Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (226-235)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

JURNAL TEKNIK INFORMATIKA

Homepage : http://journal.uinjkt.ac.id/index.php/ti

Addressing Class Imbalance in Machine Learning for Predicting On-Time Student Graduation at The Islamic University of Riau

Akmar Efendi[1*], Sarjon Defit[2]

[1]Department of Informatics Engineering, Faculty of Engineering, Universitas Islam Riau
[2]Department of Information Technology, Faculty of Computer Science, Universitas Putra Indonesia YPTK
[1]Jl. Kaharuddin Nasution, Pekanbaru, Indonesia
[2]Jl. Lubuk Begalung, Padang, Indonesia

Article:
Accepted: July 20, 2025
Revised: March 17, 2025
Issued: October 30, 2025

***Correspondence Address:**
akmarefendi@eng.uir.ac.id

ABSTRACT

Timely graduation is an important indicator of academic performance in higher education. However, many students still fail to graduate on time, prompting the need for predictive models to support academic decision-making. This study aims to analyze the impact of class imbalance on machine learning algorithm performance in predicting student graduation at the Islamic University of Riau. Data were obtained through questionnaires and labeled into “graduated on time” and “not on time” classes, which were initially imbalanced. The Synthetic Minority Over-Sampling Technique (SMOTE) was applied during preprocessing to balance the dataset. Four machine learning algorithms were compared: Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, and Support Vector Machine. The evaluation was conducted with and without SMOTE, using accuracy, precision, recall, F1-score, and confusion matrix. Results showed significant performance improvements after applying SMOTE, with all models achieving around 99% accuracy. SVM achieved the most stable results across both conditions. The study highlights the effectiveness of SMOTE in improving classification fairness and reliability, especially in datasets with class imbalance. This work may assist universities in early intervention for students at risk of late graduation.

Keywords : *student graduation; SMOTE; classification; machine learning.*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (226-235)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

1. INTRODUCTION

Graduation timeliness remains a critical benchmark in evaluating the effectiveness of higher education institutions [1]. Universities are expected to ensure that most students complete their studies within the designated period [2]. Delays in graduation can hinder institutional accreditation processes, disrupt academic planning, and reduce the timely availability of graduates to meet labor market demands [3]. Consequently, early detection of potential graduation delays is essential for implementing proactive interventions.

Recent technological advancements have enabled the integration of machine learning methods into education, including the prediction of student graduation outcomes [4]. Algorithms such as Decision Tree, Naive Bayes, K-Nearest Neighbors (K-NN), and Support Vector Machine (SVM) have been widely adopted due to their capacity to identify patterns in historical student data [5], [6], [7], [8]. However, many educational datasets suffer from class imbalance—where the number of on-time graduates significantly outweighs those with delayed graduation [9]. This imbalance skews the performance of classifiers, often resulting in biased predictions toward the majority class and overlooking those at risk of not graduating on time [10].

Several previous studies have implemented machine learning to predict graduation outcomes. For instance, study [11] utilized Decision Tree and Naïve Bayes based on academic performance but did not consider the class distribution. Study [12] xplored various algorithms, including DT, SVM, Random Forest, ANN, KNN, and Logistic Regression, yet performance for the minority class was suboptimal. Study [13] also reported misclassification of minority class despite using Random Forest. Meanwhile, study [14] showed that applying SMOTE improved classification accuracy by addressing data imbalance effectively.

Unlike earlier studies, this research focuses on evaluating the influence of class imbalance on the performance of several machine learning algorithms using a dataset specific to the Islamic University of Riau. What sets this work apart is the emphasis on fair classification performance through the application of SMOTE and the investigation of which algorithm benefits most from such balancing. Furthermore, this study uses a real dataset gathered from 120 respondents (students and alumni) through structured questionnaires, providing context-rich and authentic information for predictive modeling.

This study also aims to go beyond simple classification by analyzing whether particular features—such as GPA, family support, or study habits—correlate with delayed graduation. This deeper insight offers new contributions not only in prediction performance but also in guiding strategic interventions. By implementing preprocessing stages such as handling missing values, normalization, encoding, and SMOTE-based balancing, the data is modeled using four classifiers: Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, and Support Vector Machine. Each algorithm is selected for its unique strength: interpretability [15], simplicity [16], instance-based learning [17], and robustness in high-dimensional spaces [18], respectively.

The outcome of this study is expected to inform the development of equitable and reliable graduation prediction tools that can assist universities in early intervention for students at risk of not graduating on time, ultimately supporting institutional efforts to enhance student success rates.

