JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

Early Warning System for Student Academic Risk Prediction Using Gradient Boosting with SMOTE-Based Class Balancing

Roberto Kaban[1)*], Jamaludin Bin Sallim[2)], Susmanto[3)], Rozlina Binti Mohamed[4)]

[1)] Teknik Informatika, Institut Teknologi dan Bisnis Indonesia, Deli Serdang, Indonesia
[2, 4)] Software Engineering & Artificial Intelligence, Universiti Malaysia Pahang Al-Sultan Abdulllah, Pahang, Malaysia
[3)]Teknik Komputer, Universitas Serambi Mekkah, Banda Aceh, Indonesia
[1)]roberto.kaban@yahoo.com, [2)]jamal@umpsa.edu.my, [3)]susmanto@serambimekkah.ac.id, [4)]rozlina@umpsa.edu.my

Submitted : 29 June 2026 | **Accepted** : 20 July 2026 | **Published** : 21 July 2026

Abstract: This study proposes a machine learning-based Early Warning Academic System for predicting student academic risk levels using historical academic performance data from ITB Indonesia. The aim of this research is to identify at-risk students at an early stage using routinely collected academic indicators, including Quiz Score, Attendance, Assignment Score, and Midterm Score. The proposed framework integrates Gradient Boosting as the classification model, Synthetic Minority Oversampling Technique (SMOTE) to address class imbalance, and Stratified 5-Fold Cross Validation to ensure robust performance evaluation. The model classifies students into three categories, namely Safe, Warning, and Risk. Experimental results show that the proposed model achieves an accuracy of 80.00%, macro F1-score of 78.24%, Cohen's Kappa of 0.6306, and Matthews Correlation Coefficient of 0.6311, indicating strong and reliable predictive performance under imbalanced data conditions. ROC-AUC analysis further confirms strong discriminative capability, particularly for identifying Risk students. Feature importance analysis reveals that Midterm Score is the most influential predictor (43.49%), followed by Quiz Score (25.60%) and Assignment Score (21.99%), while Attendance contributes 8.93%. These findings indicate that academic performance indicators, especially midterm assessments, play a critical role in early risk detection. The results demonstrate that effective early warning prediction can be achieved using a limited set of routinely available academic variables, providing a practical and scalable approach to support early intervention strategies in higher education institutions.
Keywords: Gradient Boosting; Early Warning System; Student Academic Risk; SMOTE; Educational Data Mining

INTRODUCTION

The rapid advancement of information technology has transformed various sectors, including higher education (Liu, 2025). Universities are increasingly expected to improve educational quality, student retention, and graduation rates while ensuring that students receive adequate academic support throughout their studies (Lünich et al., 2024). One of the major challenges faced by higher education institutions is the early identification of students who are at risk of experiencing academic difficulties (Albreiki et al., 2023; Osborne & Lang, 2023). Failure to detect such students in a timely manner may lead to declining academic performance, delayed graduation, increased dropout rates, and inefficient allocation of academic support resources (Albreiki et al., 2023).

Traditionally, student performance evaluation relies on cumulative grade point averages, final examination scores, or end-of-semester assessments. While these indicators provide valuable information regarding academic achievement, they often reveal problems only after students have already experienced significant learning difficulties. Consequently, academic advisors and institutions may lose the opportunity to implement preventive

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

interventions that could improve student outcomes (Akello et al., 2026). Therefore, there is a growing need for predictive systems capable of identifying students who require academic assistance at an earlier stage of the learning process.

In recent years, Educational Data Mining (EDM) and Learning Analytics have emerged as important research areas for understanding and predicting student behavior and academic performance (Ozyurt et al., 2023). By leveraging historical academic records, machine learning techniques can discover hidden patterns within educational data and generate predictive models that support evidence-based decision making (Almalawi et al., 2024; Ganage, 2026). One of the most promising applications of these techniques is the development of Early Warning Systems (EWS), which aim to classify students according to their academic risk levels and enable institutions to provide targeted interventions before academic problems become severe (Paul & Devi, 2025; Sharma & Sasikala, 2026).

Various machine learning algorithms have been employed for student performance prediction, including Decision Trees, Random Forests, Support Vector Machines, Artificial Neural Networks, and ensemble learning approaches (Ganage, 2026; Sathvik & Nagaveni Biradar, 2025; Tiwari & Jain, 2024). Among these methods, Gradient Boosting has gained significant attention due to its ability to combine multiple weak learners into a highly accurate predictive model (Li & Sheng, 2025). Unlike traditional classification techniques, Gradient Boosting can effectively model complex and non-linear relationships among academic variables, making it particularly suitable for educational datasets that often exhibit heterogeneous learning patterns and varying student characteristics (Aluso & Enyejo, 2025).

Despite the promising performance of machine learning models, one common challenge in educational datasets is class imbalance (Sujon et al., 2024). In most academic environments, the majority of students tend to achieve satisfactory performance, while only a relatively small proportion experience academic difficulties. Such imbalance can cause classification models to become biased toward majority classes and reduce their effectiveness in identifying students who require intervention (Rochmadi, 2025). Consequently, the Synthetic Minority Oversampling Technique (SMOTE) has been widely adopted to generate synthetic instances for minority classes and improve class representation during model training (Elreedy et al., 2024). The integration of SMOTE with ensemble learning algorithms has demonstrated considerable potential in improving predictive performance and minority class recognition.

Although previous studies have explored student performance prediction using various machine learning techniques, several limitations remain. Many existing studies focus on binary classification, which categorizes students only as successful or unsuccessful (Pan et al., 2025). Such approaches may not adequately represent the gradual progression of academic risk encountered in real educational environments. Furthermore, some studies rely solely on accuracy as the primary evaluation metric, despite the fact that accuracy alone may provide misleading conclusions when class distributions are imbalanced (Ge et al., 2025; Villar & Andrade, 2024). In addition, limited attention has been given to combining imbalance handling techniques, robust ensemble learning algorithms, multi-class classification strategies, and model interpretability within a unified early warning framework.

To address these limitations, this study proposes a machine learning-based Early Warning Academic System that integrates Gradient Boosting and SMOTE for multi-class student risk classification. The proposed framework classifies students into three academic categories, namely Safe, Warning, and Risk, based on their academic achievement. Academic indicators including quiz scores, attendance records, assignment scores, and midterm examination scores are utilized as predictive features. The use of three risk levels allows institutions to implement more differentiated intervention strategies according to the severity of students' academic conditions.

This research employs historical academic data obtained from students of Institut Teknologi dan Bisnis Indonesia (ITB Indonesia). The dataset contains academic performance indicators that are routinely collected during the learning process and therefore can be utilized for early prediction before final examination results become available. To ensure robust model evaluation, the proposed framework incorporates Stratified K-Fold Cross Validation, which preserves class distribution across validation folds and provides a more reliable estimation of model generalization performance. Model effectiveness is assessed using multiple evaluation metrics, including accuracy, precision, recall, F1-score, confusion matrix analysis, and Receiver Operating Characteristic–Area Under the Curve (ROC-AUC) (Aslam et al., 2025).

In addition to predictive performance evaluation, this study also investigates the contribution of individual academic variables through feature importance analysis. Such analysis provides valuable insights into which academic factors most strongly influence student risk classification. These findings can assist lecturers, academic advisors, and institutional management in designing more effective intervention strategies and support programs based on empirical evidence.

