

Comparative Analysis of Multi-Classifer Models with Resampling Techniques for Imbalanced Student Graduation Prediction

Irmawati Carolina¹, Resti Lia Andharsaputri¹, Suharjanti¹, Titin Prihatin¹, Hafis Nurdin¹

¹Universitas Bina Sarana Informatika, Jakarta, Indonesia

ARTICLE INFORMATION

Artikel History:

Received: 30-01-2026

Revised: 27-02-2026

Accepted: 03-03-2026

Available Online: 31-03-2026

Keyword:

Borderline-SMOTE
Imbalanced Data
Machine Learning
Class Imbalance Handling
Student Graduation

ABSTRACT

Student graduation prediction supports early academic intervention but commonly suffers from class imbalance, where on-time graduates dominate the dataset. This study evaluates five classifiers—Random Forest (RF), XGBoost, Logistic Regression (LR), *k*-Nearest Neighbors (*k*-NN), and Gaussian Naïve Bayes (GNB)—under five class-imbalance handling scenarios: Baseline (no resampling), Random Undersampling (RUS), SMOTE, ADASYN, and Borderline-SMOTE. Experiments were conducted on 796 student records (10 attributes) with an imbalanced distribution (634 on-time vs. 162 not on-time; ratio 1:3.9) using Stratified 5-Fold Cross-Validation. Performance was assessed using confusion-matrix metrics and AUC-ROC to reflect minority-class detection. Under baseline, RF achieved the highest accuracy (0.873) but limited minority recall (0.573), confirming majority-class bias. Resampling consistently improved minority recall across models; for example, LR recall increased to 0.802 with RUS, while GNB reached 0.833 with ADASYN, although accuracy decreased due to the sensitivity-specificity trade-off. Overall, RF and XGBoost showed the most stable discrimination across resampling scenarios based on AUC (RF: 0.870–0.883; XGBoost: 0.847–0.866). The main contribution is a systematic, reproducible comparative evaluation of classifier-resampling combinations for imbalanced graduation prediction, providing practical guidance for selecting robust models to identify students at risk of delayed graduation.

Corresponding Author:

Irmawati Carolina,
Sistem Informasi,
Universitas Bina Sarana Informatika,
Jl. Kramat Raya No.98, Senen, Jakarta, Indonesia, 10450,
Email: irmawati.imc@bsi.ac.id

INTRODUCTION

On-time student graduation is a crucial indicator in evaluating academic performance in higher education institutions and is often used as an important component in accreditation and quality assurance assessments (Sumarna et al., 2024). High rates of delayed graduation not only negatively affect institutional reputation but also result in economic and social consequences for students and academic administrators (Kahfi et al., 2024). Therefore, the ability to identify students who are at risk of failing to graduate on time at an early stage has become an important objective in data-driven academic decision-making (Nurdin et al., 2024).

Along with the rapid development of information technology, machine learning techniques have been widely applied to predict student graduation outcomes using academic and non-academic data (Nurdin et al., 2023). Previous studies have demonstrated that classification algorithms such as Random Forest, Logistic Regression, and Naïve Bayes are capable of modeling student graduation patterns with relatively good accuracy (Pangesti et al., 2024). However, most existing studies primarily focus on improving overall accuracy and tend to overlook a fundamental issue commonly found in graduation datasets, namely imbalanced class distribution (Iqbal et al., 2025).

DOI: <https://doi.org/10.31294/p.v28i1.11976>



This work is licensed under a [Creative Commons Attribution-ShareAlike 4.0 International License](https://creativecommons.org/licenses/by-sa/4.0/)

Imbalanced data occurs when the number of instances in one class significantly exceeds those in another class (Fulazzaky et al., 2024). In the context of student graduation prediction, students who graduate on time typically constitute the majority class, while students who fail to graduate on time represent a minority class (Salmi et al., 2024). In this study, 634 students graduated on time, whereas only 162 students did not, resulting in an imbalanced class ratio of approximately 1:3.9. Such imbalance can bias machine learning models toward the majority class, leading to misleadingly high accuracy while reducing the model's ability to correctly identify students at risk of delayed graduation (Gurcan & Soylu, 2024). Several studies have confirmed that imbalanced class distributions often result in poor minority class recall and unreliable predictive performance (Altalhan et al., 2025).

To address class imbalance, various strategies have been proposed in the machine learning literature. These approaches can generally be categorized into data-level methods, such as resampling techniques, and algorithm-level methods, including cost-sensitive learning and ensemble strategies (Huang & Dai, 2021). Among these, resampling techniques are widely adopted due to their flexibility and ease of integration with different classification algorithms (Yang et al., 2024). Synthetic oversampling methods, particularly SMOTE and its variants, remain dominant in imbalanced learning research and have shown promising results across multiple domains (Siregar & Arifin, 2024). However, the effectiveness of these techniques is highly dependent on data characteristics and the classification algorithm employed (Ayodele, 2023).

Empirical evidence from the healthcare domain shows that combining resampling techniques with ensemble-based models can significantly improve recall and AUC values, especially for minority classes (Muraru et al., 2024). Despite these promising findings, the application of resampling-based approaches in educational data mining, particularly for student graduation prediction, remains less systematically explored compared to healthcare and financial domains (Glučina et al., 2023). At the national level, several studies have explored machine learning for academic performance prediction (Ridwansyah et al., 2024). Nevertheless, most of these studies rely on a single classifier or do not comprehensively examine the impact of different resampling techniques on model performance (Rukiastindari et al., 2024).

Based on these gaps, this study addresses the following research question:

Which combination of classification algorithm and class-imbalance handling technique provides the most stable and reliable predictive performance for identifying students at risk of delayed graduation?

To guide empirical validation, the following hypotheses are formulated:

H1: Applying class-imbalance handling techniques significantly improves minority-class recall and F1-score compared to the baseline (no resampling) condition.