Previous studies have shown various approaches to predicting students' on-time graduation. Some studies utilized the Support Vector Machine (SVM) algorithm, achieving an accuracy of around 71.20%, but faced challenges with imbalanced data, particularly in predicting students who did not graduate on time. Meanwhile, other studies used Random Forest and Neural Networks on large datasets (132,734 students), achieving 76% accuracy and 79% AUC, with GPA identified as the most influential factor [19]. Research on student loyalty using Random Forest also showed strong results (90.9% accuracy) based on questionnaire data related to campus service quality perception [20].

Other approaches, such as Decision Tree and Naive Bayes, also delivered competitive results. A study using Decision Tree to predict graduation at Dian Nuswantoro University achieved high accuracy, reaching 91% [21]. Naive Bayes, using 14 academic and demographic variables, recorded 85% accuracy

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (226-235)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

with a high f1-score for on-time graduation predictions but lower performance in identifying delays [22]. Several studies combined AdaBoost with Decision Tree, achieving f1-scores up to 0.82 [4]. Even in imbalanced data scenarios, Decision Tree enhanced with SMOTE demonstrated excellent performance (96.67% accuracy) [23]. Additionally, research using an educational data mining approach with Decision Tree for core computing courses achieved 88.9% accuracy. Historical student data across multiple cohorts have also been successfully analyzed using various classification algorithms [24].

Overall, these findings indicate that predictive models based on classical machine learning algorithms such as Decision Tree, Naive Bayes, K-Nearest Neighbors, and Support Vector Machine can play an important role in improving academic management and supporting early intervention for at-risk students, although improvements are still needed, particularly in handling imbalanced datasets involving delayed graduation cases.

This study aims to predict students' on-time graduation using a machine learning approach. Data were collected through questionnaires distributed to students and alumni of the Islamic University of Riau. After the labeling process, data were preprocessed through stages that included handling missing values, normalization, and label encoding. To address class imbalance between students who graduated on time and those who did not, the SMOTE (Synthetic Minority Over-sampling Technique) method was applied.

The model was then built using several classical machine learning algorithms, namely Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, and Support Vector Machine (SVM). Each model was evaluated in classifying students into two categories: Graduated on Time and Did Not Graduate On Time. The classification results are expected to support academic policy-making and provide early intervention for students at risk of delayed graduation.

2. METHODS

This study follows a series of systematic stages to investigate how class imbalance affects the performance of machine learning algorithms in predicting student graduation outcomes. The methodological flow is presented in Figure 1.

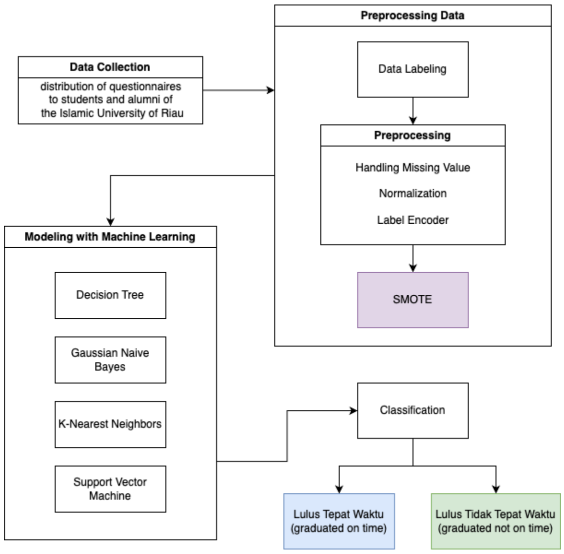


Figure 1. *Research flow diagram*

The research process consists of the following stages:

2.1. Dataset

The dataset used in this research was obtained from questionnaires distributed to students and alumni of the Islamic University of Riau. The questionnaire included parameters such as GPA, number of credits taken per semester, semester of enrollment, organizational involvement, scholarship status, financial condition, and motivation to study. These variables are commonly associated with student academic progression and timely graduation.

The initial dataset consisted of 275 records, which were categorized into two classes: "Graduated On Time" (68 records) and "Not Graduated On Time" (207 records), showing a class imbalance ratio of approximately 1:3. This imbalance potentially biases models towards the majority class. The distribution is illustrated in Figure 2.