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

The contributions of this study are threefold. First, it develops a multi-class early warning framework capable of classifying students into Safe, Warning, and Risk categories using academic performance indicators. Second, it addresses the class imbalance problem through the implementation of SMOTE to improve minority class detection. Third, it combines predictive analytics and explainable machine learning through feature importance analysis, thereby supporting transparent and data-driven academic decision making. The proposed framework is expected to contribute to the advancement of educational data mining research and provide practical benefits for improving student success and academic quality assurance in higher education institutions.

LITERATURE REVIEW

Machine Learning for Classification

Machine Learning (ML) is a branch of Artificial Intelligence (AI) that enables computer systems to learn patterns from historical data and make predictions without being explicitly programmed for specific tasks (Yu et al., 2023). Based on the learning paradigm, machine learning algorithms are generally categorized into supervised, unsupervised, semi-supervised, and reinforcement learning (Kaban et al., 2026). Among these paradigms, supervised learning is the most widely adopted for predictive modeling because it utilizes labeled datasets to construct models capable of predicting unseen observations (Alomari et al., 2025).

Classification is a supervised learning task that assigns observations into predefined categories according to their feature characteristics (Chukwura, 2023). Numerous classification algorithms have been developed, including Logistic Regression, Decision Trees, Support Vector Machines, Artificial Neural Networks, and ensemble learning approaches. The selection of an appropriate classifier depends on various factors such as dataset characteristics, predictive performance, computational efficiency, and model interpretability (Ling, 2023).

Ensemble Learning and Gradient Boosting

Ensemble learning improves predictive performance by combining multiple base learners into a single model (Sharanya, 2024). The two principal ensemble strategies are bagging and boosting, where bagging constructs independent learners in parallel while boosting sequentially trains learners to correct errors generated by previous models (Bagawade & Thirupurasundari, 2025). This iterative learning mechanism enables boosting algorithms to capture complex nonlinear relationships more effectively than many conventional classifiers.

Among boosting-based methods, Gradient Boosting has received considerable attention because of its high predictive accuracy, robustness, and ability to model heterogeneous data distributions (Nguyen et al., 2024). Previous comparative studies have reported that Gradient Boosting consistently achieves competitive performance across various structured classification problems, including educational data analysis (Ersozlu et al., 2024). Furthermore, tree-based Gradient Boosting models provide feature importance information, allowing researchers to identify influential predictor variables and improve model interpretability (Rakitin et al., 2025).

Synthetic Minority Oversampling Technique (SMOTE)

Class imbalance is a common challenge in classification problems where one class contains substantially fewer observations than other classes. Under such conditions, classification algorithms tend to favor majority classes, resulting in poor recognition of minority instances despite achieving high overall accuracy (Lokanan, 2024). Consequently, evaluation metrics such as precision, recall, F1-score, and ROC-AUC are generally considered more informative than accuracy alone for assessing imbalanced classification models.

Synthetic Minority Oversampling Technique (SMOTE) is one of the most widely used approaches for handling class imbalance (Călin, 2025; Mujahid et al., 2024). Instead of duplicating minority observations, SMOTE generates synthetic samples by interpolating neighboring minority instances, thereby improving class representation while reducing overfitting. Numerous studies have demonstrated that integrating SMOTE with ensemble learning algorithms significantly improves minority class detection and overall predictive performance (Idhmad et al., 2024; Rahmi et al., 2024; Yuningsih et al., 2023).

Related Works

Student performance prediction has become an important topic in Educational Data Mining because it supports early identification of academically at-risk students (Lu, 2024). Previous studies have applied conventional machine learning algorithms such as Decision Trees, Logistic Regression, and Support Vector Machines (Angeioplastis et al., 2025; Charaya & Sonia, 2026; Rajasekhar & Fernandes, 2026; Rico-Juan et al., 2024). More recently, ensemble learning methods, including Random Forest, XGBoost, AdaBoost, and Gradient Boosting, have demonstrated superior predictive performance (Nafea et al., 2023; Rawat & Pant, 2025; Villar & Andrade, 2024).

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

Academic indicators such as attendance, quizzes, assignments, and examination scores are widely recognized as important predictors of student achievement (Teresa & et. al, 2025). To address class imbalance, several studies have employed SMOTE to improve the identification of at-risk students (Giap et al., 2024; Kabumba et al., 2025).

However, most existing studies focus on binary classification and primarily emphasize predictive accuracy, while limited attention has been given to multi-class prediction, model interpretability, and validation using Indonesian higher education datasets.

Research Gap

Although previous studies have achieved promising results, several research gaps remain. First, most studies formulate student performance prediction as a binary classification problem, which may not adequately represent varying levels of academic risk. Second, the integration of Gradient Boosting and SMOTE for multi-class prediction within an Early Warning Academic System has received limited attention. Third, existing studies mainly focus on predictive performance while providing limited model interpretability for academic decision-making (Giap et al., 2024; Kabumba et al., 2025; Nafea et al., 2023).

Therefore, this study proposes a machine learning-based Early Warning Academic System that integrates Gradient Boosting and SMOTE to classify students into Safe, Warning, and Risk categories while providing interpretable insights through feature importance analysis.

METHOD

Research Framework

This study proposes a machine learning-based Early Warning Academic System for predicting students' academic risk levels using historical academic performance data. The proposed framework consists of six major stages: data collection, data preprocessing, class balancing using Synthetic Minority Oversampling Technique (SMOTE), model development using Gradient Boosting, model validation through Stratified K-Fold Cross Validation, and performance evaluation. The overall workflow of the proposed methodology is illustrated in Figure 1.

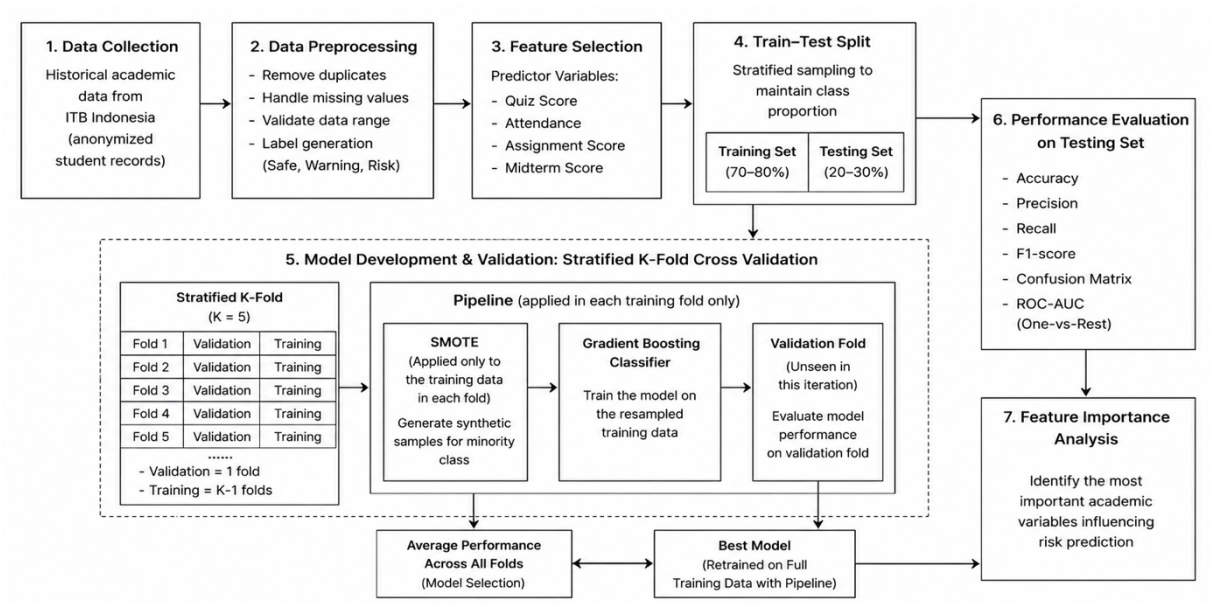

Fig.1 Research Framework

Figure 1 illustrates the proposed research framework. The workflow begins with collecting historical academic data from ITB Indonesia. The collected data are then preprocessed to remove inconsistencies and prepare the dataset for analysis. Since the original dataset exhibits an imbalanced class distribution, SMOTE is employed to generate synthetic minority samples. The balanced dataset is subsequently used to train a Gradient Boosting classifier. Model performance is validated using Stratified K-Fold Cross Validation and evaluated through multiple classification metrics. Finally, feature importance analysis is performed to identify the academic variables that contribute most significantly to student risk prediction.