H2: Ensemble-based classifiers (Random Forest and XGBoost) demonstrate more stable discriminatory performance across imbalance-handling scenarios, as reflected by consistently higher AUC values.

To answer this question, this study proposes a systematic comparative evaluation of five classification algorithms—Random Forest, XGBoost, Logistic Regression, k-Nearest Neighbors, and Gaussian Naïve Bayes—under five imbalance-handling scenarios: baseline (no resampling), Random Undersampling, SMOTE, ADASYN, and Borderline-SMOTE. By focusing on minority-class detection and performance stability rather than accuracy alone, this research aims to provide practical guidance for higher education institutions in designing more reliable early-warning systems for delayed graduation..

This study contributes at an applied-comparative level by delivering a controlled evaluation of five widely used classifiers combined with five class-imbalance handling scenarios on a real higher-education dataset. Unlike prior graduation-prediction works that often report accuracy-dominant results or evaluate a single model, this research emphasizes minority-class detection using AUC-ROC and F1-score and reports the stability of ensemble methods across resampling variants. The findings provide actionable guidance for academic decision support systems, specifically for early identification of students at risk of delayed graduation.

RESEARCH METHOD

This research focuses on analyzing the capabilities of various resampling techniques combined with machine learning models in building a student graduation prediction system based on available academic data in the case of imbalanced data. In general, the research was conducted through the stages of sourcing and collecting student graduation datasets, followed by data preprocessing, imbalanced data handling, classification modeling, model evaluation, and methodological contributions. Furthermore, various resampling techniques were applied in a validation scheme to address the problem of class imbalance that commonly occurs in student graduation data, where the number of students who graduate on time tends to be more dominant than the minority class. Various classification models were trained and evaluated using Stratified 5-Fold Cross-Validation to ensure stable and unbiased performance estimation. To prevent data leakage, all resampling techniques (SMOTE, ADASYN, Borderline-SMOTE, and Random Undersampling) were applied exclusively to the training partition within each fold, while the corresponding test partition retained the original class distribution. Feature scaling was also fitted only on the

training data and subsequently applied to the test data.. The final stage of the research focused on a comprehensive analysis of the evaluation results to identify the most effective combination of resampling techniques and classification algorithms in improving student graduation prediction performance. All stages of this research are presented visually in Fig. 1.

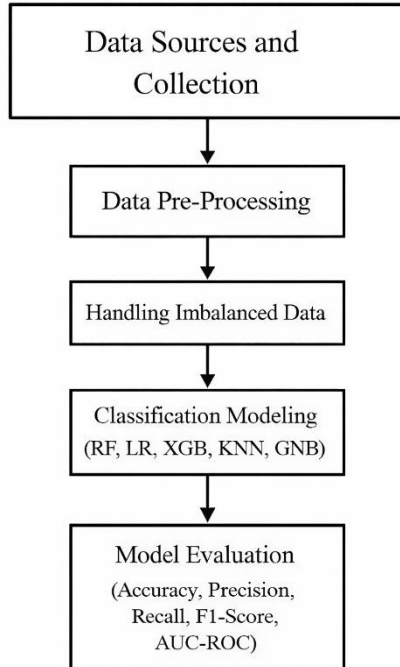


Figure 1. Research Method Flowchart

Data Sources and Collection

The research data was obtained from a private university academic database, which includes demographic information, academic data, and student graduation status as the target variable, which indicates whether students graduated on time or not, with a total of 796 student records. The data collection process was conducted in a structured manner, with attention to completeness and consistency (Geron, 2023).

Data Pre-Processing

The pre-processing stage included data cleaning, handling missing values, and normalizing numeric attributes, as previously conducted by researchers. The dataset used in this study was clean and curated. The pre-processing stage focused on checking data consistency, analyzing class distributions, and performing necessary feature transformations prior to the modeling process. Next, the class distributions were analyzed to identify the level of data imbalance.

The data used in this study revealed significant class imbalance in the student graduation target variable. Of the 796 student data sets, 162 students (approximately 20.4%) were categorized as not graduating on time, while 634 students (approximately 79.6%) were categorized as graduating on time. The class imbalance ratio of approximately 1:3.9 indicates

that the number of students graduating on time significantly outnumbers those not graduating on time.

This imbalance has the potential to bias machine learning models, such as Logistic Regression and Random Forest, toward the majority class. As a result, although the model can produce high overall accuracy, its ability to identify students at risk of failing to graduate on time is low. In certain scenarios, a classification model without data imbalance management can even produce very low recall in the minority class, resulting in a large proportion of students who do not graduate on time being misclassified as graduating on time. To address the problem of class imbalance in student graduation data, resampling techniques such as SMOTE can be applied to the target graduation variable. SMOTE works by generating new synthetic samples in the minority class, namely students who do not graduate on time, through a linear interpolation process between one minority sample and its nearest neighbor in the feature space. This approach aims to improve the representation of the minority class so that the classification model can learn the pattern of failure risk more effectively.

In addition to SMOTE, advanced oversampling techniques such as ADASYN and Borderline-SMOTE were also used in this study due to their ability to tailor the synthetic sample generation process based on the classification difficulty of the minority class. ADASYN specifically focuses synthetic data generation on minority samples located in complex decision regions, while Borderline-SMOTE focuses on minority samples located around the class boundaries. This approach is expected to increase the model's sensitivity in identifying students at risk of not graduating on time without reducing the model's overall performance (Wibowo & Fatichah, 2021).

Handling Imbalanced Data

This study used five data handling scenarios. The first scenario is the baseline, which uses the original data without resampling techniques as a benchmark. This approach illustrates the model's natural performance when the class distribution is imbalanced. The second scenario uses Random Undersampling (RUS), which works by randomly reducing the number of samples in the majority class until the class distribution becomes more balanced. Mathematically, RUS can be expressed as the process of selecting a subset $D^l \subset D$ from the majority class, such that $D^l_{majoritas} \approx D_{minoritas}$ (Carvalho et al., 2025). While simple, this method risks losing important information from the majority class.