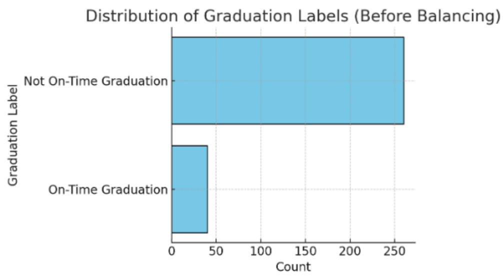


Figure 2. *Dataset before balancing*

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (226-235)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

2.2. Preprocessing

The preprocessing stage aimed to transform raw questionnaire responses into a structured format suitable for machine learning modeling. The steps included:

a. Handling Missing Values: Missing data were handled using mean imputation for numerical features and mode imputation for categorical features [25].
b. Encoding: Categorical attributes such as gender, financial aid, and organizational activity were encoded using label encoding [26].
c. [27].
d. Balancing Data: The Synthetic Minority Over-sampling Technique (SMOTE) was applied to generate synthetic samples in the minority class, resulting in a balanced dataset of 414 records (207 per class) [28]. This balanced distribution is shown in Figure 3.

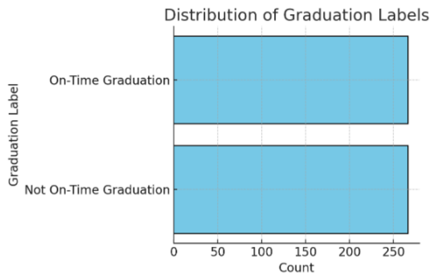


Figure 3. *Dataset after balancing*

2.3. Modeling

The classification models were developed using four machine learning algorithms selected for their interpretability, classification characteristics, and common use in educational data mining:

a. Decision Tree (DT): A rule-based learner that splits the dataset based on information gain or Gini index. Suitable for mixed data types and interpretable models [23].
b. Gaussian Naive Bayes (GNB): A probabilistic classifier assuming a normal distribution of features, known for its computational efficiency [29].
c. K-Nearest Neighbors (K-NN): A non-parametric method classifying based on the majority vote among k nearest neighbors. In this study, k = 5 was chosen based on empirical testing [30].
d. Support Vector Machine (SVM): Utilizes a linear kernel to construct an optimal hyperplane that maximizes the margin between classes. This model was selected for its performance in high-dimensional datasets [31].

Each model was implemented using the Scikit-learn library in Python. The dataset was split using an 80:20 train-test split, stratified to maintain the proportion of each class in both sets.

2.4. d. Evaluation Model

Each model was evaluated using a test dataset that had been separated from the training set to objectively measure performance on unseen data. Evaluation was performed using five key metrics: accuracy, precision, recall, F1-score, and the confusion matrix [32], [33].

a. Accuracy measures the proportion of correct predictions out of all predictions. However, since the dataset initially suffered from class imbalance, accuracy alone is not a reliable metric especially for minority class performance.
b. Precision indicates the proportion of correct positive predictions (i.e., how many predicted “on-time graduates” were actually on time).
c. Recall reflects the model's ability to capture all true instances of the positive class.
d. F1-score is the harmonic mean of precision and recall and is crucial when dealing with imbalanced datasets because it balances both metrics.

The confusion matrix was also used as a visual tool to show the number of correct and incorrect predictions per class, helping identify whether the model was biased toward the majority class. Evaluation was conducted in two scenarios:

a. Using the original (imbalanced) dataset
b. Using the balanced dataset processed with SMOTE

This dual-scenario evaluation served as the basis for analyzing the impact of class imbalance on the effectiveness of each machine learning algorithm. It also helped to assess how class distribution influenced the model’s ability to produce accurate and fair classifications in the context of predicting student graduation outcomes.

Jurnal Teknik Informatika Vol. 18 No. 2, October 2025 (226-235)
ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

2.5. Result Analysis

The analysis compared model performance under both dataset conditions (imbalanced vs balanced). The focus was on how accuracy and other metrics changed with the application of SMOTE and how each algorithm performed in identifying minority class samples. The most robust model was determined based not only on overall accuracy but also on balanced performance across all metrics, particularly F1-score for the minority class. This approach highlights how machine learning models can be adjusted for fairer and more reliable predictions in educational settings, with practical implications for early intervention programs targeting students at risk of delayed graduation.

3. RESULTS AND DISCUSSION

The initial stage of this study was testing the dataset before applying any balancing techniques, as shown in Table 1.

