SinkrOn
JURNAL & PENELITIAN TEKNIK INFORMATIKA
Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Optimization of Machine Learning Models in Student Graduation Prediction Systems Using Ensemble Learning with PSO and SMOTE

Hamdani[1)*], Susanti[2)], M. Khairul Anam[3)], Rahman Pradipta[4)], Lathifah[5)]
[1,2)] Universitas Sains dan Teknologi Indonesia, Indonesia
[3)] Universitas Teknokrat Indonesia, Indonesia, [4,5)]Universitas Samudra, Indonesia
[1)]hamdani@usti.ac.id, [2)]Susanti@usti.ac.id, [3)]lathifah@teknokrat.ac.id [4)]khairulanam@unsam.ac.id, [5)]rahmanpradipta@unsam.ac.id

Submitted :Sep 14, 2025 | **Accepted** : Oct 6, 2025 | **Published** : Oct 15, 2025

Abstract: The timely graduation of students is a key metric in evaluating the academic effectiveness of higher education institutions. However, accurately identifying students at risk of delayed graduation remains challenging due to imbalanced data distributions and the instability of single-model prediction approaches. This study proposes an optimized ensemble-based machine learning system for predicting on-time graduation among university students. The model integrates C4.5, K-Nearest Neighbor (KNN), and Random Forest algorithms through a hard voting classifier, which is further optimized using Particle Swarm Optimization (PSO) to determine the most effective weighting configuration. To address class imbalance, the Synthetic Minority Over-sampling Technique (SMOTE) is implemented, ensuring balanced representation between timely and delayed graduates. A dataset of 809 student academic records from Universitas Sains dan Teknologi Indonesia (USTI) was used, and performance was evaluated using 5-fold cross-validation. The proposed ensemble model achieved an average accuracy of 93.70%, a precision of 0.94, a recall of 0.93, and an F1-score of 0.94, outperforming each individual classifier. These results confirm that the combination of ensemble learning, PSO-based optimization, and data balancing effectively improves both accuracy and model stability. The findings highlight the system's potential as a reliable decision-support tool for educational institutions to anticipate delayed graduations and improve academic supervision strategies.

Keywords: Graduation Prediction; Ensemble Learning; SMOTE; Particle Swarm Optimization; Voting Classifier

INTRODUCTION

Timely graduation of students is one of the key indicators in evaluating the effectiveness of higher education delivery (Herianto et al., 2024). At the Universitas Sains dan Teknologi Indonesia (USTI), achieving timely graduation is a strategic priority due to its contribution to study program accreditation and the institution's public reputation. Delayed graduation can result in a waste of resources for both students and the institution and may hinder the academic regeneration process (Bakri et al., 2022). Therefore, an accurate predictive system is needed to identify students at risk of not graduating on time, serving as a support tool for more adaptive and anticipatory academic supervision policies (Dwinanda et al., 2023).

Various studies have been conducted to build student graduation prediction systems using machine learning algorithms. For instance, (Riadi et al., 2024) employed the K-Nearest Neighbor (KNN) algorithm with an accuracy of 78%, while (Junaidi et al., 2023) applied the Naïve Bayes algorithm with an accuracy of 88%. Research by (Hasibuan & Mahdiana, 2023) utilized the C4.5 algorithm and recorded an accuracy of 75.52%, whereas (Moerdyanto & Nuryana, 2023) built a prediction system using a decision tree algorithm and achieved an accuracy of 75.95%. Although these results are promising, most of the studies are still limited to using single algorithms, which tend to be unstable against data variation and often fail to generalize well when applied to new datasets. Another limitation is that many of these studies did not address the issue of imbalanced data, even though, in practice, the distribution between students who graduate on time and those who do not is often highly skewed.

*name of corresponding author

This highlights a clear research gap in which previous studies relied mainly on single models that were unstable when applied to different datasets and did not take into account the problem of imbalanced data distribution. Such limitations can cause biased prediction results that tend to favor the majority class, making it difficult to detect students who are most at risk of delayed graduation. Therefore, this study emphasizes the importance of developing a more stable, fair, and generalizable model that can overcome these limitations through the integration of ensemble learning and data balancing techniques.

Using a single algorithm without adjusting for class distribution poses the risk of producing a model biased toward the majority class (Chen et al., 2024; Van FC et al., 2025). For example, a model may accurately predict students who graduate on time but fail to identify those at risk of delayed graduation a group that is arguably the most critical to detect. Therefore, ensemble-based approaches and data balancing techniques offer potential solutions for improving the accuracy and fairness of classification outcomes (Anam, Lestari, Yenni, et al., 2025; Omotehinwa & Oyewola, 2023; Pavitha & Sugave, 2022; Yin & Li, 2022).