Next, the Synthetic Minority Over-sampling Technique (SMOTE) is used to generate synthetic samples in the minority class. New samples are formed through linear interpolation between one minority data x_{ixi} and one of its nearest neighbors $x_{nnx}_{\{nn\}x_{nn}}$, as shown in Equation (1):

$$x_{new} = x_i + \lambda(x_{nn} - x_i), \lambda \in (0, 1) \quad (1)$$

This approach aims to enrich the representation of minority classes without directly duplicating data.

The Adaptive Synthetic Sampling (ADASYN) technique is a development of SMOTE that adjusts the number of synthetic samples based on the classification difficulty of each minority data set (Salifu et al., 2025). ADASYN generates more synthetic samples in areas with high misclassification rates, allowing the model to focus more on learning complex patterns in the minority class (Miftahshudur et al., 2025). The final scenario uses Borderline-SMOTE, which specifically generates synthetic samples from minority data located near the decision boundary between classes. This approach is considered effective in increasing the model's sensitivity to students at risk of failing to graduate on time without excessively increasing the overlap between classes (Nouas et al., 2025).

Classification Modeling

This study evaluated five classification algorithms—Random Forest, XGBoost, Logistic Regression, k-Nearest Neighbors (KNN), and Gaussian Naïve Bayes—under various class imbalance handling scenarios. To ensure fair comparison across resampling strategies, fixed hyperparameter settings were adopted based on commonly recommended configurations in prior literature and standard library defaults. No extensive hyperparameter optimization (e.g., GridSearch or RandomSearch) was performed, as the primary objective of this study was to assess the comparative robustness of classifiers under different imbalance-handling techniques rather than to maximize the performance of each individual model. This approach enhances reproducibility and ensures consistent evaluation across experimental conditions.

Random Forest and XGBoost are ensemble methods with good generalization capabilities, but in imbalanced data distributions, the model formation process tends to be more influenced by the majority class (Chen & Guestrin, 2016). Meanwhile, Logistic Regression, a basic statistical model (Geron, 2023). KNN, a distance-based method, and Naïve Bayes, a probabilistic approach, are also vulnerable to data imbalance because the learning and decision-making processes are highly dependent on the distribution of the training data (Fernández et al., 2018).

Model Evaluation

Model evaluation was conducted using a confusion matrix, including accuracy, precision, recall, F1-score, and AUC-ROC. Confusion matrix-based metrics were used to assess the model's ability to classify each class, particularly the minority class. Furthermore, AUC-ROC was used because it provides a comprehensive overview of model performance on imbalanced data, without relying on a specific decision threshold (Çorbacıoğlu & Akse, 2023).

In this study, class label 0 represents students graduating on time, the majority class, while label 1

represents students not graduating on time, the minority class.

RESULTS AND DISCUSSION

After preprocessing and handling data imbalance, model performance was tested using several classification algorithms and predetermined resampling scenarios. Model evaluation was conducted based on the confusion matrix, as described in the methods section, to assess the model's ability to classify student graduation data, particularly for minority classes. The results of this test were used to analyze the effect of the resampling technique on the performance of each algorithm and to compare model performance under baseline conditions (no resampling) and balanced data.

Before discussing the model performance evaluation results, the initial condition of the preprocessed dataset was examined. The dataset was verified to contain no missing values across all attributes; therefore, no data imputation was required, and the dataset could be directly used for model training and testing.

In addition to confirming data completeness, an exploratory analysis was conducted to better understand the relationships among variables. Figure 2 presents the correlation heatmap of all numerical features, illustrating the strength and direction of pairwise relationships within the dataset.

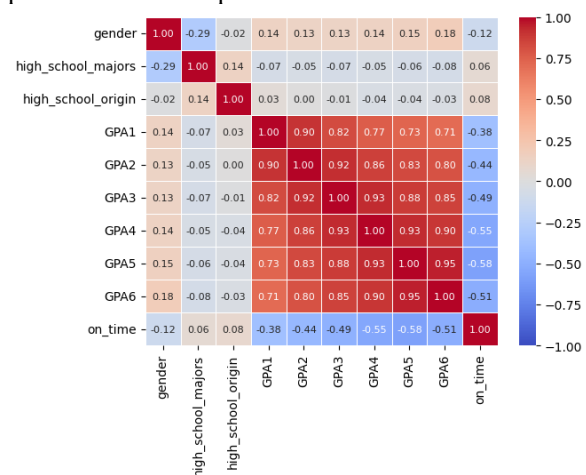


Fig. 2 Graduation Attribute Correlation Matrix

Figure 2 illustrates the pairwise Pearson correlation coefficients among all variables in the dataset. The heatmap reveals that the cumulative GPA variables (GPA1–GPA6) exhibit strong positive correlations with each other, with correlation values ranging from 0.82 to 0.95. This strong interrelationship is expected, as students' academic performance across consecutive semesters tends to be consistent and progressively related.

Regarding the target variable (On Time Graduation), moderate positive correlations are observed particularly with higher-semester GPA variables, such as GPA4 ($r = 0.55$), GPA5 ($r = 0.58$), and GPA6 ($r = 0.51$). This suggests that academic performance in later semesters has a stronger

association with timely graduation compared to early-semester performance.

Meanwhile, demographic variables such as Gender and Major show weak correlations with the target variable ($|r| < 0.20$), indicating limited direct linear influence on graduation outcomes. Similarly, the Entry Path variable demonstrates minimal correlation with most features.

In addition to checking for missing values, an initial dataset analysis was also conducted to examine the class distribution of the student graduation data. This analysis aimed to identify any imbalances in the number of data points between classes that could potentially impact the performance of the classification model. The distribution of data points in each class is shown in Fig. 3.