Dataset Description

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

The dataset used in this research was obtained from the academic information system of ITB Indonesia. The collected data consist of historical academic records from undergraduate students enrolled in several study programs. To maintain student privacy, all personally identifiable information was removed before analysis.

Each observation represents one student's academic performance during a semester. Four academic indicators were selected as predictor variables because they are routinely collected throughout the learning process and become available before the final examination. These variables are summarized in Table 1.

Table 1
Predictor Variables Used in the Study

Variable	Description
Quiz Score	Average score obtained from quizzes
Attendance	Percentage of class attendance
Assignment Score	Average assignment score
Midterm Examination Score	Midterm examination (UTS) score

The target variable is the student's academic risk category. The class labels used for prediction are presented in Table 2.

Table 2
Academic Risk Categories

Label	Description
Safe	Students showing satisfactory academic performance
Warning	Students requiring academic monitoring
Risk	Students requiring immediate academic intervention

Data Preprocessing

Before model construction, several preprocessing steps were performed to improve data quality. First, incomplete and duplicate records were identified and removed to ensure dataset consistency. Missing values were examined and handled according to the characteristics of each variable. Numerical variables were verified to ensure that all values were within valid academic score ranges. Second, the academic performance labels were generated based on predefined institutional criteria, producing three risk categories: Safe, Warning, and Risk. Finally, exploratory analysis was conducted to examine the class distribution. The analysis revealed that the dataset was imbalanced because the majority of students belonged to the Safe category, while considerably fewer students were classified as Warning and Risk.

Handling Class Imbalance Using SMOTE

Since the dataset exhibited class imbalance, the Synthetic Minority Oversampling Technique (SMOTE) was applied before model training. SMOTE generates synthetic samples by interpolating existing minority observations rather than simply duplicating them. This approach increases minority class representation while reducing the likelihood of overfitting. The synthetic sample is generated as (Călin, 2025; Kabumba et al., 2025):

$$x_{new} = x_i + \delta(x_{nn} - x_i), \quad 0 \leq \delta \leq 1 \tag{1}$$

where

- x_i is the minority instance,
- x_{nn} denotes one of its nearest minority neighbors,
- δ is a random value between 0 and 1.

The balanced dataset obtained after SMOTE was subsequently used during model training.

3.5 Gradient Boosting Classification

The balanced dataset was used to train a Gradient Boosting classifier. Gradient Boosting constructs a strong predictive model by sequentially combining multiple weak learners (Li & Sheng, 2025). During each iteration, a new decision tree is trained to reduce the prediction errors produced by the previous ensemble.

The model can be expressed as :

$$F_m(x) = F_{(m-1)}(x) + \eta\, h_m(x) \tag{2}$$

where

- $F_m(x)$ represents the ensemble model after iteration mmm,
- $h_m(x)$ denotes the newly constructed weak learner,

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

- η is the learning rate controlling the contribution of each learner.

The iterative learning process enables Gradient Boosting to model complex nonlinear relationships among academic variables and improve prediction accuracy.

Model Validation

To obtain reliable performance estimation, Stratified K-Fold Cross Validation was employed. Unlike conventional K-Fold Cross Validation, Stratified K-Fold preserves the proportion of each class in every fold, thereby reducing evaluation bias caused by imbalanced datasets. The dataset was divided into *k* mutually exclusive subsets. During each iteration, one subset served as testing data while the remaining subsets were used for training. The average performance across all folds was calculated to estimate model generalization capability.

Performance Evaluation

The performance of the proposed model is evaluated using multiple complementary metrics to ensure a comprehensive and reliable assessment, particularly under class imbalance conditions. These metrics include Accuracy, Precision, Recall, F1-score, Cohen's Kappa, Matthews Correlation Coefficient (MCC), and Receiver Operating Characteristic (ROC) analysis.

Accuracy is defined as the proportion of correctly classified instances over the total number of instances (Albreiki et al., 2023; Aslam et al., 2025):

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Although accuracy provides a general overview of model performance, it may not be sufficient for imbalanced datasets. Therefore, additional evaluation metrics are required. Precision measures the proportion of correctly predicted positive observations among all predicted positive observations.

$$Precision = \frac{TP}{TP+FP} \tag{4}$$

Recall evaluates the ability of the model to correctly identify all relevant instances of a given class.

$$Recall = \frac{TP}{TP+FN} \tag{5}$$

The F1-score represents the harmonic mean of Precision and Recall, providing a balanced measure between the two metrics.

$$F_1 = \frac{2\times Precision\times Recall}{Precision+Recall} \tag{6}$$

To provide a more robust evaluation under class imbalance, Cohen's Kappa is employed. This metric measures the level of agreement between predicted and actual labels while accounting for agreement that may occur by chance.

$$\kappa = \frac{p_o-p_e}{1-p_e} \tag{7}$$

where,

$$p_o = \frac{1}{N}\sum_{k=1}^{K} C_{kk} \tag{8}$$

$$p_e = \sum_{k=1}^{K} \frac{(C_{k\cdot}\cdot C_{\cdot k})}{N^2} \tag{9}$$

- C_{kk} denotes the number of correctly classified instances for class *k* (the diagonal element of the confusion matrix).
- $C_{k\cdot}$ denotes the total number of actual instances in class *k* (row sum).
- $C_{\cdot k}$ denotes the total number of instances predicted as class *k* (column sum).
- K is the total number of classes.
- N is the total number of instances.

Cohen's Kappa is widely used in multi-class classification problems due to its ability to provide a chance-corrected measure of performance. In addition, the Matthews Correlation Coefficient (MCC) is utilized as a comprehensive performance measure that considers all elements of the confusion matrix. MCC is regarded as a balanced metric even when the class distribution is highly skewed. For binary classification, MCC is defined as:

$$MCC = \frac{TP\cdot TN-FP\cdot FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \tag{11}$$

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

For multi-class classification, MCC is computed as a generalized correlation coefficient based on the confusion matrix. The MCC value ranges from -1 to +1, where +1 indicates perfect prediction, 0 indicates performance equivalent to random guessing, and -1 indicates total disagreement between predictions and actual labels.

Furthermore, Receiver Operating Characteristic (ROC) curves are used to evaluate the discriminative capability of the classifier. The Area Under the Curve (AUC) is computed to quantify the overall separability of the model. For multi-class classification, ROC analysis is performed using a One-vs-Rest (OvR) strategy, where each class is evaluated against all other classes.

$$AUC = \int_0^1 \mathrm{TPR}(\mathrm{FPR})\ d(\mathrm{FPR}) \tag{12}$$

Collectively, these evaluation metrics provide a comprehensive assessment of model performance in terms of accuracy, robustness, class imbalance sensitivity, and discriminative ability.