This study proposes a novel approach by combining three machine learning algorithms K-Nearest Neighbor (KNN), Random Forest, and C4.5 through a hard voting ensemble method. This voting mechanism is selected because it leverages the diversity of base classifiers to reduce variance and improve generalization, resulting in more stable and accurate predictions compared to individual models (Suandi et al., 2024). Furthermore, the ensemble is enhanced with Particle Swarm Optimization (PSO), which is used to determine the optimal weights for each classifier in the voting process. PSO was chosen for its efficiency in finding global optima in complex search spaces, ensuring that the contribution of each algorithm is proportionate to its performance (Pirapong et al., 2024). To address the issue of class imbalance, which often leads to biased models, the Synthetic Minority Over-sampling Technique (SMOTE) is applied. SMOTE was selected because it generates synthetic examples of the minority class rather than simply duplicating data, thus enriching the feature space and improving the model's ability to learn minority class patterns effectively (Chopannejad et al., 2024).

To ensure the reliability of outcomes, evaluation is conducted using 5-fold cross-validation. This technique provides a more consistent and robust assessment of model performance by minimizing variance and avoiding overfitting to a specific subset (Azis et al., 2020). The main objective of this research is to develop an optimized ensemble learning model that integrates PSO and SMOTE to improve predictive accuracy, stability, and fairness in student graduation prediction. In addition, this study aims to evaluate the effectiveness of the proposed model through comprehensive testing and analysis using multiple performance metrics. The contribution of this study lies in three main aspects, namely methodological advancement through the integration of ensemble learning, optimization, and data balancing; empirical validation through rigorous evaluation; and practical implications by providing a reliable tool to assist academic institutions in early identification and intervention for students at risk of delayed graduation.

LITERATURE REVIEW

Previous studies on student graduation prediction indicate that most approaches still rely on single algorithms. (Dina Amalia Putri et al., 2025), for instance, applied K-Nearest Neighbor (KNN) and reported an accuracy of 76%. In contrast, (Mehta, 2023) employed Naïve Bayes and achieved a higher accuracy of 85%. (Prayitno et al., 2021) utilized the C4.5 algorithm with an accuracy of 79%, while (Co & Casillano, 2021) obtained 88.9% using a decision tree. These findings show that single models can provide initial insights, but their predictive stability and generalization remain limited when exposed to diverse data.

Another challenge that has often been overlooked is the issue of class imbalance. In academic settings, the proportion of students who graduate on time is typically far greater than those who do not, leading models to overpredict the majority class (Latief et al., 2024). Consequently, students at risk of delayed graduation are frequently misclassified. The Synthetic Minority Over-Sampling Technique (SMOTE) has been widely recommended as a remedy, as it creates synthetic samples of minority classes. Unlike conventional oversampling, SMOTE enriches the feature space and improves the model’s capacity to capture minority class patterns (Mubarak et al., 2023).

To overcome the limitations of single algorithms, ensemble learning has emerged as a promising alternative. By combining multiple classifiers, prediction outcomes become more stable. The hard voting classifier, for example, reduces variance by aggregating the decisions of diverse base models (Anam, Lestari, Efrizoni, et al., 2025). However, without weight adjustments, the influence of each model is often disproportionate. This gap has led to the adoption of metaheuristic optimization techniques such as Particle Swarm Optimization (PSO), which determines optimal voting weights so that stronger classifiers contribute more significantly to the final decision (Susanto & Suparwito, 2023). The comparison of previous studies is summarized in Table 1.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Table 1 Comparison of Previous Studies on Student Graduation Prediction

Author(s)	Method	Dataset	Accuracy
Dina Amalia Putri et al., 2025	K-Nearest Neighbor (KNN)	Academic dataset (students)	76%
Mehta, 2023	Naïve Bayes	Academic dataset	85%
Prayitno et al., 2021	C4.5 Decision Tree	Academic dataset	79%
Co & Casillano, 2021	Decision Tree	Academic dataset	88.9%

Overall, the literature highlights a clear research gap in integrating ensemble learning, data balancing, and weight optimization through metaheuristics. The novelty of this study lies in proposing a hybrid approach that combines ensemble voting with Particle Swarm Optimization (PSO) and SMOTE. Unlike prior studies that relied on single algorithms or ensembles without optimization and balancing, this research offers a more stable, accurate, and fair prediction system for student graduation.

METHOD

This study adopts a quantitative experimental approach aimed at developing a predictive model for on-time student graduation by combining machine learning algorithms with optimization and data balancing techniques. The methodological workflow consists of several key stages, as illustrated in Figure 1: data collection, labeling, data balancing using SMOTE, data validation, model construction, model integration through a voting classifier, optimization using Particle Swarm Optimization (PSO), and final model evaluation.