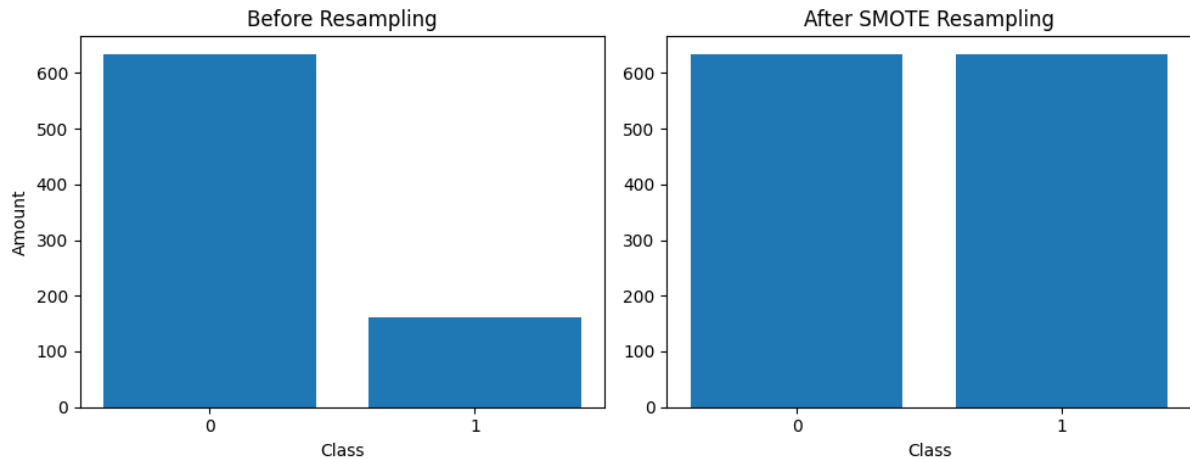


Fig. 3 Distribution Results of Student Graduation Classes

Based on Fig. 3, the distribution of student graduation data before resampling indicated a class imbalance. Class 0, the majority class, had a significantly larger number of data points than Class 1, the minority class: 634 data points for the majority class and 162 data points for the minority class. This significant difference in data size could potentially cause the classification model to learn patterns from the majority class, resulting in low prediction accuracy for the minority class.

After applying the SMOTE method, the data distribution across the two classes became balanced,

with each class having an equal number of data points. This balancing process aims to reduce model bias toward the majority class and improve the model's ability to recognize patterns in the minority class more optimally.

To ensure that the proportion of majority and minority classes is maintained throughout each validation process, this study employed the Stratified K-Fold Cross-Validation technique with a K value of 5. The class distribution in each fold is shown in Fig. 4.

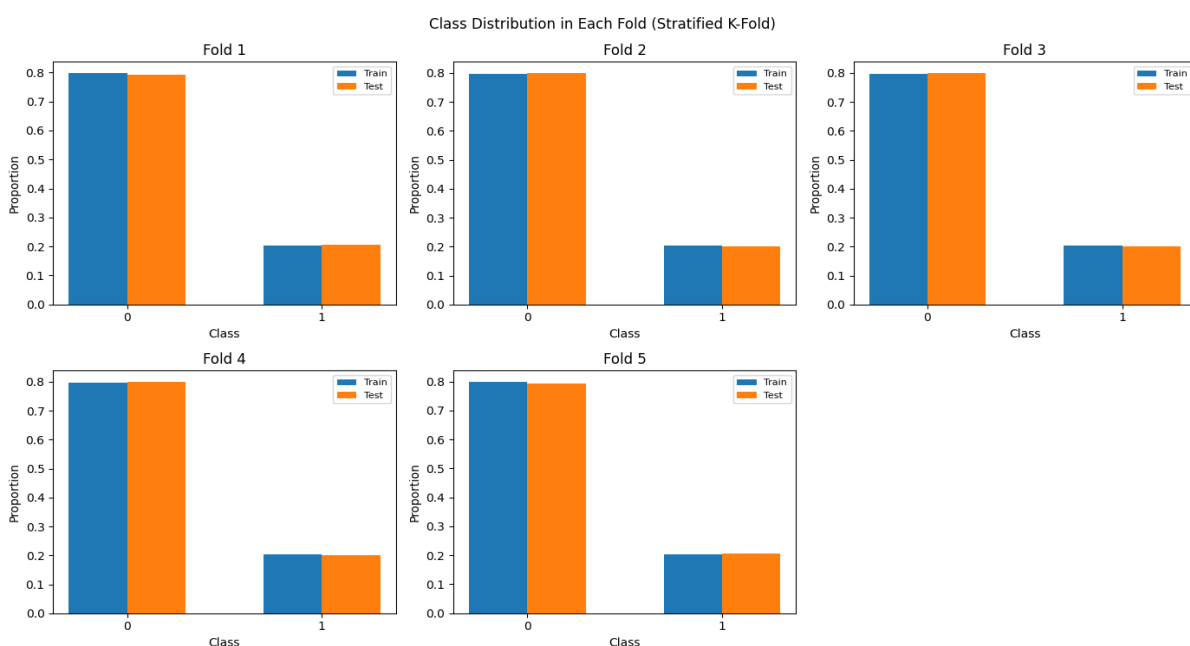


Fig. 4 Class Distribution on Each Fold Using Stratified K-Fold (K = 5)

Fig. 4 shows the class distribution in each fold using the Stratified K-Fold Cross Validation method with $K = 5$ in the study of predicting student graduation on time. The dataset is divided into five folds, where in each fold the data is divided into training data and test data. It can be seen that the proportion of student classes graduating on time (class 0) and not graduating on time (class 1) remains consistent in each fold, which is around 80% for class 0 and 20% for class 1. This indicates that the stratification process successfully

maintains a balanced class distribution, so that model performance evaluation can be carried out objectively and reduces bias due to data imbalance.