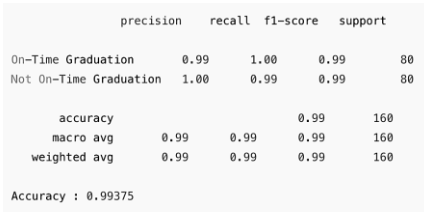
Table 1. *Testing without SMOTE*

Algorithm	Acc	Precision	Recall	F1-Score
DT	89.87	85.67	87.34	88.70
GNB	80.82	76.89	60.45	75.43
KNN	85.46	80.54	76.46	77.67
SVM	92.43	80.23	88.67	91.99

The results from testing without using balancing techniques such as SMOTE indicate that although accuracy values appear high, not all algorithms provide balanced performance. The Support Vector Machine (SVM) algorithm recorded the highest accuracy at 92.43%, with a recall of 88.67% and an F1-score of 91.99%, indicating relatively stable performance even under imbalanced conditions. Meanwhile, the Decision Tree (DT) algorithm also performed well with an accuracy of 89.87% and an F1-score of 88.70%, showing a decent balance between precision and recall. In contrast, Gaussian Naive Bayes (GNB), despite achieving 80.82% accuracy, had the lowest recall at 60.45%, suggesting that the model failed to capture most data in the minority class. As a result, its F1-score reached only 75.43%, far below those of SVM and DT. The K-Nearest Neighbors (KNN) algorithm achieved an accuracy of 85.46% and an F1-score of 77.67%, indicating moderate but more stable performance compared to GNB.

Overall, these findings confirm that accuracy alone is not a sufficient metric for evaluating model performance in imbalanced datasets. Greater attention should be paid to recall and F1-score, which better represent a model's success in detecting the minority class. This underscores the need for data balancing methods such as SMOTE during the preprocessing stage to achieve more fair and accurate classifications across all classes.

The next phase of testing involved the use of SMOTE. Figure 4 presents the classification report for the SVM algorithm after applying SMOTE.



	precision	recall	f1-score	support
On-Time Graduation	0.99	1.00	0.99	80
Not On-Time Graduation	1.00	0.99	0.99	80
accuracy			0.99	160
macro avg	0.99	0.99	0.99	160
weighted avg	0.99	0.99	0.99	160

Accuracy : 0.99375

Figure 4. *Classification report of SVM after SMOTE*

As shown in Figure 4, the classification report for SVM after being applied to a balanced dataset demonstrates excellent model performance in classifying both classes. The metrics—precision, recall, and F1-score—are nearly perfect for both classes. For the "On-Time Graduation" class, the model achieved a precision of 0.99, recall of 1.00, and F1-score of 0.99. For the "Not On-Time Graduation" class, all metrics were similarly high. The overall model accuracy reached 99.375%, with both macro and weighted averages also reflecting excellent performance. These results clearly show that applying SMOTE was highly effective in improving model performance, particularly in recognizing the previously underrepresented minority class.

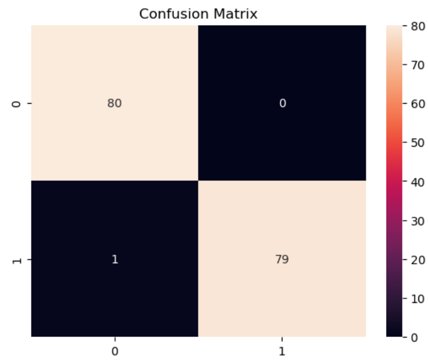


Figure 5. *Confusion matrix of SVM*

ISSN: p-ISSN 1979-9160 (Print)| e-ISSN 2549-7901 (Online)
DOI: https://doi.org/10.15408/jti.v18i2.45913

Figure 5 also includes the confusion matrix for the SVM model on the balanced dataset. It visually confirms the classification performance: 80 instances were correctly classified as "On-Time Graduation," and 79 were correctly classified as "Not On-Time Graduation." There was only one misclassification—an instance of "Not On-Time Graduation" being predicted as "On-Time Graduation." The absence of false positives for the majority class and the presence of only a single false negative demonstrate the model's strong precision and sensitivity after balancing. This reinforces the classification report results, emphasizing that SMOTE significantly enhanced the model's overall fairness and precision in detecting both classes.

Subsequent tests were conducted on the other algorithms using SMOTE, and the results are shown in Table 2.