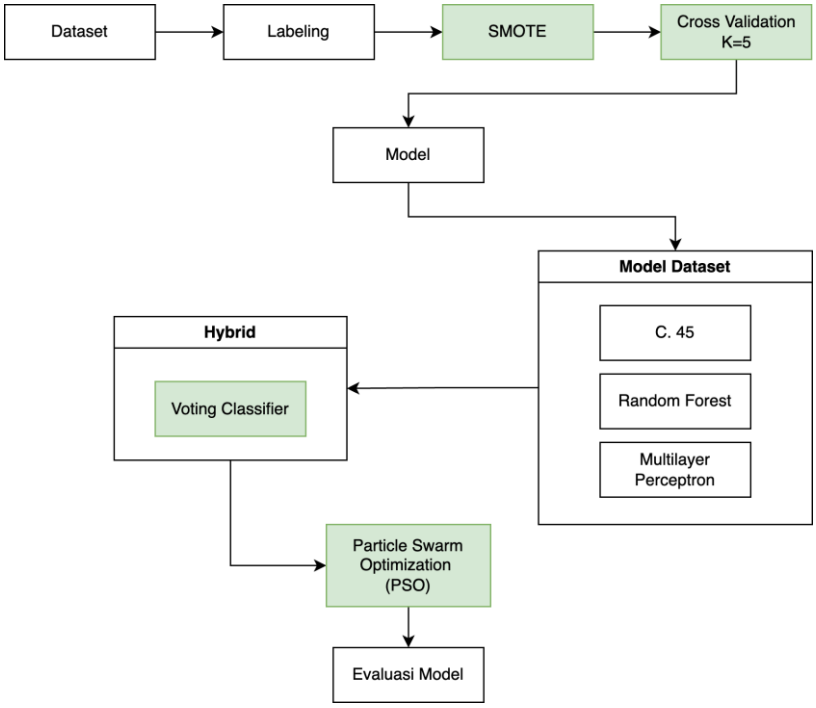

Figure 1. Flow Methodology

Research Dataset

The dataset used in this study consists of academic records of USTI students from the 2017–2019 cohorts, totaling 809 entries. The attributes include the Grade Point Average (GPA) for semesters 6, 7, and 8; the number of credits (SKS) taken in each of those semesters; thesis completion status; and thesis score. The target class (label) consists of two categories: “Graduated On Time” and “Did Not Graduate On Time.” The dataset contains no missing values, hence no imputation was required. This clean and representative data structure allows the model training process to be conducted efficiently and accurately.

Data Preprocessing

Before training the model, it was necessary to examine the class distribution of the dataset to identify potential imbalance problems that could affect prediction performance. The initial analysis revealed a disproportionate number of students graduating on time compared to those who did not. To address this issue, the Synthetic Minority

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

Over-sampling Technique (SMOTE) was applied to generate additional synthetic samples for the minority class. Figures 2 and 3 illustrate the class distribution before and after the application of SMOTE.

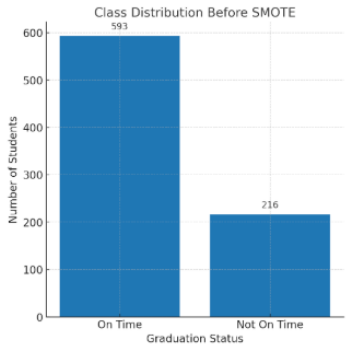

Figure 2. Class Distribution Before SMOTE

Figure 2 shows that the dataset was highly imbalanced, consisting of 593 students who graduated on time and 216 students who did not graduate on time. This uneven ratio indicates that the "On Time" class dominated the dataset. Such imbalance can lead the model to bias toward the majority class, thereby reducing its ability to accurately identify students at risk of delayed graduation. This imbalance motivated the use of SMOTE in the preprocessing stage.

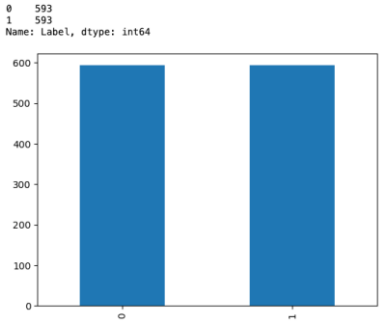

Figure 3. Class Distribution After SMOTE

Figure 3 presents the distribution after applying SMOTE. The technique generated synthetic examples for the minority class ("Not On Time") based on existing feature relationships, resulting in a balanced dataset of 593 samples per class. This balance allows the model to learn patterns from both classes more effectively and fairly, leading to improved accuracy, stability, and generalization capability (Erlin et al., 2022).

Data Validation

To assess the consistency and reliability of the model, the K-Fold Cross Validation technique was used with K=5. The dataset was randomly split into five folds. In each iteration, one-fold served as the test set, while the remaining four folds were used for training. This process was repeated five times, and the evaluation results were averaged to obtain a stable measure of model performance (Chamorro-Atalaya et al., 2023).

Based Model Construction

Three machine learning algorithms were used as base classifiers:

- C4.5, a decision tree algorithm, produces an interpretable rule-based model that enables understanding of feature importance and decision logic (Saputra et al., 2023);
- Random Forest, an ensemble algorithm composed of multiple decision trees, is recognized for its robustness and ability to reduce overfitting by aggregating multiple weak learners into a strong predictor (Putra & Erwin Harahap, 2024);
- K-Nearest Neighbor (KNN), a distance-based algorithm, classifies instances according to their proximity in the feature space, making it sensitive to local data structure and nonlinear relationships (Prayoga et al., 2023).