After ensuring that the class proportions remained stable in each fold through the application of Stratified K-Fold Cross-Validation, model performance was evaluated using various classification algorithms and predefined resampling scenarios. A summary of the model performance evaluation results based on metrics is presented in Table 1.

Table 1. Model Performance Evaluation Results on Various Classification Algorithms and Resampling Techniques

| Resampling | Model | Accuracy | Precision | Recall | F1-score | AUC |
|--------------------------|---------|----------|-----------|----------|----------|----------|
| Baseline (No Resampling) | RF | 0.873058 | 0.736306 | 0.572538 | 0.640275 | 0.882486 |
| Baseline (No Resampling) | XGBoost | 0.861792 | 0.699835 | 0.554735 | 0.613610 | 0.864443 |
| Baseline (No Resampling) | LR | 0.864261 | 0.740926 | 0.505303 | 0.598411 | 0.888206 |
| Baseline (No Resampling) | GNB | 0.830354 | 0.569521 | 0.727841 | 0.636733 | 0.861674 |
| Baseline (No Resampling) | k-NN | 0.837909 | 0.678514 | 0.406629 | 0.505296 | 0.837714 |
| SMOTE | RF | 0.841690 | 0.594089 | 0.696970 | 0.639267 | 0.879299 |
| SMOTE | XGBoost | 0.837909 | 0.595720 | 0.634848 | 0.609960 | 0.861679 |
| SMOTE | LR | 0.811564 | 0.526185 | 0.751705 | 0.617064 | 0.887428 |
| SMOTE | GNB | 0.798954 | 0.504249 | 0.789394 | 0.614899 | 0.860511 |
| SMOTE | k-NN | 0.782626 | 0.482857 | 0.752841 | 0.586984 | 0.837238 |
| ADASYN | RF | 0.829119 | 0.561640 | 0.709280 | 0.625143 | 0.877403 |
| ADASYN | XGBoost | 0.840432 | 0.596585 | 0.671591 | 0.629263 | 0.866033 |
| ADASYN | LR | 0.782649 | 0.479422 | 0.788826 | 0.595590 | 0.882739 |
| ADASYN | GNB | 0.761250 | 0.452417 | 0.832576 | 0.586236 | 0.860888 |
| ADASYN | k-NN | 0.761297 | 0.451129 | 0.789773 | 0.573703 | 0.838468 |
| Borderline-SMOTE | RF | 0.840440 | 0.593152 | 0.690720 | 0.636160 | 0.870090 |
| Borderline-SMOTE | XGBoost | 0.842948 | 0.607547 | 0.640909 | 0.621188 | 0.862926 |
| Borderline-SMOTE | LR | 0.795212 | 0.499046 | 0.789015 | 0.610144 | 0.885184 |
| Borderline-SMOTE | GNB | 0.758742 | 0.448962 | 0.832576 | 0.583299 | 0.860415 |
| Borderline-SMOTE | k-NN | 0.772547 | 0.465995 | 0.728030 | 0.567651 | 0.825577 |
| Random Undersampling | RF | 0.807775 | 0.519534 | 0.789773 | 0.625494 | 0.872870 |
| Random Undersampling | XGBoost | 0.790197 | 0.488758 | 0.771023 | 0.596364 | 0.846879 |
| Random Undersampling | LR | 0.801502 | 0.508778 | 0.801894 | 0.621912 | 0.883347 |
| Random Undersampling | GNB | 0.797689 | 0.506355 | 0.770644 | 0.609089 | 0.862805 |
| Random Undersampling | k-NN | 0.817783 | 0.539560 | 0.746402 | 0.626175 | 0.847846 |

Table 1 shows that the performance of the classification model in predicting student graduation is significantly influenced by the use of resampling techniques. The evaluation was conducted using five classification algorithms: Random Forest (RF), XGBoost, Logistic Regression (LR), Gaussian Naïve Bayes (GNB), and k-Nearest Neighbor (k-NN), with five scenarios for handling data imbalance: no resampling (baseline), SMOTE, ADASYN, Borderline-SMOTE, and Random Undersampling.

In this study, class label 0 represents students graduating on time as the majority class, while class label 1 represents students not graduating on time as the minority class. Therefore, the recall value for the minority class (class 1) is the primary indicator in evaluating the model's ability to identify students at risk of not graduating on time.

In the baseline scenario without resampling, the Random Forest model performed best with an accuracy of 0.873, an AUC of 0.882, and an F1-score

of 0.640. However, the relatively low recall value (0.573) indicates that the model still has limitations in detecting students who do not graduate on time as a minority class. A similar situation also occurs with other algorithms, where high accuracy values are more influenced by the dominance of the majority class, resulting in a bias in favor of students who graduate on time.

After applying the SMOTE technique, the model's ability to identify students who do not graduate on time improved. This is indicated by an increase in recall values for almost all algorithms, particularly for Logistic Regression (0.752) and Gaussian Naïve Bayes (0.789). However, this increased sensitivity to the minority class was accompanied by a decrease in accuracy, indicating a tradeoff between the model's global performance and its ability to detect graduation risks.

In the ADASYN scenario, a significant increase in recall for the minority class was also observed, particularly for the Gaussian Naïve Bayes algorithm (0.833) and Logistic Regression (0.789). However, this method tended to produce a greater decrease in accuracy than SMOTE, indicating that the synthetic data distribution generated by ADASYN can increase model sensitivity but potentially decrease predictive stability.

The Borderline-SMOTE technique demonstrated a more balanced balance between

accuracy and recall. The Random Forest and XGBoost algorithms in this scenario maintained accuracy values above 0.84, with recalls of 0.691 and 0.641, respectively, and a relatively stable F1-score. This indicates that Borderline-SMOTE is effective in improving the model's ability to identify students who do not graduate on time, particularly at the boundary between classes.

