

Optimization of Student Graduation Timeliness Prediction at Royal University Using Particle Swarm Optimization-Based Naive Bayes Algorithm

Rudi Hermawan, Muhammad Iqbal

Abstract

Graduation timeliness is a primary indicator in the accreditation assessment of higher education institutions. Royal University faces complexities in monitoring student graduation trends due to the vast volume of heterogeneous academic data. This study aims to enhance the performance of graduation timeliness prediction by implementing Feature Selection optimization using the Particle Swarm Optimization (PSO) algorithm on Naive Bayes classification. The dataset successfully extracted comprises 4,909 observations of alumni academic records, utilized to ensure the validity of target label attribution (On-time vs. Delayed). The PSO optimization process is employed to reduce irrelevant features that may degrade model performance. The research results indicate that the integration of PSO significantly improves accuracy from 79.55% in the standard Naive Bayes algorithm to 87.16% in the optimized model. The dominant selected features include School Origin, Final GPA, and Attendance Rate. These findings prove that the hybrid PSO-Naive Bayes model is effective as an Early Warning System for university management to improve the timely graduation ratio.

Keywords: *Naive Bayes, Particle Swarm Optimization, Graduation Timeliness, Feature Selection, Royal University*

Rudi Hermawan¹

¹Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: rudirk60@gmail.com¹

Muhammad Iqbal²

²Information Technology, Universitas Pembangunan Panca Budi, Indonesia
e-mail: muhammadiqbal@dosen.pancabudi.ac.id²

2nd International Conference on Islamic Community Studies (ICICS)

Theme: History of Malay Civilisation and Islamic Human Capacity and Halal Hub in the Globalization Era

<https://proceeding.pancabudi.ac.id/index.php/ICIE/index>

Introduction

Higher education institutions are required to guarantee the quality of their graduates, one of which is measured through the parameter of graduation timeliness. At Royal University, the increasing student population demands a more intelligent monitoring system to identify students at risk of academic delay [1]. The utilization of Machine Learning in the realm of education (Educational Data Mining) has been proven to provide deep predictive insights regarding student retention and academic performance [2].

However, the primary challenge in academic data classification is the presence of redundant or irrelevant attributes (noise), which, if included, can diminish prediction accuracy [3]. The Naive Bayes algorithm is frequently chosen for its ease of implementation and the assumption that the features within the data are independent of one another [4]. Recent research on individual academic characteristics indicates that targeted feature selection within predictive models, such as Naive Bayes, provides a robust framework for accurately identifying student graduation timelines and improving institutional accreditation [5]. Therefore, this study proposes a hybrid approach using Particle Swarm Optimization (PSO) to optimize feature selection for Naive Bayes to achieve maximum accuracy at Royal University.

Literature Review

The Naive Bayes algorithm is a probability-based machine learning method derived from Bayes' Theorem, operating under the strong assumption of feature independence within a class. In the field of Educational Data Mining (EDM), such automated techniques are essential for comprehension of learners and their environments, enabling institutions to predict student success or failure in required courses effectively [2]. The Bayes equation is expressed as follows:

$$P(H|E) = \frac{P(E|H) \cdot P(H)}{P(E)} \dots (1)$$

Despite its simplicity, Naive Bayes exhibits robust performance in classifying tabular academic data, provided the dataset has undergone a rigorous feature selection process using metaheuristic algorithms such as PSO. This effectiveness is evidenced by recent studies in 2024 which achieved perfect accuracy scores through this hybrid approach [4].

PSO is a metaheuristic optimization algorithm inspired by the social behavior of bird flocking (swarming) [6]. In the search space, each "particle" represents a potential solution (feature subset) moving based on its position and velocity. In this study, PSO acts as a feature selection agent to identify the most significant subset of variables [7]. Velocity updates for particles are governed by the following equation:

$$v_{id}^{t+1} = w \cdot v_{id}^t + c_1 \cdot r_1 \cdot (pbest_{id} - x_{id}^t) + c_2 \cdot r_2 \cdot (gbest_{id} - x_{id}^t) \dots (2)$$

Recent research in 2024 indicates that academic variables such as early-semester GPA and attendance rates are the strongest predictors of on-time graduation [8]. Furthermore, recent research conducted in 2025 indicates that the integration of the Random Forest algorithm with Particle Swarm Optimization (PSO)-based feature selection for predicting student graduation achieves exceptional performance, reaching an accuracy rate of up to 99.63%. This demonstrates that metaheuristic methods are highly effective in enhancing the generalization capabilities of classification models when applied to complex academic datasets [9].

