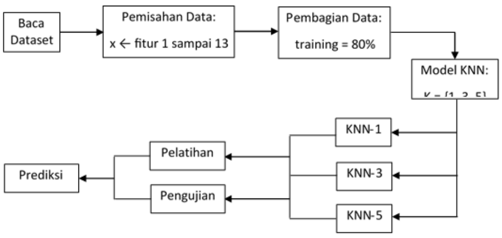
From the utilized dataset, it is evident that there is an imbalance within the predictive target class, where the number of data points labeled as "Pass" is significantly higher than those labeled as "Fail." To address this issue, the Synthetic Minority Over-sampling Technique (SMOTE) is employed. SMOTE is utilized to increase the number of samples in the minority class, in this case, "Fail," thereby creating a balance between the two classes. The steps undertaken to handle data imbalance are as follows:

1. **Data Preparation** involves separating features 1 – 13 as the dataset, and identifying the imbalance between the target categories "Pass" and "Withdrawn.".
2. **Application of SMOTE** entails generating new synthetic samples based on existing samples in the minority category using interpolation techniques.
3. **Evaluation** entails assessing the results of applying SMOTE after examining the dataset to ensure that the data imbalance between target categories has been reduced/eliminated.

D. Pemodelan

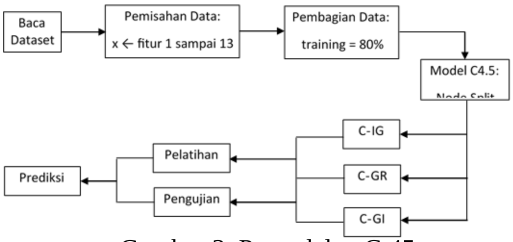
The modeling conducted in this research was done gradually, which means it was carried out in stages. :

1. K-Nearest Neighbor (KNN) modeling was conducted, with the values of K set at 1, 3, and 5. The prediction steps are illustrated in the figure below:



Gambar 2. Pemodelan KNN

2. C4.5 algorithm modeling, which employs three variations for determining optimal splits at each node in the decision tree: information gain, gain ratio, and Gini index.



Gambar 3. Pemodelan C.45

Model Evaluation In this modeling process, after the models are built, they are evaluated using the Kappa Measure to assess the performance of each constructed prediction model. The steps for evaluating with the Kappa Measure are as follows: Testing the KNN Model, Testing the C4.5 Model, and finally, evaluating with the Calculation of Kappa Measure. For the Kappa Measure calculation process, the following are the steps:

a. Calculate the contingency matrix (confusion matrix) between the predicted results and the actual results.
b. Use the contingency matrix to calculate the Kappa Measure value using the formula:

$$Kappa = \frac{P(A)-P(E)}{1-P(E)}$$

In which:

$P(A)$ is the percentage agreement between the predicted results and the actual results, P(E) is the percentage of agreement expected by chance.

.

Tabel 2. Sampel Confusion Matrix

	Passed	Unpass
Passed	514	8
Unpass	4	359

For example, if one of the models produces a confusion matrix as shown in Table 2, then the calculation steps for the kappa measure value from that table are as follows:

➢ Menghitung nilai P(A)
- P(A) adalah nilai yang diperoleh dengan menggunakan formula berikut:
- $P(A) = \frac{TP+TN}{N} = \frac{514+359}{885} = 0.986$

➢ Calculate the value of P(E)
- P(A) It is the value obtained using the following formula:
- $P(E) = \frac{(TP+TN)*TP}{N} + \frac{(FP+FN)*TN}{N}$
- $= \frac{(514+359)*514}{885} + \frac{(8+4)*359}{885}$
- $= 508.34 + 4.87$
- $= 0.579$

➢ Calculate the value of K
- The value of K is obtained using the previously mentioned Kappa Measure formula, resulting in the following outcome:
- $Kappa = \frac{0.986 - 0.579}{1-0.579} = \frac{0.407}{0.421} = 0.966$

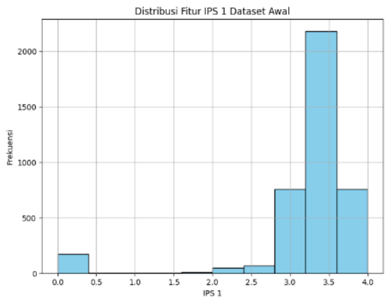
c. Evaluate the results by comparing the Kappa Measure values of the KNN and C4.5 models to determine their relative performance evaluation. The higher the Kappa Measure value, the better the model is at making predictions consistent with the actual results.