Feature Importance Analysis

Besides predictive performance, this study also investigates the contribution of individual academic variables through feature importance analysis. Feature importance scores were extracted directly from the trained Gradient Boosting model. Variables with higher importance values contribute more significantly to the prediction process and therefore represent stronger indicators of students' academic risk. The resulting feature importance ranking provides interpretable insights for lecturers and academic advisors regarding which academic indicators should receive greater attention when monitoring student progress.

RESULT

Descriptive Statistics and Class Distribution

The dataset used in this study comprises academic records from 271 undergraduate students at ITB Indonesia. Table 2 presents the descriptive statistics of the four predictor variables and the final examination score (UAS).

Table 3
Descriptive Statistics of Academic Variables (N = 271)

Variable	N	Mean	Std	Min	Q1	Median	Q3	Max
Quiz Score	271	69.02	10.78	30.0	65.0	70.0	75.0	91.0
Attendance (%)	271	78.83	19.42	25.0	69.0	86.0	94.0	100.0
Assignment Score	271	70.53	8.45	51.0	62.0	75.0	75.0	85.0
Midterm Score	271	71.68	6.32	56.0	70.0	70.0	75.0	85.0
UAS Score	271	73.39	9.50	51.0	65.5	75.0	80.0	95.0

As presented in Table 3, Quiz Score has a mean of 69.02 (SD = 10.78), which is the lowest mean among the score-based variables and also exhibits the widest spread, with values ranging from 30 to 91. Attendance shows high variability (Mean = 78.83%, SD = 19.42), indicating substantial differences in student engagement across the dataset. Assignment Score averages 70.53 (SD = 8.45), while Midterm Score records a mean of 71.68 (SD = 6.32), with a narrower distribution compared to Quiz Score. The UAS score has the highest mean at 73.39 (SD = 9.50), reflecting generally better performance in the final examination.

The distribution of the target variable (academic risk category) is summarized in Table 3 and visualized in Figure 2.

Table 4
Class Distribution of Student Academic Risk

Class	Count	Percentage (%)
Safe	159	58.67
Warning	87	32.10
Risk	25	9.23

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

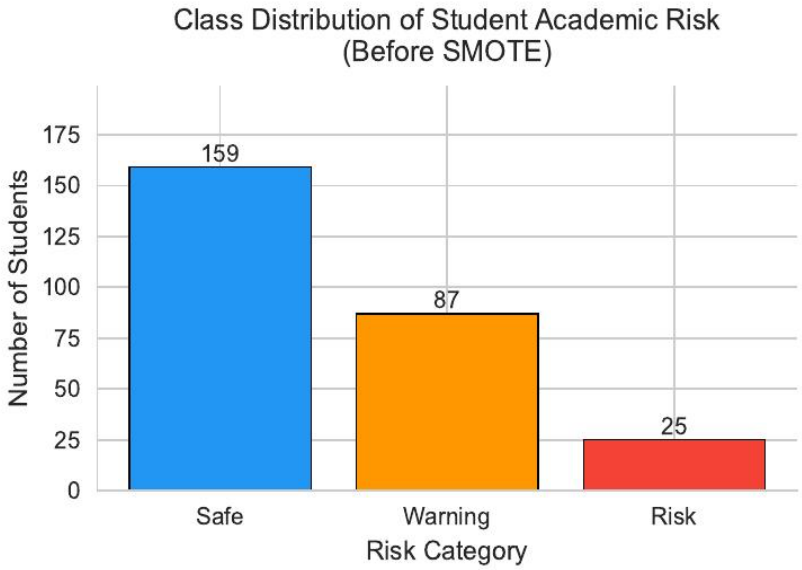

Fig.2 Class distribution of student academic risk before SMOTE (N = 271).

As presented in Table 4 and Figure 2, the dataset exhibits a pronounced class imbalance. The Safe class is the largest group, comprising 159 students (58.67%), followed by Warning with 87 students (32.10%), and Risk with only 25 students (9.23%). This imbalance means that a naive classifier could achieve approximately 58.67% accuracy merely by predicting all instances as Safe, which would render the model practically useless for identifying at-risk students. This finding confirmed the necessity of applying SMOTE prior to model training.

Class Balancing Using SMOTE

To address the class imbalance identified in the training partition, the Synthetic Minority Oversampling Technique (SMOTE) was applied exclusively to the training set before model fitting. Figure 3 presents a comparison of class distribution in the training set before and after SMOTE.

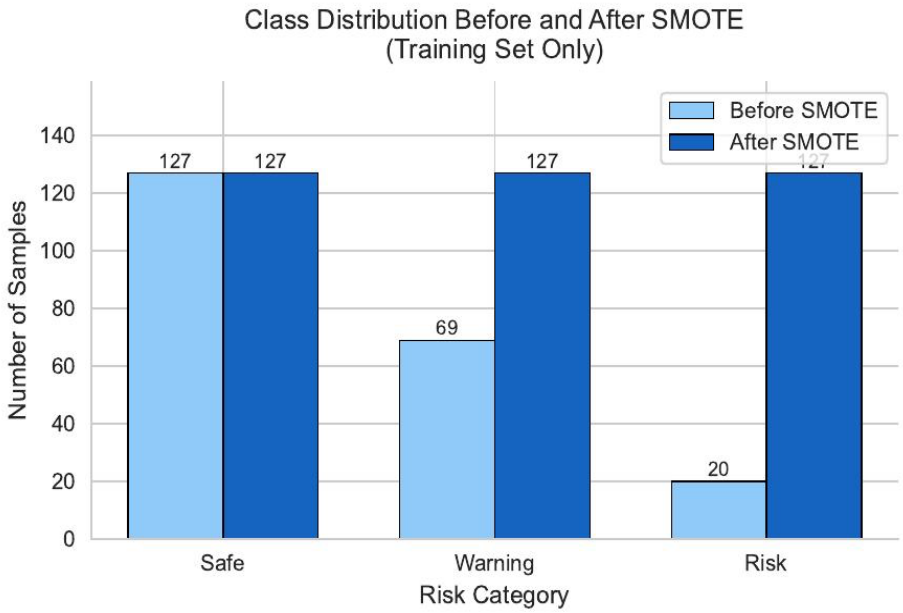

Fig. 3 Comparison of class distribution in the training set before and after SMOTE.

As illustrated in Figure 3, prior to SMOTE the training set contained 127 Safe, 69 Warning, and 20 Risk samples, reflecting the original imbalance. After applying SMOTE, synthetic samples were generated for the Warning and Risk classes until all three classes reached 127 samples each, yielding a fully balanced training set of 381 samples in total. Importantly, SMOTE was applied only within the training partition to prevent any

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

information from the test set from influencing the synthetic sample generation process, thereby avoiding data leakage.

Gradient Boosting Model Configuration

The Gradient Boosting classifier was trained using the SMOTE-balanced training set. Table 4 summarizes the hyperparameter configuration applied to the final model.