Table 2. *Testing With SMOTE*

Algorithm	Acc	Precision	Recall	F1-Score
DT + SMOTE	99.37	99.37	99.37	99.37
GNB + SMOTE	98.75	98.75	98.75	98.75
KNN + SMOTE	98.75	98.75	98.75	98.75

Table 2 shows that after applying SMOTE, the performance of all machine learning algorithm Decision Tree (DT), Gaussian Naive Bayes (GNB), and K-Nearest Neighbors (KNN) significantly improved. Each algorithm achieved exceptionally high accuracy, precision, recall, and F1-score values. DT with SMOTE recorded 99.37% across all evaluation metrics, while both GNB and KNN achieved 98.75%.

These results further support the conclusion that SMOTE is effective in balancing class distribution, allowing the models to recognize patterns from both classes more fairly. The noticeable improvement in recall indicates that the models became more capable of identifying the minority class, which was previously underrepresented in the imbalanced dataset.

Compared to the SVM model, as presented in Figures 2 and 3, SVM with SMOTE also achieved highly competitive results, reaching 99.375% accuracy and nearly perfect precision, recall, and F1-score. The confusion matrix showed only one misclassified instance out of 160 test samples, indicating a very low error rate.

Thus, when compared with other tested algorithms, SVM remained the most stable and consistently high-performing model both before and after data balancing. This shows that SVM is not only effective in addressing classification tasks on high-dimensional datasets but is also highly responsive to improvements in data distribution through SMOTE. Therefore, in the context of this research, SVM can be considered the most robust and reliable algorithm for classifying on-time student graduation, especially when supported by proper preprocessing strategies.

The final step of analysis was comparing the current study's results with prior studies related to timely student graduation prediction.

Table 3. *Comparison with previous studies*

Researcher	Algorithm	Dataset	Accuracy
[19]	SVM	Student Graduation	71%
[34]	SVM	Student Graduation	96%
[35]	C4.5	Student Graduation	90%
[36]	SVM + PSO	Student Graduation	79%
[37]	C4.5	Student Graduation	94%
[38]	Random Forest + Binning + SMOTE	Student Graduation	91%
[39]	Naïve Bayes	Student Graduation	85%
This Study	SVM + SMOTE	Student Graduation	99%
This Study	GNB + SMOTE	Student Graduation	99%
This Study	KNN + SMOTE	Student Graduation	99%
This Study	DT + SMOTE	Student Graduation	99%

Based on the comparison presented in Table 3, this study shows a significant improvement over previous research in predicting on-time student graduation. A variety of algorithms were used in prior studies, including Support Vector Machine (SVM), C4.5, Naïve Bayes, and Random Forest, with varying accuracy levels. For instance, studies [29] and [32] using SVM reported 71% and 79% accuracy respectively, even when [32] used SVM combined with Particle Swarm Optimization (PSO). Study [30], which also used SVM, achieved a high accuracy of 96%, while C4.5 in studies [31] and [33] produced accuracies of 90% and 94%, respectively. Study [34], which employed Random Forest combined with Binning and SMOTE, reached 91%, and study [35] using Naïve Bayes achieved 85%.

In contrast, the current study reached 99% accuracy across all four models used: SVM + SMOTE, GNB + SMOTE, KNN + SMOTE, and DT + SMOTE. This demonstrates that consistently applying data balancing techniques such as SMOTE can significantly enhance the performance of machine learning algorithms when dealing with class imbalance in graduation datasets. Furthermore, SMOTE not only improved SVM performance but also significantly boosted other algorithms such as Naïve Bayes and Decision Tree, which had shown lower performance in earlier studies. These findings reinforce this study's contribution in showing that proper preprocessing strategies play a crucial role in producing accurate and reliable classification models—particularly in higher education contexts.

CONCLUSION

This study demonstrates that class imbalance in student graduation data has a significant impact on the performance of machine learning classification algorithms. Without the application of balancing techniques such as SMOTE, the algorithms tend to be biased toward the majority class and fail to detect students at risk of not graduating on time. By applying SMOTE, all evaluated algorithms Decision Tree, Gaussian Naive Bayes, K-Nearest Neighbors, and Support Vector Machine showed a substantial improvement in performance, achieving accuracy levels of around 99%. Among the four models, SVM exhibited the most stable performance both before and after data balancing. These results confirm that appropriate preprocessing strategies, particularly in addressing data imbalance, play a crucial role in developing accurate, fair, and reliable graduation prediction systems. The findings of this study can serve as a basis for academic policy-making aimed at improving on-time graduation rates in higher education institutions. For future research, it is recommended to explore hyperparameter tuning to further optimize model performance, and to apply ensemble methods such as bagging, boosting, or stacking in order to enhance robustness and generalization of the classification results across different datasets.