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

The selection of these three algorithms was intentionally based on their methodological diversity. C4.5 represents a rule-based approach with strong interpretability, Random Forest emphasizes ensemble robustness and variance reduction, and KNN contributes a similarity-based perspective. Combining these heterogeneous models ensures that the ensemble system captures different learning patterns, reduces bias and variance, and enhances predictive stability when applied to complex academic data. Each algorithm was trained using the balanced dataset and validated through the K-Fold cross-validation scheme.

Model Integration with Voting Classifier

After building the base models, they were integrated using the hard voting classifier technique. In this method, each model casts a vote for the predicted class. The final prediction is determined by the majority vote. This ensemble approach aims to combine the strengths of each base model to improve prediction accuracy and stability (Li, 2024).

Model Optimization with PSO

To further optimize the ensemble result, the Particle Swarm Optimization (PSO) algorithm was used to find the best weight configuration for each model in the voting process. PSO simulates the behavior of particle swarms in finding a global solution by evaluating different weight combinations based on their prediction performance. This process helps assign proportionate influence to each model in the final decision-making process (Gupta & Rattan, 2023). The following is pseudocode for the development carried out.

Input: Training data (X_train, y_train), Base models {C4.5, KNN, RF}
Output: Optimized ensemble classifier with PSO-based weights
1. Initialize particle population with random weights [w1, w2, w3]
2. For each particle:
a. Normalize weights so that w1 + w2 + w3 = 1
b. Train base models (C4.5, KNN, RF) on X_train
c. Predict using weighted voting:
*y_pred = argmax(w1*C4.5(x) + w2*KNN(x) + w3*RF(x))*
d. Evaluate fitness using accuracy or F1-score
3. Update particle velocity and position based on PSO rules
4. Repeat until maximum iteration or convergence
5. Select the best weight combination for final ensemble model

The proposed algorithm begins by taking the training data (X_{train},y_{train}) and three base models, namely C4.5, K-Nearest Neighbor (KNN), and Random Forest (RF), as its input. The output of this process is an optimized ensemble classifier whose voting weights are determined using the Particle Swarm Optimization (PSO) technique. Initially, a population of particles is created, where each particle represents a possible combination of voting weights [w_1, w_2, w_3] assigned to the three base models. These weights are then normalized so that their total equals one, ensuring proportional contribution among classifiers.

For every particle, the base models (C4.5, KNN, and RF) are trained using the training dataset. Each trained model produces predictions that are subsequently combined using a weighted voting mechanism, where the final predicted class is determined by the model receiving the highest weighted score:

$$y_{pred} = argmax(w_1 \times C4.5(x) + w_2 \times KNN(x) + w_3 \times RF(x)) \quad (1)$$

The performance of each particle is then evaluated using an objective fitness function, such as accuracy or F1-score, which measures how well that particular weight configuration performs.

Next, the PSO algorithm updates each particle's velocity and position according to the swarm's best-known positions, allowing particles to iteratively move toward better solutions in the search space. This process continues until a predefined number of iterations is reached or the fitness score converges. In the final step, the best-performing particle, representing the optimal combination of weights, is selected as the final configuration for the ensemble model. As a result, the ensemble classifier achieves an optimal balance among the three base models, improving overall prediction accuracy and stability compared to conventional unweighted voting.

Model Performance Evaluation

The final model was evaluated using classification metrics including accuracy, precision, recall, and F1-score. The evaluation was conducted on each fold during the cross-validation process, and the results were averaged to

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

reflect the overall model performance. This evaluation provides insights into how effectively the system can predict students who are likely to graduate on time as well as those at risk of delay.

RESULT

The experimental results are summarized to demonstrate the comparative performance of the individual base models (C4.5, KNN, and Random Forest) and the proposed Voting Classifier enhanced with Particle Swarm Optimization (PSO). Each model was trained and evaluated using 5-Fold Cross Validation to ensure reliable and consistent performance measurement. As illustrated in Figure 7, the proposed Voting + PSO model achieved the highest average accuracy of 93.70%, surpassing C4.5 with 92.69%, KNN with 92.51%, and Random Forest with 92.30%. Although the numerical differences among the models appear modest, the improvement achieved by the Voting + PSO method is significant in terms of consistency and robustness. This indicates that the integration of ensemble learning with PSO successfully optimized the weight contribution of each classifier, leading to better stability and predictive reliability compared to single models.