Table 5
Hyperparameter Configuration of the Gradient Boosting Model

Parameter	Description	Value
n_estimators	Number of estimators (M)	300
learning_rate	Learning rate (η)	0.03
max_depth	Maximum tree depth	3
min_samples_split	Minimum samples required to split a node	5
min_samples_leaf	Minimum samples required in a leaf node	2
subsample	Subsample ratio of training data	0.8
random_state	Random seed for reproducibility	42

As shown in Table 5, the model employs 300 sequential weak learners (n_estimators = 300), each trained to correct the residual errors of the previous ensemble. A low learning rate of 0.03 (η = 0.03) was chosen to allow gradual and controlled refinement of predictions, reducing the risk of overfitting. Individual trees are constrained to a maximum depth of 3, ensuring each learner remains weak and promotes diversity within the ensemble. The minimum sample thresholds (min_samples_split = 5; min_samples_leaf = 2) further restrict tree growth. A subsample ratio of 0.8 introduces stochastic variability into each training iteration, which enhances generalization. All experiments were conducted with random_state = 42 to ensure full reproducibility of results.

Cross-Validation Results

Stratified 5-Fold Cross Validation was employed to evaluate the generalization capability of the model across the full dataset. By preserving the class proportion in each fold, this strategy ensures that every partition is a representative sample of the overall distribution. Table 6 reports the per-fold and aggregate results, while Figure 4 displays the distribution of each metric as box plots.

Table 6
Stratified 5-Fold Cross-Validation Results

Metric	Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Mean	Std
Accuracy	0.7091	0.6667	0.7593	0.7963	0.7963	0.7455	0.0508
Precision	0.7067	0.6340	0.7118	0.8370	0.7266	0.7232	0.0653
Recall	0.6880	0.5610	0.7440	0.7740	0.7714	0.7077	0.0796
F1-Score	0.6889	0.5862	0.7257	0.8013	0.7407	0.7086	0.0711

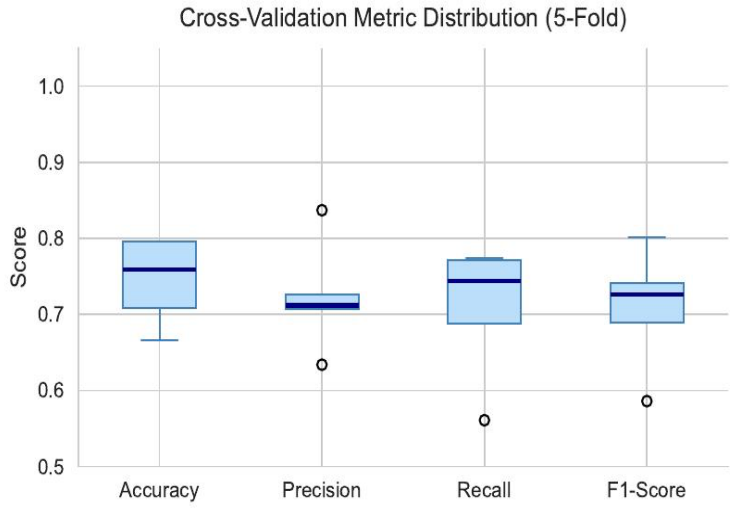

Fig. 4 Distribution of cross-validation performance metrics across five folds.

As presented in Table 5 and Figure 4, the model achieves a mean accuracy of 74.55% (SD = 5.08%), mean precision of 72.32% (SD = 6.53%), mean recall of 70.77% (SD = 7.96%), and mean F1-Score of 70.86% (SD =

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

7.11%) across all five folds. The box plots in Figure 3 confirm that most metric distributions are compact and centered in the 0.70–0.80 range, with outliers primarily attributable to Fold 2, which records the lowest performance across all metrics. This may be due to an unfavorable class arrangement within that fold's test partition. Folds 4 and 5 consistently yield the strongest results, with accuracy reaching approximately 79.63%. The overall consistency across folds confirms that the model generalizes well and is not overfitting to any particular data partition.

Model Evaluation on the Test Set

The final Gradient Boosting model was evaluated on a held-out test set consisting of 55 student records that were not used during training or cross-validation. Table 7 presents the complete per-class classification report alongside overall performance metrics.

Table 7
Classification Report and Overall Performance Metrics on the Test Set

	Precision	Recall	F1-Score	Support
Safe	0.8485	0.8750	0.8615	32
Warning	0.7059	0.6667	0.6857	18
Risk	0.8000	0.8000	0.8000	5
Macro Average	0.7848	0.7806	0.7824	55
Weighted Average	0.7974	0.8000	0.7984	55
Overall Accuracy	0.8000	—	—	—
Cohen's Kappa (κ)	0.6306	—	—	—
Matthews Correlation Coefficient (MCC)	0.6311	—	—	—

As presented in Table 7, the model attains an overall accuracy of 80.00% on the test set. At the class level, the Safe category achieves the highest F1-Score of 86.15%, with precision of 84.85% and recall of 87.50%, reflecting strong and balanced performance for the majority class. The Warning class records an F1-Score of 68.57%, which is lower due to the inherent difficulty of distinguishing borderline-performing students from those in adjacent categories. The Risk class achieves an F1-Score of 80.00%, with equal precision and recall of 80.00%, demonstrating that the model reliably identifies students who require immediate academic intervention despite the small sample size of this class in the test set.

Beyond raw accuracy, Cohen's Kappa of 0.6306 and MCC of 0.6311 provide chance-corrected measures of agreement between predicted and actual labels. Both values fall within the range conventionally interpreted as substantial agreement, confirming that the model's performance is genuinely superior to random classification and remains reliable across all three risk categories under multi-class imbalanced conditions.

The confusion matrix in Figure 5 provides a class-level breakdown of the model's predictions on the test set.

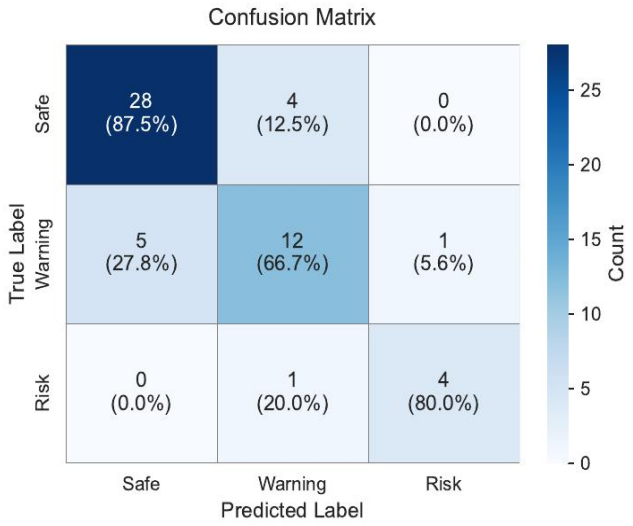

Fig.5 Confusion matrix of the Gradient Boosting model on the test set (n = 55).

As illustrated in Figure 5, the model correctly classifies 28 of 32 Safe students (87.5%), 12 of 18 Warning students (66.7%), and 4 of 5 Risk students (80.0%). The most frequent misclassification occurs at the Safe–Warning boundary: 5 Warning students are incorrectly predicted as Safe (27.8%), and 4 Safe students are predicted

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

as Warning (12.5%). This pattern is expected given the overlapping academic characteristics between these two adjacent risk groups. Notably, no Safe student is misclassified as Risk, and no Risk student is misclassified as Safe, indicating that the model maintains clear separation between the extreme risk categories a critical property for any early warning system intended for practical academic intervention.

ROC Curve Analysis

To assess the discriminative capability of the Gradient Boosting model, Receiver Operating Characteristic (ROC) curves were generated using a One-vs-Rest (OvR) strategy for each class. Figure 6 presents the ROC curves alongside the micro-average curve.