Meanwhile, with the Random Undersampling technique, the recall value for the minority class increased significantly across all algorithms, with Logistic Regression achieving the highest value of 0.802. However, reducing the amount of data for the majority class resulted in a decrease in overall accuracy, indicating the loss of some important information during the undersampling process.

Overall, the evaluation results showed that Random Forest with SMOTE-based oversampling and Borderline-SMOTE provided the most stable performance in predicting student graduation. This combination is able to maintain a balance between accuracy, F1-score, and AUC, while improving the model's ability to identify students who are likely to fail to graduate on time as a minority class.

To clarify the differences in model performance across various resampling scenarios, Figure 5 presents a comparison of the performance of classification algorithms based on the metrics outlined in the methods section.

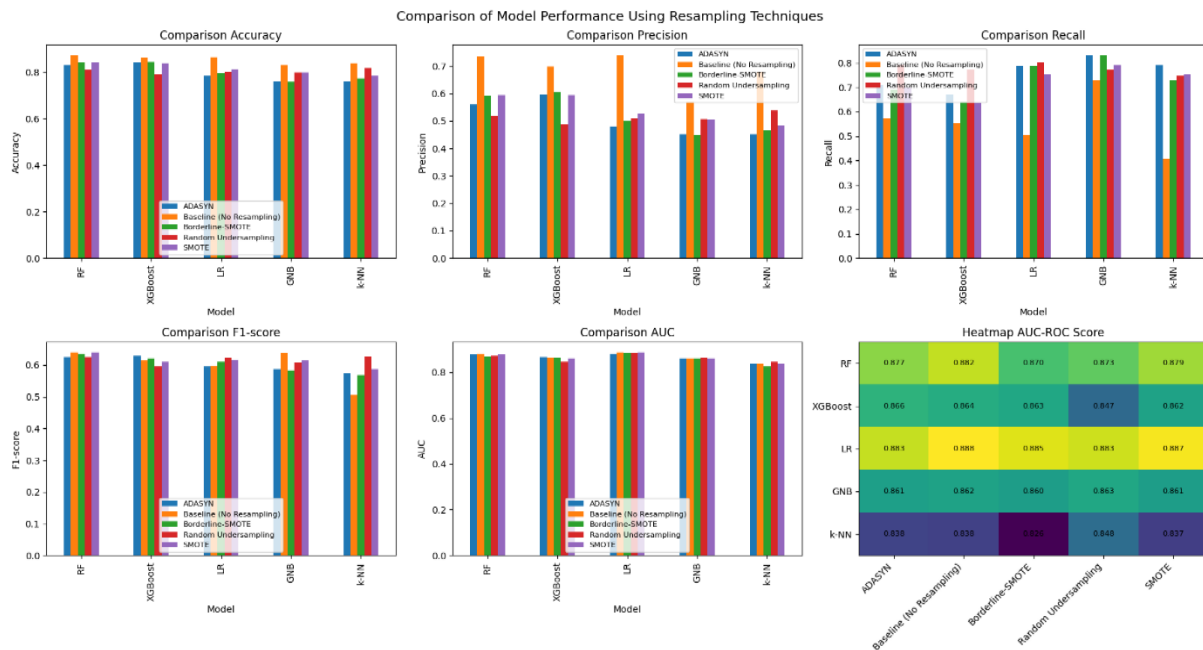


Fig. 5 Model Performance Comparison

Based on the evaluation results, it can be seen that the performance of classification models in predicting student graduation is significantly influenced by the type of algorithm used and the approach used to handle class imbalance. As shown in Table 1 and Figure 5, models trained on original data without resampling techniques (baseline) generally

produce relatively high accuracy scores. However, this performance is largely dominated by the model's success in predicting the majority class, namely students who graduate on time.

Under baseline conditions, the Random Forest, XGBoost, and Logistic Regression algorithms demonstrated high accuracy and recall scores.

However, because the positive class in this study represents students who graduate on time, these high recall scores reflect the model's tendency to overrecognize the majority class. This indicates a model bias toward the dominant class, while its ability to identify students who do not graduate on time as the minority class is still limited.

After the implementation of resampling techniques, the model's performance pattern changed. In some algorithms, recall scores decreased compared to the baseline. This decrease does not indicate a decline in model quality, but rather indicates a reduction in the dominance of the majority class in the classification process. With a more balanced data distribution, the model no longer focuses too much on the majority class and begins to pay more proportional attention to the minority class.

Improvements in model performance after resampling are more accurately analyzed using the F1-score and AUC-ROC metrics. These two metrics are able to represent the balance between precision and recall, as well as the model's overall discriminatory ability under imbalanced data conditions. Test results show that Random Forest and XGBoost have relatively stable AUC values across various resampling scenarios, indicating good generalization ability despite changes in the data distribution.

In terms of data balancing techniques, oversampling methods such as SMOTE, ADASYN, and Borderline-SMOTE tend to produce better performance balance than Random Undersampling. Undersampling techniques can reduce the dominance of the majority class, but they risk losing important information. Conversely, synthetic oversampling techniques enrich the representation of the minority class without removing the original data, allowing the model to learn classification patterns more optimally.

Furthermore, algorithm characteristics also play a crucial role in determining the effectiveness of resampling techniques. Ensemble-based algorithms such as Random Forest and XGBoost have proven more stable than Logistic Regression, k-Nearest Neighbors, and Naïve Bayes. The findings are consistent with prior studies on the same student graduation dataset, where reported accuracies ranged from 84% to 88% depending on the algorithm and optimization method. Optimization-based approaches such as SVM-PSO and Decision Tree-PSO slightly improved performance compared to standard models.