Research Methodology

This research follows a classification framework consisting of several systematic stages:

2.1 Data Collection

The dataset was extracted directly from the Royal University database by integrating the mhs (student) and krs (course enrollment) tables. This study utilizes a dataset of 4,909 graduate records that have undergone an extraction process to ensure the validity of target label classifications, specifically on-time and delayed graduation.

2.2 Data Preprocessing

This stage involves data cleaning, transformation, and feature selection to improve model performance [10]. The steps are:

- Cleaning: Removing data with missing student IDs (NIM) and handling missing values.
- Feature Engineering: Recalculating GPA and total credits from KRS transaction data.
- Transformation: Converting categorical data into numerical formats through Label Encoding.

2.3 Naive Bayes Algorithm

This method works based on Bayes' theorem with the assumption of feature independence [6]. The implementation of metaheuristic algorithms in optimizing Naive Bayes classification models has been proven to significantly increase efficiency and accuracy in various academic data case studies [5].

2.4 Particle Swarm Optimization (PSO)

PSO is a metaheuristic optimization algorithm inspired by the social behavior of bird flocks searching for food [8]. In the context of feature selection, each particle moves through the search space based on its personal best (pbest) and global best (gbest) to minimize classification errors and select the most relevant features [11].

Results

Experiments were conducted on 4,909 data points with a split of 70% for training data and 30% for testing data:

3.1 Feature Optimization Results

Based on the PSO iteration results over 20 stages, it was found that feature reduction is crucial in improving model quality. From the 6 initial features, PSO successfully identified 3 dominant features: School Origin, Final GPA, and Attendance Rate for Semesters 1-4. Demographic variables, such as Parents' Occupation, were eliminated as they were deemed not to contribute significantly to the predictive power [9].

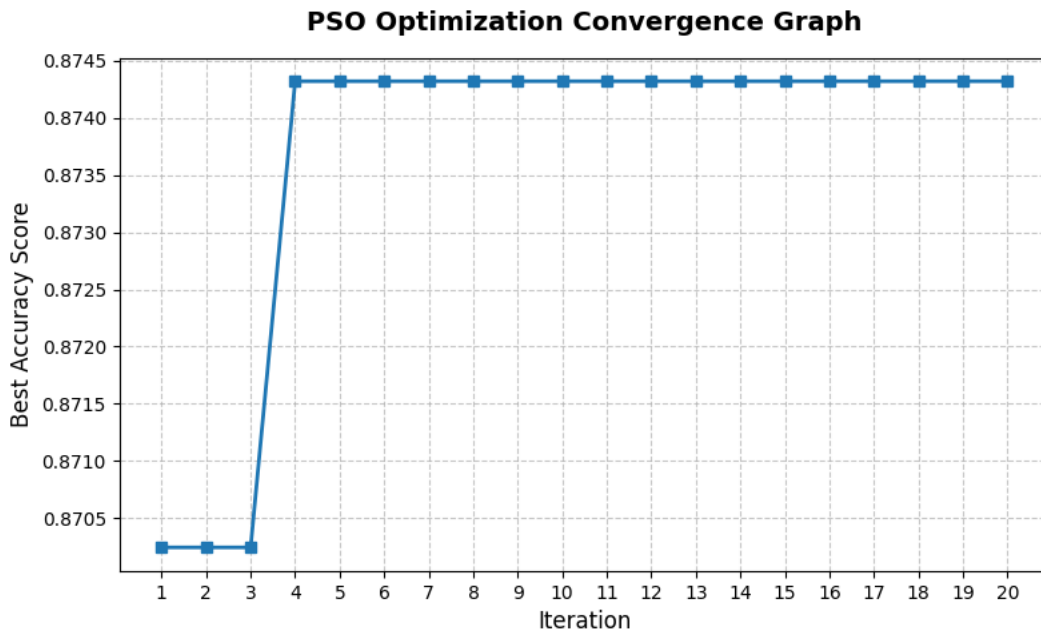


Figure 1. PSO Optimization Convergence Graph Demonstrating Accuracy Improvement Across Iterations

The optimization progress visualized in Figure 1 illustrates the convergence of the PSO algorithm. The graph depicts a significant increase in accuracy during the initial iterations before reaching a stable peak of 87.16%. This stability indicates that the algorithm efficiently identified the optimal feature subset required for high-precision classification.