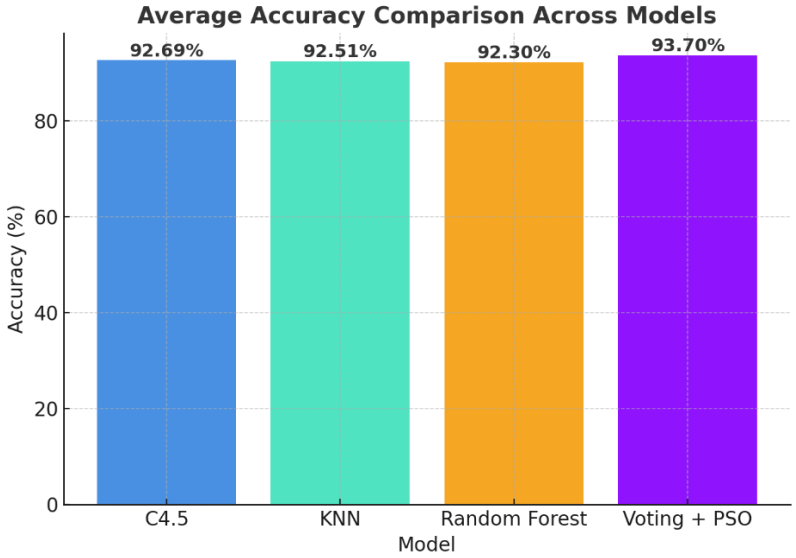

Figure 4. Summary of Model Performance Metrics

Table 3 presents a detailed summary of the classification metrics, including Accuracy, Precision, Recall, and F1-score, for each model. The C4.5 model achieved an accuracy of 92.69%, precision of 0.93, recall of 0.92, and F1-score of 0.93. The KNN model recorded an accuracy of 92.51%, precision of 0.92, recall of 0.91, and F1-score of 0.92. The Random Forest model attained slightly lower values with an accuracy of 92.30%, precision of 0.91, recall of 0.90, and F1-score of 0.91. In contrast, the proposed Voting + PSO model outperformed all three with an accuracy of 93.70%, precision of 0.94, recall of 0.93, and F1-score of 0.94.

Table 2
Summary of Model Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
C4.5	92.69	0.93	0.92	0.93
KNN	92.51	0.92	0.91	0.92
Random Forest	92.30	0.91	0.90	0.91
Voting + PSO	**93.70**	**0.94**	**0.93**	**0.94**

The balanced performance across these metrics reflects that the Voting + PSO approach not only enhances accuracy but also maintains high precision and recall, ensuring fairness in identifying both timely and delayed graduates. The improvement achieved by the Voting + PSO model highlights the benefit of combining algorithmic diversity (C4.5 for interpretability, KNN for proximity-based reasoning, and Random Forest for ensemble robustness) with metaheuristic optimization. This synergy allows the model to capture complementary decision boundaries and reduce individual classifier biases. Consequently, the optimized ensemble delivers a more generalizable and dependable performance in predicting student graduation outcomes. These findings validate the hypothesis that integrating ensemble learning with PSO and SMOTE can effectively address instability and imbalance issues inherent in previous single-model approaches, thereby providing a more reliable decision-support framework for higher education analytics.

*name of corresponding author

Sinkron : Jurnal dan Penelitian Teknik Informatika
Volume 9, Issue 4, October 2025
DOI : https://doi.org/10.33395/sinkron.v9i4.15335
e-ISSN : 2541-2019
p-ISSN : 2541-044X

DISCUSSIONS

In this section, the researchers can give a simple discussion related to the results of the research trials.This section contains the author's opinion about the research results obtained. Common features of the discussion section include the comparison between measured and modeled data or comparison among various modeling methods, the results obtained to solve a specific engineering or scientific problem, and further explanation of new and significant findings.

The findings indicate that integrating C4.5, KNN, and Random Forest within a Voting Classifier optimized using PSO achieved an average accuracy of 93.70%. This performance surpasses that of the individual algorithms, with C4.5 reaching 92.69%, KNN 92.51%, and Random Forest 92.30%. Although the numerical differences may appear modest, the consistent improvement across all validation folds highlights that the combined model offers stronger generalization capabilities.

The role of PSO proved significant in balancing the contribution of each algorithm. Without weight optimization, the ensemble relies solely on majority voting, which can distort the influence of weaker-performing models. Through PSO, optimal weights are determined based on predictive quality, allowing the final model to adjust the influence of each algorithm proportionately. This aligns with the findings of (Gupta & Rattan, 2023), who emphasized PSO's effectiveness in locating global solutions within complex search spaces.

The application of SMOTE also played a critical role in improving performance. The model not only classified students who graduated on time effectively but also provided fairer detection of students at risk of delayed graduation. This addresses the bias often observed in single-algorithm approaches (Van FC et al., 2025). Consequently, the system becomes more relevant for supporting academic decision-making by offering early warning mechanisms for at-risk students.

The practical implication of this study lies in the potential adoption of ensemble-optimization–based prediction systems within higher education institutions. Such systems can serve as decision-support tools for academic monitoring, enabling more effective guidance and timely interventions. Nonetheless, the study remains limited by its relatively small dataset (809 students). Future research should expand testing with larger datasets, across multiple universities, and incorporate non-academic attributes (e.g., socio-economic factors) to further enhance the generalizability of the model. This study also indicates that the results obtained obtained quite high results compared to some previous studies. Table 3 is a comparison with previous research.