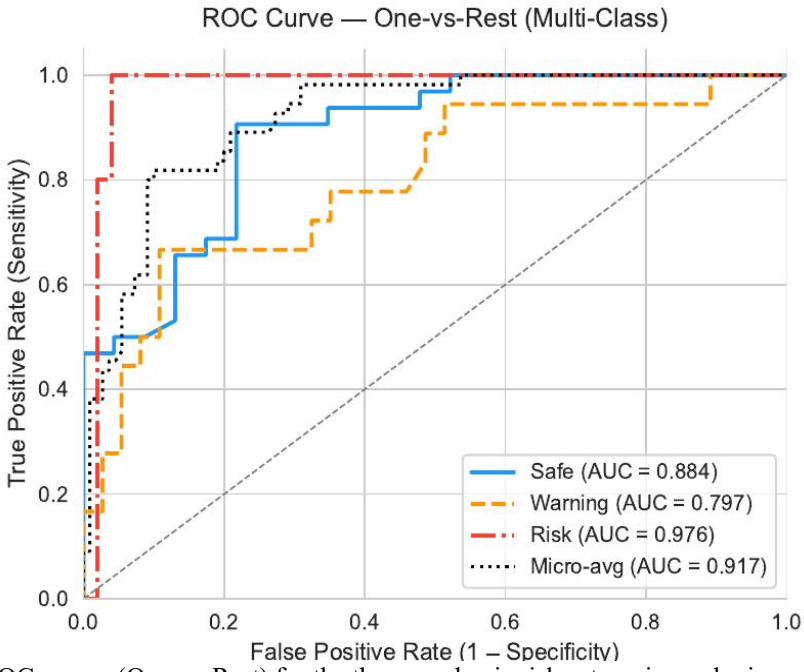

Fig.6 ROC curves (One-vs-Rest) for the three academic risk categories and micro-average.

As shown in Figure 6, the Risk class achieves the highest AUC of 0.976, indicating near-perfect discriminative ability for identifying students in the highest risk category. The Safe class attains an AUC of 0.884, reflecting strong separability between Safe and non-Safe students. The Warning class yields an AUC of 0.797, which remains acceptable and reflects the inherent challenge in separating borderline students from the adjacent Safe and Risk groups. The micro-average AUC of 0.917 aggregates performance across all classes, confirming the model's overall strong discriminative capability. Collectively, AUC values above 0.79 for all classes indicate that the Gradient Boosting model provides reliable probabilistic estimates across the full risk spectrum, making it well-suited for deployment in an early warning system.

Feature Importance Analysis

To interpret the contribution of each academic variable to the model's predictions, feature importance scores were extracted directly from the trained Gradient Boosting model. Table 8 and Figure 7 present the complete feature importance ranking.

Table 8
Feature Importance Scores Extracted from the Gradient Boosting Model

Feature	Importance Score	Rank
Midterm Score	0.4349	1
Quiz Score	0.2560	2
Assignment Score	0.2199	3
Attendance	0.0893	4

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

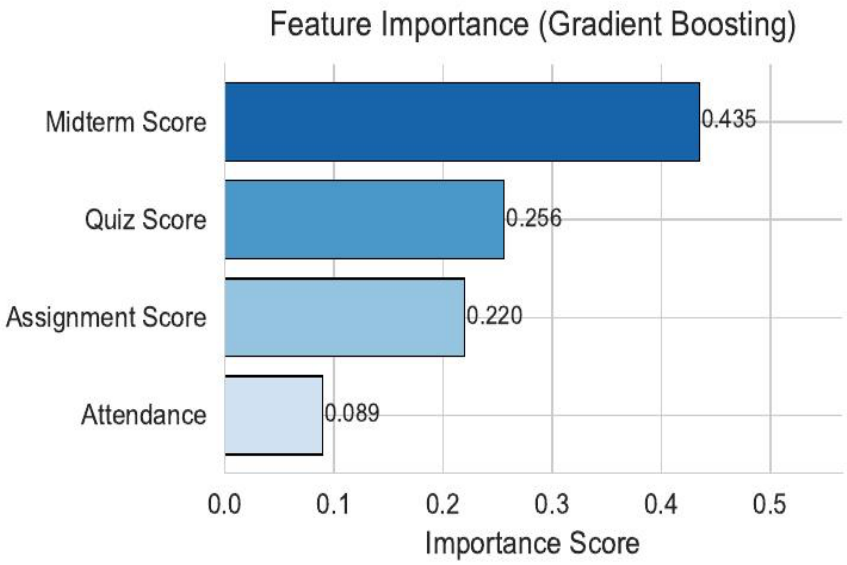

Fig.7 Feature importance ranking from the trained Gradient Boosting model.

As shown in Table 8 and Figure 7, Midterm Score emerges as the most influential predictor with an importance score of 0.4349 (43.49%), making it the dominant variable in the model's risk classification decisions. Quiz Score ranks second at 0.2560 (25.60%), followed by Assignment Score at 0.2199 (21.99%). Attendance contributes the least among the four variables at 0.0893 (8.93%), though it still provides a meaningful supplementary signal.

Taken together, the three score-based variables Midterm Score, Quiz Score, and Assignment Score account for approximately 91.07% of the model's total predictive weight. This finding underscores that academic performance indicators, particularly examination and quiz results, are the primary determinants of a student's risk classification. The dominance of Midterm Score suggests that students' performance in the mid-semester examination is the single strongest signal of academic risk, which is practically significant because midterm results become available well before the final examination.

Attendance, while ranked last, should not be dismissed entirely. Its contribution of approximately 8.93% indicates that class participation provides complementary information that supplements the score-based predictors. Students with low attendance may exhibit risk patterns that are not fully captured by their quiz or assignment scores alone.

From a practical standpoint, these results offer clear and actionable guidance for academic advisors: prioritizing the monitoring of midterm examination results, followed by quiz and assignment performance, enables earlier and more targeted identification of at-risk students. Intervention strategies aligned with these indicators such as supplemental tutoring triggered by low midterm scoresmay substantially improve student retention and academic outcomes.

DISCUSSIONS

Model Performance

The proposed early warning academic system based on Gradient Boosting with SMOTE achieved an overall accuracy of 80.00% on the held-out test set, with a macro-average F1-Score of 78.24%, Cohen's Kappa of 0.6306, and MCC of 0.6311. These results indicate that the model provides reliable and consistent predictive performance across three academic risk categories: Safe, Warning, and Risk.

In the context of class imbalance, the achieved performance is particularly meaningful. A naïve majority-class classifier would yield approximately 58.67% accuracy by always predicting the Safe class. Compared with a naïve majority-class classifier (58.67% accuracy), the proposed model improves predictive accuracy by 21.33 percentage points, demonstrating that SMOTE effectively mitigates class imbalance and enables the model to learn discriminative patterns from minority classes.

Cross-validation results further confirm model stability. The mean accuracy across five stratified folds was 74.55% (SD = 5.08%), with a mean F1-Score of 70.86% (SD = 7.11%). The relatively low variance indicates stable generalization performance across different data partitions, with only Fold 2 showing reduced performance due to likely minority-class variability in that split.

Class-Level Analysis

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

Safe Class

The Safe class achieved the strongest performance, with an F1-Score of 86.15%, precision of 84.85%, and recall of 87.50%. This result is consistent with its dominance in the dataset (58.67%), which provides sufficient representative patterns for learning. High recall indicates that most well-performing students are correctly identified, minimizing unnecessary intervention.