In this study, the highest baseline accuracy reached 87.3% using Random Forest, comparable to several optimized models reported previously. Beyond accuracy, this study also evaluated precision, recall, F1-score, and AUC-ROC to ensure a more comprehensive assessment under imbalanced conditions. Logistic Regression achieved the highest AUC (0.888), while Random Forest showed stable AUC values (0.870–0.882) across resampling scenarios.

Unlike earlier works that focused primarily on accuracy improvement through metaheuristic optimization, this study emphasizes balanced evaluation metrics, particularly for minority-class detection. The Friedman test further confirmed significant differences among classifiers ($p < 0.05$).

Table 2. Compare Models within each

| Resampling | p-value |
|----------------------|---------|
| Baseline | 0.0022 |
| SMOTE | 0.0046 |
| ADASYN | 0.0081 |
| Borderline-SMOTE | 0.0065 |
| Random Undersampling | 0.0186 |

The Friedman test results indicate statistically significant differences among classification models across all resampling scenarios. As shown in Table 2, all p-values are below the 0.05 significance threshold (Baseline: 0.0022; SMOTE: 0.0046; ADASYN: 0.0081; Borderline-SMOTE: 0.0065; Random Undersampling: 0.0186).

Table 3. Compare Resampling within each

| Model | Resampling | p-value |
|-------|------------|---------|
| Model | RF | 0.326 |
| | XGB | 0.417 |
| | LR | 0.364 |
| | GNB | 0.335 |
| | kNN | 0.475 |

In contrast, when comparing resampling techniques within each individual model, the Friedman test does not reveal statistically significant differences. All p-values exceed 0.05 (RF: 0.326; XGB: 0.417; LR: 0.364; GNB: 0.335; kNN: 0.475).

The Friedman test demonstrates that significant differences exist among classifiers within each resampling setting ($p < 0.05$ for all scenarios). However, no statistically significant differences are observed when comparing resampling techniques within the same classifier ($p > 0.05$ for all models).

These findings imply that classifier architecture plays a more dominant role than resampling strategy in determining performance under imbalanced data conditions.

Overall, the results of this study confirm that model evaluation in imbalanced data cannot rely solely on accuracy or recall for the majority class. A more comprehensive evaluation using F1-score and AUC-ROC is needed to more objectively describe the model's performance, particularly in identifying students at risk of failing to graduate on time. Therefore, selecting a combination of classification algorithms and techniques for handling imbalanced data is a key factor in building a reliable and balanced student graduation prediction system.

CONCLUSION

This study investigated the performance of multiple classification algorithms for predicting student graduation under imbalanced data conditions. The findings highlight that while baseline models may achieve high overall accuracy, comprehensive evaluation using F1-score and AUC-ROC is essential to properly assess minority-class detection. The comparative analysis demonstrates that classifier selection has a greater influence on performance variation than the choice of resampling technique, as supported by statistical validation using the Friedman test. Ensemble-based approaches consistently exhibited robust and stable performance across evaluation scenarios. The primary contribution of this work lies in providing a systematic and statistically validated comparison of classifiers and imbalance-handling strategies within the same dataset. Nonetheless, the study is limited to a single dataset and does not incorporate cost-sensitive learning or advanced interpretability analysis, which remain important directions for future research.

REFERENCES

- Altalhan, M., Algarni, A., & Monia, T. H. A. (2025). Imbalanced Data Problem in Machine Learning: A Review. *Turki Hadj Alouane Monia*, 11. <https://doi.org/10.1109/ACCESS.2025.3531662>
- Ayodele, A. (2023). A comparative study of ensemble learning techniques for imbalanced classification problems. *World Journal of Advanced Research and Review*, 19(1), 1633–1643. <https://doi.org/10.30574/wjarr.2023.19.1.1202>
- Carvalho, Miguel., Pinho, Armando J., & Brás, Susana. (2025). Resampling approaches to handle class imbalance: a review from a data perspective. *Journal of Big Data*, 12(71). <https://doi.org/10.1186/s40537-025-01119-4>
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 785–794). Association for Computing Machinery. <https://doi.org/10.1145/2939672.2939785>
- Çorbacioğlu, Ş. K., & Aksel, G. (2023). Receiver operating characteristic curve analysis in diagnostic accuracy studies. *Turkish Journal of Emergency Medicine*, 23(4). https://doi.org/10.4103/tjem.tjem_182_23
- Fernández, Alberto., García, Salvador., Galar, Mikel., Prati, Ronaldo C., Krawczyk, Bartosz., Herrera, Francisco. (2018). Learning from Imbalanced Data Sets. Springer Nature Link.
- Fulazzaky, T., Saefuddin, A., & Soleh, A. M. (2024). Evaluating Ensemble Learning Techniques for Class Imbalance in Machine Learning: A Comparative Analysis of Balanced Random Forest, SMOTE-RF, SMOTEBoost, and RUSBoost. *Scientific Journal of Informatics*, 11(4). <https://doi.org/https://doi.org/10.15294/sji.v11i4.15937>
- Geron, A. (2023). *Hands on Machine learning with Scikit Learn Keras and Tensor Flow Concepts, Tools, and Techniques to Build Intelligent Systems* (3rd ed.). O'Reilly Media, Inc.
- Glučina, M., Ariana Lorencin, Nikola Anđelić, & Ivan Lorencin. (2023). Cervical Cancer Diagnostics Using Machine Learning Algorithms and Class Balancing Techniques. *Applied Sciences*, 13(12). <https://doi.org/10.3390/app13021061>
- Gurcan, F., & Soylu, A. (2024). Learning from Imbalanced Data: Integration of Advanced Resampling Techniques and Machine Learning Models for Enhanced Cancer Diagnosis and Prognosis. *Cancers*, 16(19). <https://doi.org/10.3390/cancers16193417>
- Huang, C. Y., & Dai, H. L. (2021). Learning from class-imbalanced data: review of data driven methods and algorithm driven methods. *Data Science in Finance and Economics*, 1(1), 21–36. <https://doi.org/10.3934/DSFE.2021002>
- Iqbal, M., Yumi Novita Dewi, Lisnawanty, Maisyaroh, & Suhardjono. (2025). Optimalisasi Prediksi Dalam Kelulusan Berbasis Deep Learning: Perbandingan Kinerja Multi- Layer Perceptron dan Deep Neural Network. *Infotek : Jurnal Informatika Dan Teknolog*, 8(2), 630–641. <https://doi.org/DOI : 10.29408/jit.v8i2.30756>
- Kahfi, A. H., Prihatin, T., Yudhistira, Sudradjat, A., & Wijaya, G. (2024). The Right Steps Towards Graduation: Nb-Pso Smart Combination For Student Graduation Prediction. *Jurnal Teknik Informatika (Jutif)*, 5(2), 607–614. <https://doi.org/10.52436/1.jutif.2024.5.2.1889>
- Miftahushudur, Tajul., Sahin, Halil Mertkan., Grieve, Bruce., Yin, Hujun. (2025). A Survey of Methods for Addressing Imbalance Data Problems in Agriculture Applications. *Remote Sensing*, 17(3). <https://doi.org/10.3390/rs17030454>
- Mulugeta, G., Zewotir, T., Tegegne, A. S., Juhar, L. H., & Muleta, M. B. (2023). Classification of imbalanced data using machine learning algorithms to predict the risk of renal graft failures in Ethiopia. *BMC Medical Informatics and Decision Making*, 23(98). <https://doi.org/10.1186/s12911-023-02185-5>
- Muraru, M. M., Simó, Z., & Iantovics, L. B. (2024). Cervical Cancer Prediction Based on Imbalanced Data Using Machine Learning Algorithms with a Variety of Sampling Methods. *Applied Sciences*, 14(22). <https://doi.org/10.3390/app142210085>
- Nouas, Sihem., Oukid, Lamia., Boumahdi, Fatima., (2025). Enhancing imbalanced text classification: an overlap-based refinement approach. *Data Science and Management*, 8(4). <https://doi.org/10.1016/j.dsm.2025.03.001>
- Nurdin, H., Carolina, I., Andharsaputri, R. L.,