Table 3
Comparison with Previous Research

Researcher	Model	Improvement	Accuracy
(Wahyudi et al., 2023)	C4.5	-	88.92%
(Moerdyanto & Nuryana, 2023)	Random Forest	-	75.95%
(Junaidi et al., 2024)	Naïve Bayes	-	87%
(Rachardian & Sediyono, 2024)	KNN	-	75%
(Sari et al., 2024)	Naïve Bayes	-	71.24%
This Research	Voting Classifier	SMOTE and PSO	93.69%

The comparison presented in Table 2 highlights the significant performance improvement achieved by this research compared to several previous studies that employed single machine learning algorithms for graduation prediction. For instance, Wahyudi et al. (2023) used the C4.5 algorithm and achieved an accuracy of 88.92%, while Moerdyanto and Nuryana (2023) applied Random Forest with a notably lower accuracy of 75.95%. Similarly, (Junaidi et al., 2024) and (Sari et al., 2024) implemented Naïve Bayes, reaching 87% and 71.24%, respectively. Rechardian and Sediyono (2024), using K-Nearest Neighbor (KNN), reported an accuracy of 75%.

In contrast, this study outperformed all of them by combining Voting Classifier, SMOTE for data balancing, and Particle Swarm Optimization (PSO) for model weight optimization, resulting in a superior accuracy of 93.69%. The ensemble approach, coupled with advanced preprocessing and optimization techniques, proves to be highly effective in handling data imbalance and increasing the robustness of predictions. This comparison emphasizes the contribution of hybrid modeling and algorithmic tuning in advancing the predictive accuracy of machine learning-based graduation prediction systems.

CONCLUSION

This study successfully developed a robust and accurate machine learning-based prediction system for on-time student graduation. By combining the strengths of three foundational algorithms (C4.5, KNN, and Random Forest) through a hard voting mechanism and optimizing their contribution weights using Particle Swarm Optimization (PSO), the model achieved a superior average accuracy of 93.70%. The use of SMOTE effectively balanced the dataset, allowing the model to fairly learn from both majority and minority classes. Comparative analysis with

*name of corresponding author

previous studies confirmed that the proposed ensemble system significantly outperforms individual baseline models, especially in educational contexts where early detection of at-risk students is crucial for targeted academic interventions. This predictive system not only supports institutional performance evaluation but also offers practical benefits for proactive academic management.

Future studies could explore the integration of deep learning models such as LSTM or Transformer architectures to capture temporal patterns in academic progression. Additionally, incorporating non-academic features such as socio-economic background, attendance records, or extracurricular involvement may further enhance predictive power. Moreover, real-time deployment of the system as an academic monitoring tool, supported by a web or mobile interface, could bring practical benefits to academic advisors and faculty in implementing early interventions.

ACKNOWLEDGMENT

Thanks are due to the Directorate of Research, Technology and Community Service of the Ministry of Higher Education, Science, and Technology of the Republic of Indonesia for providing funding in the 2025 Beginner Lecturer Research Scheme, as well as the Universitas Sains dan Teknologi Indonesia for providing facilities to conduct research activities.

REFERENCES

Anam, M. K., Lestari, T. P., Efrizoni, L., Handayani, N. S. & Andhika, I. (2025). Sentiment Analysis Optimization Using Ensemble of Multiple SVM Kernel Functions. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, *9*(4), 905–914. https://doi.org/10.29207/resti.v9i4.6708

Anam, M. K., Lestari, T. P., Yenni, H., Nasution, T. & Firdaus, M. B. (2025). Enhancement of Machine Learning Algorithm in Fine-grained Sentiment Analysis Using the Ensemble. *ECTI Transactions on Computer and Information Technology (ECTI-CIT)*, *19*(2), 159–167. https://doi.org/10.37936/ecti-cit.2025192.257815

Azis, H., Purnawansyah, P., Fattah, F. & Putri, I. P. (2020). Performa Klasifikasi K-NN dan Cross Validation Pada Data Pasien Pengidap Penyakit Jantung. *ILKOM Jurnal Ilmiah*, *12*(2), 81–86. https://doi.org/10.33096/ilkom.v12i2.507.81-86

Bakri, R., Astuti, N. P. & Ahmar, A. S. (2022). 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*, *4*(2), 259–265. https://doi.org/10.35877/454ri.asci1581

Chamorro-Atalaya, O., Arévalo-Tuesta, J., Balarezo-Mares, D., Gonzáles-Pacheco, A., Mendoza-León, O., Quipuscoa-Silvestre, M., Tomás-Quispe, G. & Suarez-Bazalar, R. (2023). K-Fold Cross-Validation through Identification of the Opinion Classification Algorithm for the Satisfaction of University Students. *International Journal of Online and Biomedical Engineering*, *19*(11), 140–158. https://doi.org/10.3991/ijoe.v19i11.39887