REFERENCES

[1] Herianto, B. Kurniawan, Z. H. Hartomi, Y. Irawan, and M. K. Anam, "Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction," *Journal of Applied Data Sciences*, vol. 5, no. 3, pp. 1272–1285, Sep. 2024, doi: 10.47738/jads.v5i3.316.

[2] R. Bakri, N. P. Astuti, and A. S. Ahmar, "Machine Learning Algorithms with Parameter Tuning to Predict Students' Graduation-on-time: A Case Study in Higher Education," *Journal of Applied Science, Engineering, Technology, and Education*, vol. 4, no. 2, pp. 259–265, Dec. 2022, doi: 10.35877/454ri.asci1581.

[3] R. Sepriansyah and S. D. Purnamasari, "Prediction of Student Graduation Using Naïve Bayes," *Budapest International Research and Critics Institute-Journal (BIRCI-Journal)*, vol. 5, no. 3, pp. 24255–24268, 2022, doi: 10.33258/birci.v5i3.6447.

[4] A. Desfiandi and B. Soewito, "Student Graduation Time Prediction Using Logistic Regression, Decision Tree, Support Vector, And Adaboost Ensemble Learning," *International Journal of Information System and Computer Science) IJISCS*, vol. 7, no. 3, pp. 195–199, 2023, doi: 10.56327/ijiscs.v7i2.1579.

[5] M. K. Anam, M. B. Firdaus, F. Suandi, Lathifah, T. Nasution, and S. Fadly, "Performance Improvement of Machine Learning Algorithm Using Ensemble Method on Text Mining," in *ICFTSS 2024 - International Conference on Future Technologies for Smart Society*, Kuala Lumpur: Institute of Electrical and Electronics Engineers Inc., Sep. 2024, pp. 90–95. doi: 10.1109/ICFTSS61109.2024.10691363.

[6] A. N. Ulfah and M. K. Anam, "Analisis Sentimen Hate Speech Pada Portal Berita Online Menggunakan Support Vector Machine (SVM)," *JATISI (Jurnal Teknik Informatika dan Sistem Informasi)*, vol.

7, no. 1, pp. 1–10, 2020, doi: 10.35957/jatisi.v7i1.196.

[7] L. L. Van FC, M. K. Anam, S. Bukhori, A. K. Mahamad, S. Saon, and R. L. V. Nyoto, "The Development of Stacking Techniques in Machine Learning for Breast Cancer Detection," *Journal of Applied Data Sciences*, vol. 6, no. 1, pp. 71–85, Jan. 2025, doi: 10.47738/jads.v6i1.416.

[8] F. Suandi *et al.*, "Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning Algorithms," in *International Conference on Applied Engineering*, Atlantis Press, 2024, pp. 126–138. doi: 10.2991/978-94-6463-620-8_10.

[9] F. P. Arifianti and A. Salam, "XGBoost and Random Forest Optimization using SMOTE to Classify Air Quality," *Advance Sustainable Science, Engineering and Technology*, vol. 6, no. 1, pp. 1–8, Nov. 2024, doi: 10.26877/asset.v6i1.18136.

[10] W. Satria and M. Riasetiawan, "Essay Answer Classification with SMOTE Random Forest And Adaboost In Automated Essay Scoring," *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, vol. 17, no. 4, p. 359, Oct. 2023, doi: 10.22146/ijccs.82548.

[11] A. Tholib, M. N. Fadli Hidayat, S. Yono, R. Wulanningrum, and E. Daniati, "Comparison of C4.5 and Naive Bayes for Predicting Student Graduation Using Machine Learning Algorithms," *International Journal of Engineering and Computer Science Applications (IJECSA)*, vol. 2, no. 2, pp. 65–72, Sep. 2023, doi: 10.30812/ijecsa.v2i2.3364.

[12] M. Arifin, F. Helmi, and R. B. Hikmawansyah, "Analisis Metode dan Algoritma Dalam Sistem Pendukung Keputusan untuk Memprediksi Kelulusan," *Jurnal Advance Research Informatika*, vol. 3, no. 1, p. 73, Dec. 2024, doi: 10.24929/jars.v3i1.4045.

[13] M. Putra and Erwin Harahap, "Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest," *Jurnal Riset Matematika*, vol. 4, no. 2, pp. 127–136, Dec. 2024, doi: 10.29313/jrm.v4i2.5102.