- Wuryanto, A., & Ridwansyah. (2024). Forward Selection as a Feature Selection Method in the SVM Kernel for Student Graduation Data. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(October), 2531–2537. <https://doi.org/10.33395/sinkron.v8i4.14172>
- Nurdin, H., Sartini, Sumarna, Maulana, Y. I., & Riyanto, V. (2023). Prediction of Student Graduation with the Neural Network Method Based on Particle Swarm Optimization. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 8(4), 2353–2362. <https://doi.org/10.33395/sinkron.v8i4.12973>
- Pangesti, W. E., Ariyati, I., Priyono, Sugiono, & Suryadithia, R. (2024). Utilizing Genetic Algorithms To Enhance Student Graduation Prediction With Neural Networks. *Sinkron : Jurnal Dan Penelitian Teknik Informatika*, 9(1), 276–284. <https://doi.org/10.33395/sinkron.v9i1.13161> e-ISSN
- Ridwansyah, Iqbal, M., Destiana, H., Sugiono, & Hamid, A. (2024). Data Mining Berbasis Machine Learning Untuk Analitik Prediktif Dalam Kelulusan. *SemanTIK*, 10(2), 1–10. <https://doi.org/10.55679/semantik.v10i2.67>
- Rukiastiandari, S., Rohimah, L., Aprillia, A., & Mutia, F. (2024). Predicting Graduation Outcomes: Decision Tree Model Enhanced with Genetic Algorithm. *Paradigma - Jurnal Komputer Dan Informatika*, 26(1), 1–6. <https://doi.org/10.31294/p.v26i1.3165>
- Salifu, Daisy., Chepkemoi, Lorna., Ibrahim, Eric Ali., Nkoba, Kiatoko., Tonnang, Henri E.Z. (2025). Data Augmentation and Machine Learning algorithms for multi-class imbalanced morphometrics data of stingless bees. *Heliyon*, 11(3). <https://doi.org/10.1016/j.heliyon.2025.e42214>
- Salmi, M., Atif, D., Oliva, D., Abraham, A., & Sebastian Ventura. (2024). Handling imbalanced medical datasets: review of a decade of research. *Artificial Intelligence Review (Springer Nature)*, 57(10). <https://doi.org/10.1007/s10462-024-10884-2>
- Siregar, A. Y., & Arifin, A. S. (2024). Enhancing XGBoost Classification with SVM-SMOTE & EasyEnsemble for Imbalanced Telemedicine Sentiment Data. *Jurnal Indonesia Sosial Teknologi*, 5(10). <https://doi.org/10.59141/jist.v5i10.1160>
- Sumarna, Nawawi, I., Suhardjono, Hari Sugiarto, & Yuliandari, D. (2024). Meningkatkan Akurasi Prediksi Kelulusan Mahasiswa Menggunakan Metode Algoritma Genetika. *Jurnal Informatika, Manajemen Dan Komputer*, 16(2). <https://doi.org/10.36723/juri.v16i2.706>
- Wibowo, Prasetyo. & Fatichah, Chastine. (2021). An in-depth performance analysis of the oversampling techniques for high-class imbalanced dataset. *Institut Teknologi Sepuluh Nopember*, 7(1). <https://doi.org/10.26594/register.v7i1.2206>
- Yang, Y., Khorshidi, H. A., & Aickelin, U. (2024). A review on over-sampling techniques in classification of multi-class imbalanced datasets: insights for medical problems. *Frontiers in Digital Health*, 6(1430245). <https://doi.org/10.3389/fdgth.2024.1430245>