III. THE RESULT AND CONCLUSION

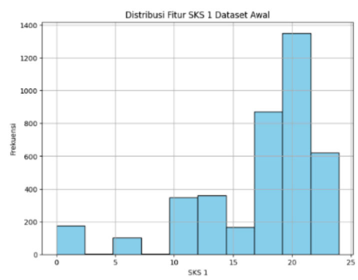
This section discusses the research findings obtained from the model training and testing processes, as well as the analysis of these results.

A. *The Normalized Result*

In normalization, data on student lecture activities were processed, totaling 3997 records. The initial distribution data on numerical distribution can be seen in the Initial Feature Distribution figure below.:



Gambar 4. Distribusi Data Pada Fitur IPS 1



Gambar 5. Distribusi Data Pada Fitur SKS 1

From the above figure, statistical distribution data on the dataset were obtained, including the count, mean, standard deviation (std), minimum value (min), first quartile (25%), second quartile (50%), and third quartile (75%), as detailed in the table below.

Tabel 3. Statistik Distribusi Data Pada Dataset

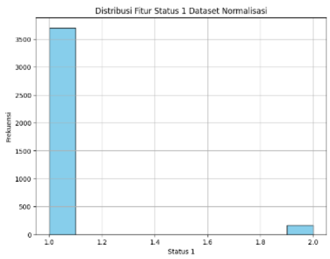
Fitur	Statistik							
	count	mean	std	min	25%	50%	75%	max
IPS 1	3994.00	3.22	0.75	0.00	3.18	3.36	3.53	4.00
SKS 1	3994.00	17.35	5.46	0.00	15.00	19.00	21.00	24.00
IPS 2	3945.00	3.02	1.15	0.00	3.25	3.47	3.61	4.00
SKS 2	3945.00	15.09	6.71	0.00	11.00	18.00	20.00	24.00
IPS 3	3898.00	3.04	1.23	0.00	3.16	3.55	3.69	4.00
SKS 3	3898.00	13.52	6.53	0.00	11.00	14.00	20.00	22.00
IPS 4	3868.00	2.64	1.47	0.00	2.92	3.37	3.51	4.00
SKS 4	3868.00	13.71	8.56	0.00	4.00	20.00	20.00	24.00
IPK	3868.00	2.88	1.19	0.00	3.00	3.38	3.54	4.00

From Table 4.1 above, it can be observed that the average Semester Grade Point Averages (IPS) vary between 2.64 to 3.22. This indicates that overall, students tend to achieve good academic performance. Additionally, quartile analysis shows a consistent improvement from one semester to the next, with the 75th quartile of each semester indicating an increase in academic achievement. Regarding the number of Credit Semester Units (SKS) taken, the data shows variation from 13.52 to 17.35 for each semester, with the maximum number of SKS taken in each semester reaching 24 SKS.
This dataset is then normalized to convert non-numeric data into numeric data, facilitating further classification processes. In this study, with rules as seen in the table below

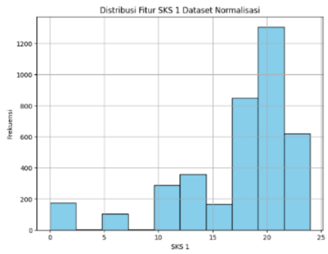
Tabel 4. Normalization Rules

Variabel	Categories	Value Of Normalized
Status 1	Active	1
	Non-Active	2
	Waiting for Competency Test	1
	Paid leave	2
	Passed	2
Status 2	Passed	1
	Non-Active	2
	Waiting for Competency Test	1
	Paid leave	2
	Passed	2
Status 3	Passed	1
	Non-Active	2
	Waiting for Competency Test	1
	Paid Leave	2
	Passed	2
Status 4	Active	1
	Non-Active	2
	Waiting for Competency Test	1
	Paid Leave	2
	Passed	2
Status	Passed	1
	Resigned	-1

The normalized dataset of student lecture activities in this study is presented in the figure below:



Gambar 6. Distribusi Data Normalisasi Pada Fitur Status 1



Gambar 7. Distribusi Normalisasi Pada Fitur SKS 1

The result of the normalization distribution yields statistical data on the normalized dataset, including the count, mean, standard deviation (std), minimum value (min), quartile 1 (25%), quartile 2 (50%), and quartile 3 (75%), as detailed in the table below.