[14] B. N. Aisyah and I. Gunawan, "Penerapan Machine Learning Untuk Memprediksi Kelulusan Mahaiswa Menggunakan Algoritma Decision Tree," in *Seminar Nasional Hasil Penelitian & Pengabdian Masyarakat Bidang Ilmu Komputer*, 2024, pp. 1–6.

[15] P. P. Putra, M. K. Anam, S. Defit, and A. Yunianta, "Enhancing the Decision Tree Algorithm to Improve Performance Across Various Datasets," *INTENSIF: Jurnal Ilmiah Penelitian dan Penerapan Teknologi Sistem Informasi*, vol. 8, no. 2, pp. 200–212, Aug. 2024, doi: 10.29407/intensif.v8i2.22280.

[16] M. Vedaraj, C. S. Anita, A. Muralidhar, V. Lavanya, K. Balasaranya, and P. Jagadeesan, "Early Prediction of Lung Cancer Using Gaussian Naive Bayes Classification Algorithm," *International Journal of Intelligent Systems and Applications in Engineering IJISAE*, vol. 11, no. 6s, pp. 838–848, 2023, [Online]. Available: www.ijisae.org

[17] M. K. Anam, B. N. Pikir, M. B. Firdaus, S. Erlinda, and Agustin, "Penerapan Naïve Bayes Classifier , K-Nearest Neighbor dan Decision Tree untuk Menganalisis Sentimen pada Interaksi Netizen dan Pemeritah," *Matrik: Jurnal Manajemen, Teknik Informatika, dan Rekayasa Komputer*, vol. 21, no. 1, pp. 139–150, 2021, doi: 10.30812/matrik.v21i1.1092.

[18] M. K. Anam, S. Sumijan, K. Karfindo, and M. B. Firdaus, "Comparison Analysis of HSV Method, CNN Algorithm, and SVM Algorithm in Detecting the Ripeness of Mangosteen Fruit Images," *Indonesian Journal of Artificial Intelligence and Data Mining*, vol. 7, no. 2, pp. 348–356, May 2024, doi: 10.24014/ijaidm.v7i2.29739.

[19] S. Abdillah, G. J. Yanris, and V. Sihombing, "Implementation Of The Support Vector Machine Method In Predicting Student Graduation," *International Journal of Science*, vol. 6, no. 1, pp. 263–270, Jan. 2025, doi: 10.46729/ijstm.v6i1.1265.

[20] Q. Widayati, K. Adi, R. R. Isnanto, E. P. Agustini, D. R. R. Julianto, and F. B. Prakasa, "Predicting Student Loyalty in Higher Education Using Machine

Learning: A Random Forest Approach," *Journal of Information Systems and Informatics*, vol. 7, no. 1, pp. 63–77, Mar. 2025, doi: 10.51519/journalisi.v7i1.977.

[21] D. A. Rachmawati, N. A. Ibadurrahman, J. Zeniarja, and N. Hendriyanto, "Implementation of The Random Forest Algorithm In Classifying The Accuracy of Graduation Time For Computer Engineering Students At Dian Nuswantoro University," *Jurnal Teknik Informatika (Jutif)*, vol. 4, no. 3, pp. 565–572, Jun. 2023, doi: 10.52436/1.jutif.2023.4.3.920.

[22] S. Mehta, "Playing Smart with Numbers: Predicting Student Graduation Using the Magic of Naive Bayes," *International Transactions on Artificial Intelligence (ITALIC)*, vol. 2, no. 1, pp. 60–75, 2023, doi: 10.33050/italic.v2i1.405.

[23] H. Mukhtar, J. Al Amien, and F. Dewi, "Prediction of Student Graduation Using Decision Tree Method," in *CELSciTech*, Pekanbaru, 2021, pp. 7–18.

[24] A. Santoso, H. Retnawati, Kartianom, E. Apino, I. Rafi, and M. N. Rosyada, "Predicting Time to Graduation of Open University Students: An Educational Data Mining Study," *Open Education Studies*, vol. 6, no. 1, pp. 1–14, Jan. 2024, doi: 10.1515/edu-2022-0220.

[25] T. A. Assegie, "An optimized K-Nearest neighbor based breast cancer detection," *Journal of Robotics and Control (JRC)*, vol. 2, no. 3, pp. 115–118, May 2021, doi: 10.18196/jrc.2363.