Chen, L., Sun, X., Li, Y., Jaseemuddin, M. & Kazi, B. U. (2024). Automated Hyperparameter Tuning and Ensemble Machine Learning Approach for Network Traffic Classification. *IEEE International Symposium on Broadband Multimedia Systems and Broadcasting, BMSB*, 1–6. https://doi.org/10.1109/BMSB62888.2024.10608236

Chopannejad, S., Roshanpoor, A. & Sadoughi, F. (2024). Attention-assisted hybrid CNN-BILSTM-BiGRU model with SMOTE–Tomek method to detect cardiac arrhythmia based on 12-lead electrocardiogram signals. *Digital Health*, *10*, 1–20. https://doi.org/10.1177/20552076241234624

Co, J. & Casillano, N. F. (2021). Predicting On-time Graduation based on Student Performance in Core Introductory Computing Courses using Decision Tree Algorithm. *Jurnal Pendidikan Progresif*, *11*(3), 650–658. https://doi.org/10.23960/jpp.v11.i3.202116

Dina Amalia Putri, Naza Sefti Prianita & Elkin Rilvani. (2025). Penerapan Metode C4.5 dan K-Nearest Neighbor untuk Klasifikasi Kelulusan Mahasiswa Berdasarkan Data Akademik. *Jupiter: Publikasi Ilmu Keteknikan Industri, Teknik Elektro Dan Informatika*, *3*(4), 256–267. https://doi.org/10.61132/jupiter.v3i4.1032

Dwinanda, M. W., Satyahadewi, N. & Andani, W. (2023). Classification of Student Graduation Status Using XGBoost Algorithm. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, *17*(3), 1785–1794. https://doi.org/10.30598/barekengvol17iss3pp1785-1794

Erlin, E., Desnelita, Y., Nasution, N., Suryati, L. & Zoromi, F. (2022). Dampak SMOTE terhadap Kinerja Random Forest Classifier berdasarkan Data Tidak seimbang. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, *21*(3), 677–690. https://doi.org/10.30812/matrik.v21i3.1726

Gupta, V. & Rattan, P. (2023). Improving Twitter Sentiment Analysis Efficiency with SVM-PSO Classification and EFWS Heuristic. *Procedia Computer Science*, *230*, 698–715. https://doi.org/10.1016/j.procs.2023.12.125

*name of corresponding author

Hasibuan, T. H. & Mahdiana, D. (2023). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Algoritma C4.5 Pada Uin Syarif Hidayatullah Jakarta. *SKANIKA: Sistem Komputer Dan Teknik Informatika*, *6*, 61–74. https://doi.org/10.36080/skanika.v6i1.2976

Herianto, Kurniawan, B., Hartomi, Z. H., Irawan, Y. & Anam, M. K. (2024). Machine Learning Algorithm Optimization using Stacking Technique for Graduation Prediction. *Journal of Applied Data Sciences*, *5*(3), 1272–1285. https://doi.org/10.47738/jads.v5i3.316

Junaidi, S., Anggela, R. V. & Fadhli, I. (2023). Prediksi Kelulusan Tepat Waktu Mahasiswa Menggunakan Metode Data Mining Dengan Algoritma Naïve Bayes. *Jurnal Edik Informatika*, *9*(2), 65–73. https://doi.org/10.22202/ei.2023.v9i2.7324

Junaidi, S., Anggela, R. V. & Kariman, D. (2024). Klasifikasi Metode Data Mining untuk Prediksi Kelulusan Tepat Waktu Mahasiswa dengan Algoritma Naïve Bayes, Random Forest, Support Vector Machine (SVM) dan Artificial Neural Nerwork (ANN). *Journal of Applied Computer Science and Technology*, *5*(1), 109–119. https://doi.org/10.52158/jacost.v5i1.489

Latief, M. A., Nabila, L. R., Miftakhurrahman, W., Ma'rufatullah, S. & Tantyoko, H. (2024). Handling Imbalance Data using Hybrid Sampling SMOTE-ENN in Lung Cancer Classification. *International Journal of Engineering and Computer Science Applications (IJECSA)*, *3*(1), 11–18. https://doi.org/10.30812/ijecsa.v3i1.3758

Li, H. (2024). Machine Learning-based Voting Classifier for Improving Sentiment Analysis on Twitter Data. *Transactions on Computer Science and Intelligent Systems Research*, *5*, 2960–2238. https://doi.org/10.62051/nfkz3035

Mehta, S. (2023). Playing Smart with Numbers: Predicting Student Graduation Using the Magic of Naive Bayes. *International Transactions on Artificial Intelligence (ITALIC)*, *2*(1), 60–75. https://doi.org/10.33050/italic.v2i1.405

Moerdyanto, O. P. & Nuryana, I. K. D. (2023). Prediksi Kelulusan Tepat Waktu Menggunakan Pendekatan Pohon Keputusan Algoritma Decision Tree. *Journal of Informatics and Computer Science*, *5*(1), 90–96. https://doi.org/10.26740/jinacs.v5n01.p90-96