3.2 Performance Comparison

After testing, the model showed a significant performance improvement following the implementation of PSO compared to the Naive Bayes model without optimization:

Table 1. Model Comparison Table

No.	Algorithm	Accuracy	Precision	Recall	F1-Score
1	Standard Naive Bayes	79,55%	79,69%	79,55%	79,61%
2	PSO-Naive Bayes	87,16%	88,45%	87,16%	86,61%

The increase of 7.61% proves that PSO successfully filtered out data noise. Confusion Matrix analysis shows that this hybrid model is much more accurate in detecting students at risk of graduating late, allowing for earlier intervention by the study program management.

3.3. Confusion Matrix Analysis

The detailed classification performance of the optimized model is presented through a confusion matrix to evaluate the specific prediction outcomes.

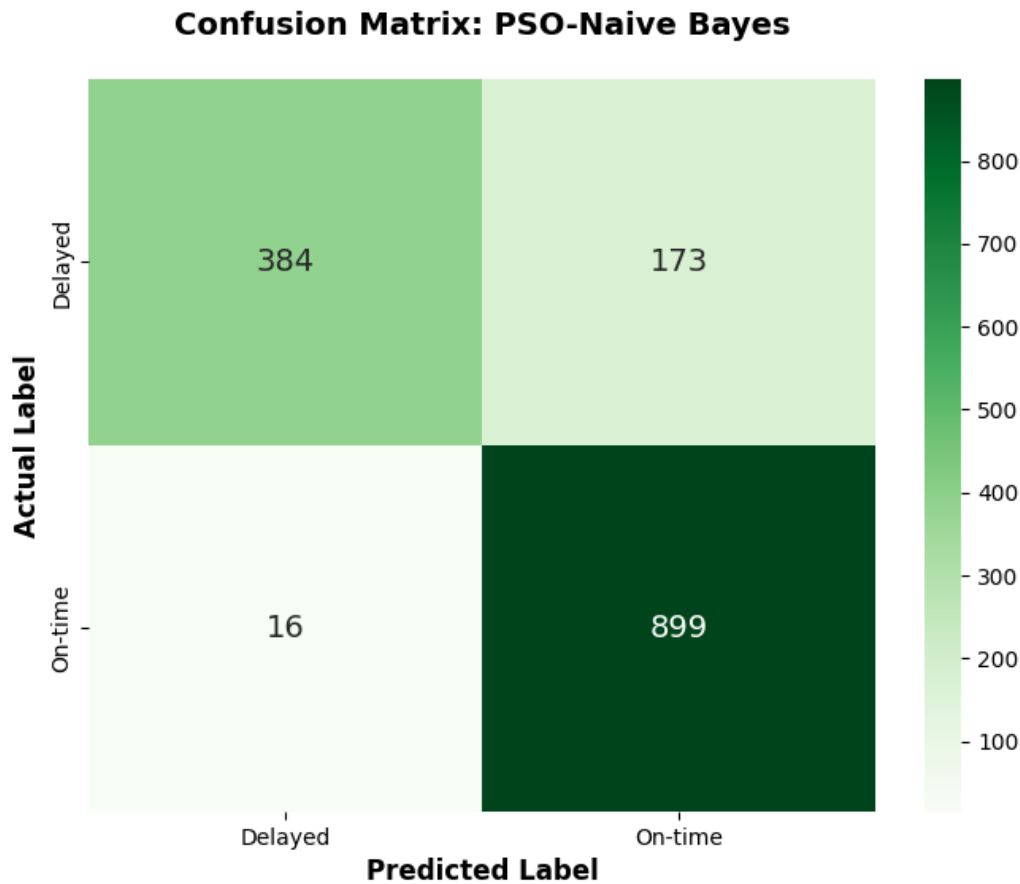


Figure 2. Confusion Matrix of the Optimized PSO-Naive Bayes Model (Delayed vs. On-time)

Conclusion

This study successfully demonstrates that the integration of Particle Swarm Optimization (PSO) with the Naive Bayes algorithm is effective in improving the prediction performance of graduation timeliness at Royal University, reaching an accuracy of 87.16%. The feature selection results show that internal academic factors (GPA and attendance) are far more decisive than demographic factors. The implementation of this model can support university management in providing a more accurate early warning system to increase the timely graduation ratio.