Warning Class

The Warning class obtained the lowest F1-Score (68.57%), with precision of 70.59% and recall of 66.67%. This reflects the inherent difficulty of distinguishing Warning students due to overlapping feature distributions with both Safe and Risk categories. Misclassification patterns confirm this overlap, particularly between Safe and Warning classes. Despite this limitation, the recall level remains practically useful for early detection, as approximately two-thirds of at-risk monitoring cases are correctly identified.

Risk Class

The Risk class achieved an F1-Score of 80.00%, with balanced precision and recall. Despite representing only 9.23% of the dataset and a very small test subset, the model maintains strong detection capability. The ROC-AUC score of 0.976 further confirms excellent separability for this critical class. Importantly, no Safe students were misclassified as Risk, indicating strong boundary preservation for high-risk identification, which is essential in early warning applications.

Effectiveness of SMOTE

SMOTE played a crucial role in improving minority class learning. Before balancing, the dataset showed severe imbalance (127 Safe vs 69 Warning vs 20 Risk in training data), which would bias the classifier toward majority predictions. By generating synthetic samples, SMOTE enabled balanced class representation (127 per class), ensuring that Gradient Boosting could optimize decision boundaries equally across all categories. The strong performance on the Risk class and high AUC value suggest that SMOTE contributed to improved minority class separability without substantially degrading majority class performance.

Feature Importance Analysis

Feature importance analysis reveals that Midterm Score is the most influential predictor (43.49%), followed by Quiz Score (25.60%) and Assignment Score (21.99%). Together, these three features account for more than 91% of total predictive contribution. This indicates that academic performance variables, particularly mid-semester evaluation results, are the primary determinants of student risk classification. The strong influence of Midterm Score suggests that mid-semester performance is a critical indicator of learning trajectory. Attendance contributes a smaller proportion (8.93%) but still provides complementary behavioral information, especially for borderline cases where score-based indicators are insufficient.

Comparison with Related Studies

The findings of this study are consistent with previous research showing that ensemble learning methods and academic performance indicators are effective for student performance prediction. However, unlike many existing studies that focus on binary classification, the proposed framework classifies students into three risk categories (Safe, Warning, and Risk), providing more actionable support for early intervention. In addition, the integration of SMOTE, Gradient Boosting, and feature importance analysis enables reliable prediction under class imbalance while offering interpretable insights for academic decision-making. Furthermore, this study contributes empirical evidence from an Indonesian higher education institution, a context that remains relatively underexplored in previous research.

Practical Implications

The proposed model can be integrated into existing academic information systems because it relies only on routinely collected academic indicators. The three-level risk classification enables institutions to prioritize academic interventions according to students' risk levels before final examination results become available, thereby supporting more timely and targeted decision-making.

Limitations

This study has several limitations. First, the dataset is relatively small (n = 271), with a particularly limited number of Risk samples, which may affect generalization. Second, only four academic variables were used, excluding other potentially important factors such as socioeconomic background, behavioral patterns, or psychological factors. Third, although SMOTE improves class balance, synthetic samples may not fully capture

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

real-world minority class distributions. Finally, external validation using datasets from other institutions was not conducted, limiting cross-institution generalizability.

CONCLUSION

This study developed an Early Warning Academic System for predicting student academic risk using Gradient Boosting combined with SMOTE-based class balancing. The proposed framework utilized four academic indicators, namely Quiz Score, Attendance, Assignment Score, and Midterm Score, to classify students into three categories: Safe, Warning, and Risk. Experimental results demonstrated that the model achieved strong predictive performance, with an accuracy of 80.00%, a macro F1-score of 78.24%, Cohen's Kappa of 0.6306, and MCC of 0.6311. These results confirm that the proposed approach is robust under imbalanced class conditions and significantly outperforms a majority-class baseline. The ROC-AUC analysis further indicates strong discriminative capability, particularly for identifying high-risk students.

Feature importance analysis revealed that Midterm Score is the most dominant predictor, followed by Quiz Score and Assignment Score, while Attendance contributes a smaller but supportive role. These findings indicate that academic performance indicators, especially mid-semester evaluation results, are sufficient to construct an effective early warning system without requiring complex or non-academic features. Future research should focus on expanding the dataset across multiple institutions, incorporating behavioral and socioeconomic variables, and evaluating more advanced ensemble models such as XGBoost and LightGBM to further enhance predictive performance and generalizability.

STATEMENT ON THE USE OF GENERATIVE AI

Generative AI was used solely to improve the grammar, clarity, and sentence flow of this manuscript. Specifically, ChatGPT (OpenAI, GPT-5.5) was used to assist in refining the academic writing and enhancing the readability of the text. The generative AI was not used to generate research ideas, collect or analyze data, interpret results, or draw scientific conclusions. All outputs generated by the AI were carefully reviewed, verified, and substantially edited by the authors, who take full responsibility for the accuracy, originality, and integrity of the final manuscript.

REFERENCES

Akello, E. F., Ijiga, O. M., & Idoko, I. P. (2026). Sequence-Aware Learning Analytics for Early Identification of At-Risk Academic Trajectories in Higher Education Using Transformer Models. *International Journal of Innovative Science and Research Technology*, 818. https://doi.org/10.38124/ijisrt/26jan563

Albreiki, B., Habuza, T., & Zaki, N. (2023). Extracting topological features to identify at-risk students using machine learning and graph convolutional network models. *International Journal of Educational Technology in Higher Education*, *20*(1), 23. https://doi.org/10.1186/s41239-023-00389-3

Almalawi, A., Soh, B., Li, A., & Samra, H. (2024). Predictive Models for Educational Purposes: A Systematic Review. *Big Data and Cognitive Computing*, *8*(12), 187. https://doi.org/10.3390/bdcc8120187

Alomari, S. A., Aldiabat, K., Hashim, F. A., Zitar, R. A., Liu, Z., Migdady, H., Hu, G., Smerat, A., & Abualigah, L. (2025). Supervised Learning: Teaching Machines with Labeled Data. In L. Abualigah, *Mastering the Minds of Machines* (1st ed., pp. 26–33). CRC Press. https://doi.org/10.1201/9781003516385-4

Aluso, L., & Enyejo, J. O. (2025). Using XGBoost and Time-Series Forecasting to Predict Student Academic Trajectories in Educational Analytics Platforms. *International Journal of Innovative Science and Research Technology*, 143. https://doi.org/10.38124/ijisrt/25dec159

Angeioplastis, A., Aliprantis, J., Konstantakis, M., & Tsimpiris, A. (2025). Predicting Student Performance and Enhancing Learning Outcomes: A Data-Driven Approach Using Educational Data Mining Techniques. *Computers*, *14*(3), 83. https://doi.org/10.3390/computers14030083

Aslam, Z., Missen, M. M. S., Ghaffar, A. A., Mehmood, A., Villar, M. G., Alvarado, E. S., & Ashraf, I. (2025). Advancing fake news combating using machine learning: A hybrid model approach. *Knowledge and Information Systems*, *67*(12), 12137–12177. https://doi.org/10.1007/s10115-025-02588-y

Bagawade, R. P., & Thirupurasundari. (2025). Ensemble Machine Learning Techniques. *International Journal of Emerging Technology and Advanced Engineering*, *15*(4), 15–23. https://doi.org/10.46338/ijetae0425_02