Tabel 5. Statistik Distribusi Data Normalisasi

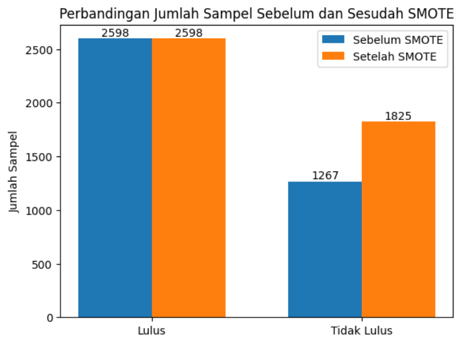
Fitur	Statistik							
	count	mean	std	min	25%	50%	75%	max
IPS 1	3865.00	3.21	0.76	0.00	3.17	3.36	3.53	4.00
SKS 1	3865.00	17.39	5.48	0.00	15.00	19.00	21.00	24.00
Status 1	3865.00	1.04	0.20	1.00	1.00	1.00	1.00	2.00
IPS 2	3865.00	3.01	1.16	0.00	3.24	3.47	3.62	4.00
SKS 2	3865.00	15.12	6.75	0.00	11.00	18.00	20.00	24.00
Status 2	3865.00	1.1	0.30	1.0	1.0	1.0	1.0	2.0
IPS 3	3865.00	3.04	1.23	0.00	3.16	3.55	3.69	4.00
SKS 3	3865.00	13.54	6.55	0.00	11.00	14.00	20.00	22.00
Status 3	3865.00	1.09	0.28	1.00	1.00	1.00	1.00	2.00
IPS 4	3865.00	2.64	1.47	0.00	2.92	3.37	3.51	4.00
SKS 4	3865.00	13.71	8.55	0.00	4.00	20.00	20.00	24.00
Status 4	3865.00	1.18	0.38	1.00	1.00	1.00	1.00	2.00

IPK	3865.00	2.88	1.19	0.00	3.00	3.38	3.54	4.00
Status	3865.00	0.34	0.94	-1.00	-1.00	1.00	1.00	1.00

From the table above, it can be observed that the relatively low standard deviation values for variables such as IPS indicate consistency in academic performance, while higher standard deviations for variables such as SKS indicate greater variation in student study loads. The minimum (min) and maximum (max) values provide information about the range of observed values, while quartiles (25%, 50%, and 75%) help describe the data distribution by dividing it into four equally sized parts

B. *The Result Of SMOTE*

Normalization conducted on the total of 3865 data reveals that there are 2598/67.218% data points categorized as "Pass" and 1267/32.782% data points categorized as "Fail". This indicates a difference of 34.436% between the two categories. This clearly indicates class imbalance, necessitating the SMOTE process to be performed.



Gambar 8. Distribusi Target Sebelum dan Sesudah SMOTE

From the results of the SMOTE process, it is observed that before SMOTE, the number of data points for the 'Pass' category is 2598, while the number of data points for the 'Fail' category is 1267. After the SMOTE process, the number of data points for the 'Pass' category remains the same, i.e., 2598, while the number of data points for the 'Fail' category increases to 1825. This indicates that the SMOTE process has successfully oversampled the minority category ('Fail'), resulting in a balanced number of data points between the two categories.

To assess whether these results meet the criteria for class balance with a maximum difference of 20%, the percentage difference between the number of data points for the 'Pass' and 'Fail' categories after SMOTE can be calculated, with the details as follows :

$$Selisih\ Persentase = \left(\frac{Data\ 'Lulus'\ Setelah\ SMOTE}{Total\ Data\ Setelah\ SMOTE} - \frac{Data\ 'Tidak\ Lulus'\ Setelah\ SMOTE}{Total\ Data\ Setelah\ SMOTE}\right) X100\%$$

$$Selisih\ Persentase = \left(\frac{2598}{4423} - \frac{1825}{4423}\right) * 100\% = 0.1747 * 100\% = 17.47\%$$

C. *Prediction Results of the K-NN Model*

Based on the K-NN model designed using variations of K values, namely K = 1, K = 3, and K = 5, the prediction results are obtained, each of which is shown in Table 6 to Table 8 in the form of a confusion matrix.