[26] T. R. Mahesh *et al.*, "Transformative Breast Cancer Diagnosis using CNNs with Optimized ReduceLROnPlateau and Early Stopping Enhancements," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, pp. 1–8, Dec. 2024, doi: 10.1007/s44196-023-00397-1.

[27] N. B. Aji, Kurnianingsih, N. Masuyama, and Y. Nojima, "CNN-LSTM for Heartbeat Sound Classification," *International Journal on Informatics Visualization*, vol. 8, no. 2, pp. 735–741, 2024, doi: 10.62527/joiv.8.2.2115.

[28] M. K. Anam *et al.*, "Sara Detection on Social Media Using Deep Learning Algorithm Development," *Journal of Applied Engineering and Technological Science*, vol. 6, no. 1, pp. 225–237, Dec. 2024, doi: 10.37385/jaets.v6i1.5390.

[29] I. F. Ashari, M. C. Untoro, E. M. Sutowo, D. Salsabila, and D. A. Zhabiyan, "Hyperparameter Tuning Feature Selection with Genetic Algorithm and Gaussian Naïve Bayes for Diabetes Disease Prediction," *Jurnal Telematika*, vol. 17, no. 1, pp. 17–23, 2022, doi: 10.61769/jurtel.v17i1.488.

[30] A. Masitha and M. Kunta Biddinika, "Preparing Dual Data Normalization for KNN Classfication in Prediction of Heart Failure," *KLIK*, vol. 4, no. 3, pp. 1227–1234, 2023, doi: 10.30865/klik.v4i3.1382.

[31] Y. A. Singgalen, "Performance Evaluation of Sentiment Classification Models: A Comparative Study of NBC, SVM, and DT with SMOTE," *Media Online*, vol. 4, no. 5, pp. 2539–2548, Apr. 2024, doi: 10.30865/klik.v4i5.1827.

[32] M. K. Anam, T. P. Lestari, H. Yenni, T. Nasution, and M. B. Firdaus, "Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble," *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, vol. 19, no. 2, pp. 159–167, Mar. 2025, doi: 10.37936/ecti-cit.2025192.257815.

[33] M. K. Anam *et al.*, "Enhancing the Performance of Machine Learning Algorithm for Intent Sentiment Analysis on Village Fund Topic," *Journal of Applied Data Sciences*, vol. 6, no. 2, pp. 1102–1115, 2025, doi: 10.47738/jads.v6i2.637.

[34] Y. Yati, Moh. A. Yaqin, and A. Y. Nadhiroh, "Application of The Support Vector Machine Algorithm for Timely Student Graduation Prediction Based on Streamlit Web at The Faculty of Informatics Engineering Nurul Jadid University," *Journal of Computer Networks, Architecture and High Performance Computing*, vol. 6, no. 3, pp. 1066–1070, Jul. 2024, doi: 10.47709/cnahpc.v6i3.3918.

[35] H. Yuliansyah, R. A. P. Imaniati, A. Wirasto, and M. Wibowo, "Predicting Students Graduate on Time Using C4.5 Algorithm," *Journal of Information Systems Engineering and Business*

Intelligence, vol. 7, no. 1, p. 67, Apr. 2021, doi: 10.20473/jisebi.7.1.67-73.

[36] S. Hidayatulloh, G. Triyono, and K. A. S. kosasih, "Prediction Model For Students' On-Time Graduation Using Algorithm Support Vector Machine (SVM) Based Particle Swarm Optimization (PSO)," *JSiI (Jurnal Sistem Informasi)*, vol. 12, no. 1, pp. 25–32, Mar. 2025, doi: 10.30656/jsii.v11i1.6936.

[37] N. A. Prahastiwi, R. Andreswari, and R. Fauzi, "Students Graduation Prediction Based on Academic Data Record Using The Decision Tree Algorithm C4.5 Method," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 8, no. 3, pp. 295–304, Aug. 2022, doi: 10.33330/jurteksi.v8i3.1680.

[38] F. Rahman and Mustikasari, "Optimalisasi Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Binning dan Synthetic Minority Oversampling Technique (SMOTE)," vol. 4, no. 1, pp. 30–36, 2024, doi: 10.24252/jagti.v4i1.77.

[39] S. Mehta, "Playing Smart with Numbers: Predicting Student Graduation Using the Magic of Naive Bayes," *International Transactions on Artificial Intelligence*, vol. 2, no. 1, pp. 60–75, Oct. 2023, doi: 10.33050/italic.v2i1.405.