References

- [1] M. Y. L. Ijlal, A. Setiawan, and D. L. Fithri, "Prediksi Ketereratan Karier Lulusan berdasarkan Aspek Akademik dan Non-Akademik berbasis Machine Learning," *Sist. J. Sist. Inf.*, vol. 14, no. 4, pp. 2540–9719, 2025, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>.
- [2] Q. Hai and C. Wang, "Optimizing Student Performance Prediction: A Data Mining Approach with MLPC Model and Metaheuristic Algorithm," *Int. J. Adv. Comput. Sci. Appl.*, vol. 15, no. 4, pp. 55–71, 2024, doi: 10.14569/IJACSA.2024.0150407.
- [3] Z. L. Aripin, A. Rahmatulloh, and E. Haerani, "Smart Payment Predictions: Evaluating Naive Bayes, XGBoost, and Decision Tree Models for Timely Tuition Fees," *Int. J. Informatics Vis.*, vol. 10, no. 1, pp. 27–32, 2026, doi: 10.62527/joiv.10.1.3717.
- [4] E. F. Wati, E. Sunita Perangin-Angin, and A. P. Sari, "Improved Naive Bayes Algorithm with Particle Swarm Optimization to Predict Student Graduation," *Int. J. Inf. Syst. Technol. Akreditasi*, vol. 7, no. 6, pp. 386–391, 2024.
- [5] K. J. Atmaja, I. P. Y. Indrawan, I. M. Dwi, P. Asana, and I. K. Wawan, "Journal of Computer Networks , Architecture and High Performance Computing Naïve Bayes-based

- Student Graduation Prediction Model : Effectiveness and Implementation to Improve Timely Graduation,” *J. Comput. Networks, Archit. High Perform. Comput.*, vol. 6, no. 3, pp. 1442–1450, 2024.
- [6] W. Shafqat, S. Malik, K. T. Lee, and D. H. Kim, “Pso based optimized ensemble learning and feature selection approach for efficient energy forecast,” *Electron.*, vol. 10, no. 18, 2021, doi: 10.3390/electronics10182188.
- [7] L. Santoso and Priyadi, “Comparative Study of Feature Engineering Techniques for Predictive Data Analytics,” *J. Technol. Informatics Eng.*, vol. 3, no. 2, pp. 417–435, 2024, doi: 10.51903/jtie.v3i2.225.
- [8] F. Tatimu, V. R. Palilingan, and I. Rianto, “Analisis Klasifikasi Kelulusan Mahasiswa menggunakan Algoritma K-Nearest Neighbor di Fakultas Teknik Universitas Negeri Manado,” *J. Educ. Method Technol.*, vol. 3, no. 1, pp. 117–123, 2025.
- [9] A. Rahman, D. Mahdiana, and A. Fauzi, “Predicting Student On-Time Graduation Using Particle Swarm Optimization and Random Forest Algorithms,” *Indones. J. Artif. Intell. Data Min.*, vol. 8, no. 1, p. 161, 2025, doi: 10.24014/ijaidm.v8i1.33577.
- [10] R. D. Deleña *et al.*, “Predicting student retention: A comparative study of machine learning approach utilizing sociodemographic and academic factors,” *Syst. Soft Comput.*, vol. 7, no. July, 2025, doi: 10.1016/j.sasc.2025.200352.
- [11] A. Nowakowski, L. Strak, and W. Wiczorek, “MAB-optimized binary PSO-based feature selection for enhanced classification performance,” *Procedia Comput. Sci.*, vol. 225, pp. 4264–4273, 2023, doi: 10.1016/j.procs.2023.10.423.