Tabel 6 Confusion Matrix Model KNN-1

	Not Passed	Passed
Not Passed	359	4
Passed	8	514

Tabel 7 Confusion Matrix Model KNN-3

	Not Passed	Passed
Not Passed	360	3
Passed	6	516

Tabel 8 Confusion Matrix Model KNN-5

	Not Passed	Pass
Not Passed	362	1
Pass	6	516

D. *Prediction Results of the C.45*

Based on the C4.5 model designed using variations of attribute selection metrics such as information gain, gain ratio, and Gini index, the prediction results are obtained, each of which is shown in Table 9 to Table 12 in the form of a confusion matrix.

Tabel 9 Confusion Matrix Model C4.5 Information Gain

	Not Passed	Passed
Not Passed	362	1
Passed	7	515

Tabel 10 Confusion Matrix Model C4.5 Gain Ratio

	Not Passed	Passed
Not Passed	362	1
Passed	7	515

Tabel 11 Confusion Matrix Model C4.5 Gini Index

	Not Passed	Passed
Not Passed	362	1
Passed	5	517

Tabel 12 Comparation Confusion Matrix Of The Model C4.5

Model	True Positive (TP)	False Positive (FP)	True Negative (TN)	False Negative (FN)
C4.5 IG	515	1	7	362
C4.5 GR	515	1	7	362

C4.5 GI	517	1	5	362

E. Hasil Kappa Measure

Kappa measure digunakan untuk mengevaluasi kinerja masing-masing model, sehingga dapat ditentukan model yang paling optimal dalam memprediksi kelulusan mahasiswa berdasarkan data aktivitas perkuliahan mahasiswa.

Hasil perhitungan *cohen's kappa* masing-masing model ini kemudian dirangkum dalam Tabel 4.12 untuk mempermudah proses analisis hasil berikutnya.

Tabel 13 Nilai Cohen's Kappa Seluruh Model

Model	Cohen's Kappa
KNN-1	0.986
KNN-3	0.99
KNN-5	0.992
C4.5 Information Gain	0.991
C4.5 Gain Ratio	0.991
C4.5 Gini Index	0.994

IV. KESIMPULAN DAN SARAN

Normalization of data successfully enhances the consistency and reliability of analysis results, facilitating more accurate and comprehensive interpretation of the data. The SMOTE process proves effective in addressing class imbalances within the dataset, thereby improving the quality of analysis and producing reliable classification outcomes. SMOTE serves as a crucial step in ensuring that the classification model built from the dataset can recognize and predict minority categories with greater accuracy.

Both the K-Nearest Neighbors (K-NN) and C4.5 models demonstrate the ability to predict student graduation with a fairly high level of accuracy. In the K-NN model, there is an improvement in performance from KNN-1 to KNN-5, indicating that increasing the number of neighbors in the K-NN algorithm enhances consistency and agreement between predictions and actual outcomes. Meanwhile, in the C4.5 model, there is no consistently superior model in terms of agreement in predicting student graduation, and all three can be considered effective in aiding decision-making related to student graduation.

Overall, Cohen's Kappa values indicate a very high level of agreement between prediction results and actual outcomes for all evaluated models and algorithms. Although there are slight differences in performance between these models, their overall performance appears similar and effective. This suggests that both K-NN and C4.5 models can be used as effective tools in assisting decision-making related to student graduation, with the appropriate selection depending on specific needs within the application context.