Mubarak, M. M. R., Chrisnanto, Y. H. & Sabrina, P. N. (2023). Implementation of Random Forest Using Smote and Smoteenn in Customer Churn Classification in E-Commerce. *Enrichment: Journal of Multidisciplinary Research and Development*, *1*(8), 463–477. https://doi.org/10.55324/enrichment.v1i8.69

Omotehinwa, T. O. & Oyewola, D. O. (2023). Hyperparameter Optimization of Ensemble Models for Spam Email Detection. *Applied Sciences (Switzerland)*, *13*(3), 1–17. https://doi.org/10.3390/app13031971

Pavitha, N. & Sugave, S. (2022). Ensemble Approach with Hyperparameter Tuning for Credit Worthiness Prediction. *2022 IEEE 3rd Global Conference for Advancement in Technology, GCAT 2022*, 1–6. https://doi.org/10.1109/GCAT55367.2022.9971879

Pirapong, P. I., Thiradet, T. S. & Sayan, S. K. (2024). Enhancing SVM Classification of Breast Cancer Using Dual-Stage PSO Optimization. *ACM International Conference Proceeding Series*, 153–157. https://doi.org/10.1145/3674658.3674683

Prayitno, J., Saputra, B. & Waluyo, R. (2021). Data Mining Implementation with Algorithm C4.5 for Predicting Graduation Rate College studentid 2 * corresponding author. *Journal of Applied Data Sciences*, *2*(3), 74–83. https://doi.org/10.47738/jads.v2i3.37

Prayoga, I., Dwifebri p, M. & Adiwijaya. (2023). Sentiment Analysis on Indonesian Movie Review Using KNN Method With the Implementation of Chi-Square Feature Selection. *Jurnal Media Informatika Budidarma*, *7*(1), 369–375. https://doi.org/10.30865/mib.v7i1.5522

Putra, M. & Erwin Harahap. (2024). Machine Learning pada Prediksi Kelulusan Mahasiswa Menggunakan Algoritma Random Forest. *Jurnal Riset Matematika*, *4*(2), 127–136. https://doi.org/10.29313/jrm.v4i2.5102

Rachardian, S. & Sediyono, E. (2024). Prediksi kelulusan tepat waktu mahasiswa untuk pemantauan program studi menggunakan metode data mining. *AITI: Jurnal Teknologi Informasi*, *21*(2), 168–182. https://doi.org/10.24246/aiti.v21i2.168-182

Riadi, I., Umar, R. & Anggara, R. (2024). Prediksi Kelulusan Tepat Waktu Berdasarkan Riwayat Akademik Menggunakan Metode K-Nearest Neighbor. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, *11*(2), 249–256. https://doi.org/10.25126/jtiik.20241127330

Saputra, A., Arita Fitri, T., Karpen & Susanti. (2023). Penerapan Data Mining Algortima C4.5 Dalam Memprediksi Predikat Kelulusan Mahasiswa Di Politeknik Kampar. *SATIN-Sains Dan Teknologi Informasi*, *9*, 149–157. https://doi.org/10.33372/stn.v9i1.990

Sari, J. S. I., Umar, E. & Momo, L. L. (2024). Prediksi Kelulusan Mahasiswa Tepat Waktu Menggunakan Metode Naïve Bayes Dan Decision Tree Pada Universitas Stella Maris Sumba. *Journal Of Informatics And Busisnes*, *3*(2), 362–368.

Suandi, F., Anam, M. K., Firdaus, M. B., Fadli, S., Lathifah, L., Yumami, E., Saleh, A. & Hasibuan, A. Z. (2024). Enhancing Sentiment Analysis Performance Using SMOTE and Majority Voting in Machine Learning

*name of corresponding author

Algorithms. *International Conference on Applied Engineering*, 126–138. https://doi.org/10.2991/978-94-6463-620-8_10

Susanto, N. W. & Suparwito, H. (2023). SVM-PSO Algorithm for Tweet Sentiment Analysis #BesokSenin. *Indonesian Journal of Information Systems (IJIS)*, *6*(1), 36–47. https://doi.org/10.24002/ijis.v6i1.7551

Van FC, L. L., Anam, M. K., Bukhori, S., Mahamad, A. K., Saon, S. & Nyoto, R. L. V. (2025). The Development of Stacking Techniques in Machine Learning for Breast Cancer Detection. *Journal of Applied Data Sciences*, *6*(1), 71–85. https://doi.org/10.47738/jads.v6i1.416

Wahyudi, A., Kusrini & Wibowo, F. W. (2023). *Predicting On-Time Graduation Of Students Using Decision Tree And Naïve Bayes Methods*. *14*(2), 132–138. https://doi.org/10.59737/jpi.v14i2.276

Yin, J. & Li, N. (2022). Ensemble learning models with a Bayesian optimization algorithm for mineral prospectivity mapping. *Ore Geology Reviews*, *145*, 1–19. https://doi.org/10.1016/j.oregeorev.2022.104916

*name of corresponding author