REFERENCES

[1] Meilani, B. D., Wahyudiana, S., Putri, A. Y. P., & Pakarbudi, A. (2019). Klasifikasi Identifikasi Faktor Penyebab Ketidaktepatan Masa Lulus Mahasiswa dengan Metode Naïve Bayes Classifier. *Seminar Nasional Sains Dan Teknologi Terapan*, 297–302. https://ejournal.itats.ac.id/sntekpan/article/view/586

[2] Kinasih, H. W., Prajanto, A., & Sartika, M. (2022). Lulus Tepat Waktu Mahasiswa Akuntansi: Faktor Apa Saja Yang Mempengaruhi. *Jurnal Riset Akuntansi (JUARA)*, *12*(2), 285–296. https://doi.org/10.36733/juara.v12i2.1700

[3] Hidayati, N., & Hermawan, A. (2021). K-Nearest Neighbor (K-NN) algorithm with Euclidean and Manhattan in classification of student graduation. *Journal of Engineering and Applied Technology*, *2*(2), 86–91. https://doi.org/10.21831/jeatech.v2i2.42777

[4] Gunawan, M., Zarlis, M., & Roslina, R. (2021). Comparative Analysis of Naïve Bayes and K-Nearest Neighbor Algorithms to Predict Student Graduation on Time. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, *5*(2), 513. https://doi.org/10.30865/mib.v5i2.2925

[5] Hakim, L. A. R., Rizal, A. A., & Ratnasari, D. (2019). Aplikasi Prediksi Kelulusan Mahasiswa Berbasis K-Nearest Neighbor (K-NN). *JTIM : Jurnal Teknologi Informasi Dan Multimedia*, *1*(1), 30–36. https://doi.org/10.35746/jtim.v1i1.11

[6] Yuliansyah, H., Imaniati, R. A. P., Wirasto, A., & Wibowo, M. (2021). Predicting Students Graduate on Time Using C4.5 Algorithm. *Journal of Information Systems Engineering and Business Intelligence*, *7*(1), 67. https://doi.org/10.20473/jisebi.7.1.67-73

[7] Etriyanti, E., Syamsuar, D., & Kunang, N. (2020). Implementasi Data Mining Menggunakan Algoritme Naive Bayes Classifier dan C4.5 untuk Memprediksi Kelulusan Mahasiswa. *Telematika*, *13*(1), 56–67. https://doi.org/10.35671/telematika.v13i1.881

[8] Hasibuan, T. H., & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *Skanika*, *6*(1), 61–74. https://doi.org/10.36080/skanika.v6i1.2976

[9] Amri, Z., Kusrini, K., & Kusnawi, K. (2023). Prediksi Tingkat Kelulusan Mahasiswa menggunakan Algoritma Naïve Bayes, Decision Tree, ANN, KNN, dan SVM. *Edumatic: Jurnal Pendidikan Informatika*, *7*(2), 187–196. https://doi.org/10.29408/edumatic.v7i2.18620

[10] Rismayati, R., Ismarmiaty, I., & Hidayat, S. (2022). Ensemble Implementation for Predicting Student Graduation with Classification Algorithm. *International Journal of Engineering and Computer Science Applications (IJECSA)*, *1*(1), 35–42. https://doi.org/10.30812/ijecsa.v1i1.1805

[11] Purwanto, E., Kusrini, K., & Sudarmawan, S. (2019). Prediksi Kelulusan Tepat Waktu Menggunakan Metode C4.5 DAN K-NN (Studi Kasus : Mahasiswa Program Studi S1 Ilmu Farmasi, Fakultas Farmasi, Universitas Muhammadiyah Purwokerto). *Techno (Jurnal Fakultas Teknik, Universitas Muhammadiyah Purwokerto)*, *20*(2), 131. https://doi.org/10.30595/techno.v20i2.5160

[12] Rau, G., & Shih, Y. S. (2021). Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes*, *53*(July), 101026. https://doi.org/10.1016/j.jeap.2021.101026

[13] Roihan, A., Sunarya, P. A., & Rafika, A. S. (2020). Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. *IJCIT (Indonesian Journal on Computer and Information Technology)*, *5*(1), 75–82. https://doi.org/10.31294/ijcit.v5i1.7951

[14] Rukmini, E., Angelina, H., & Anggreni, V. C. (2023). Indonesia higher education's online learning during the pandemic state. *International Journal of Evaluation and Research in Education*, *12*(4), 2286–2301. https://doi.org/10.11591/ijere.v12i4.25103

[15] Samsuden, M. A., Diah, N. M., & Rahman, N. A. (2019). A Review Paper on Implementing Reinforcement Learning Technique in Optimising Games Performance. *2019 IEEE 9th International Conference on System Engineering and Technology (ICSET)*, 258–263. https://doi.org/10.1109/ICSEngT.2019.8906400

.