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

Călin, S. (2025). Handling Imbalanced Data: The SMOTE Technique. *2025 17th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 1–5. https://doi.org/10.1109/ECAI65401.2025.11095450

Charaya & Sonia. (2026). Towards Accurate Student Performance Prediction: An Assessment of Machine Learning Models and Metrics. *International Journal of Latest Technology in Engineering Management & Applied Science*, *15*(1), 600–610. https://doi.org/10.51583/IJLTEMAS.2026.150100053

Chukwura. (2023). A comparative study of several classification metrics and their performances on data. *World Journal of Advanced Engineering Technology and Sciences*, *8*(1), 308–314. https://doi.org/10.30574/wjaets.2023.8.1.0054

Elreedy, D., Atiya, A. F., & Kamalov, F. (2024). A theoretical distribution analysis of synthetic minority oversampling technique (SMOTE) for imbalanced learning. *Machine Learning*, *113*(7), 4903–4923. https://doi.org/10.1007/s10994-022-06296-4

Ersozlu, Z., Taheri, S., & Koch, I. (2024). A review of machine learning methods used for educational data. *Education and Information Technologies*, *29*(16), 22125–22145. https://doi.org/10.1007/s10639-024-12704-0

Ganage, S. S. (2026). Machine Learning Techniques for Student Performance Prediction and Academic Analytics. *International Journal of Engineering Research and Science & Technology*, *22*(1(S)), 311–314. https://doi.org/10.62643/ijerst.2026.v22.i1(S).2064

Ge, X., Fang, C., Li, X., Sun, W., Wu, D., Zhai, J., Lin, S.-W., Zhao, Z., Liu, Y., & Chen, Z. (2025). Machine Learning for Actionable Warning Identification: A Comprehensive Survey. *ACM Computing Surveys*, *57*(2), 1–35. https://doi.org/10.1145/3696352

Giap, N., Nghiem, T. L., Ngo, T. H., Nguyen, M. T. L., & Phung, H. Q. (2024). Increment of Academic Performance Prediction of At-Risk Student by Dealing With Data Imbalance Problem. *Applied Computational Intelligence and Soft Computing*, *2024*(1), 4795606. https://doi.org/10.1155/2024/4795606

Idhmad, A., Kaicer, M., Nejjar, C., & Benjouad, A. (2024). Intelligent Bankruptcy Prediction Models Involving Corporate Governance Indicators, Financial Ratios and SMOTE. *Indonesian Journal of Electrical Engineering and Informatics (IJEEI)*, *12*(1), 233–244. https://doi.org/10.52549/ijeei.v12i1.5241

Kaban, R., Rachman, A. I., Sari, I. P., Zulherry, A., Basri, M., Fadilah, N., Maulidiyah, N., Sembiring, D. J., Pratama, G. U., Ronal, R., & others. (2026). *Pengantar Artificial Intelligence, Machine Learning, dan Big Data*. Yayasan Kita Menulis.

Kabumba, R. K., Mbayandjambe, A. M., Tashev, T., Slavov, V., Kambale, W. V., Ngoie, R.-B. M., Kyamakya, K., & Kasereka, S. K. (2025). Predictive Modeling of Academic Success Using Machine Learning: A Comparative Analysis with SMOTE-Based Class Balancing. *Procedia Computer Science*, *272*, 242–250. https://doi.org/10.1016/j.procs.2025.10.202

Li, G., & Sheng, H. (2025). A hybrid machine learning framework for predicting aircraft scaled sound pressure levels: A comparative study. *Journal of Engineering and Applied Science*, *72*(1), 163. https://doi.org/10.1186/s44147-025-00714-9

Ling, Q. (2023). Machine learning algorithms review. *Applied and Computational Engineering*, *4*(1), 91–98. https://doi.org/10.54254/2755-2721/4/20230355

Liu, H. (2025). Machine Learning Approaches for Predicting and Intervening in Students' Academic Status. *Applied and Computational Engineering*, *183*(1), 102–109. https://doi.org/10.54254/2755-2721/2025.BJ27254

Lokanan, M. (2024). *Exploring Resampling Techniques in Credit Card Default Prediction*. In Review. https://doi.org/10.21203/rs.3.rs-4087259/v1

Lu, X. (2024). Modern Education: Advanced Prediction Techniques for Student Achievement Data. *International Journal of Advanced Computer Science and Applications*, *15*(1). https://doi.org/10.14569/IJACSA.2024.01501126

Lünich, M., Keller, B., & Marcinkowski, F. (2024). Fairness of Academic Performance Prediction for the Distribution of Support Measures for Students: Differences in Perceived Fairness of Distributive Justice Norms. *Technology, Knowledge and Learning*, *29*(2), 1079–1107. https://doi.org/10.1007/s10758-023-09698-y

JCEIT: Journal of Computer Engineering and Information Technology
Volume 2, Number 3, July 2026
Page: 191-207
ISSN 3089-106X
(online)

Mujahid, M., Kına, E., Rustam, F., Villar, M. G., Alvarado, E. S., De La Torre Diez, I., & Ashraf, I. (2024). Data oversampling and imbalanced datasets: An investigation of performance for machine learning and feature engineering. *Journal of Big Data*, *11*(1), 87. https://doi.org/10.1186/s40537-024-00943-4

Nafea, A. A., Mishlish, M., Haban Shaban, A. M., AL-Ani, M. M., Alheeti, K. M. A., & Mohammed, H. J. (2023). Enhancing Student'sPerformance Classification Using Ensemble Modeling. *Iraqi Journal for Computer Science and Mathematics*, *4*(4). https://doi.org/10.52866/ijcsm.2023.04.04.016

Nguyen, C., Thi Tran, T., Thanh Thi Nguyen, T., Ha Thi Nguyen, T., Astarkhanova, T. S., Van Vu, L., Tai Dau, K., Nguyen, H. N., Pham, G. H., Nguyen, D. D., Prakash, I., & Pham, B. (2024). Mapping of soil erosion susceptibility using advanced machine learning models at Nghe An, Vietnam. *Journal of Hydroinformatics*, *26*(1), 72–87. https://doi.org/10.2166/hydro.2023.327

Osborne, J. B., & Lang, A. S. I. D. (2023). Predictive Identification of At-Risk Students: Using Learning Management System Data. *Journal of Postsecondary Student Success*, *2*(4), 108–126. https://doi.org/10.33009/fsop_jpss132082

Ozyurt, O., Ozyurt, H., & Mishra, D. (2023). Uncovering the Educational Data Mining Landscape and Future Perspective: A Comprehensive Analysis. *IEEE Access*, *11*, 120192–120208. https://doi.org/10.1109/ACCESS.2023.3327624

Pan, J., Zhao, Z., & Han, D. (2025). Academic Performance Prediction Using Machine Learning Approaches: A Survey. *IEEE Transactions on Learning Technologies*, *18*, 351–368. https://doi.org/10.1109/TLT.2025.3554174

Paul, C., & Devi, A. (2025). AI-Driven Early Warning Systems for Dropout Risk and Passed-Out Prediction in Higher Education. In Dr. Mohd. A. Gandhi, Dr. A. Dutt, Dr. V. Saravanan, & K. T. Umare (Eds.), *AI, Machine Learning, and IoT Applications for Academic Performance Prediction, Faculty Well-Being, and Educational Outcomes in Higher Education* (2025th ed., pp. 89–110). RADemics Research Institute. https://doi.org/10.71443/9789349552258-04

Rahmi, N. A., Defit, S., & Okfalisa. (2024). Enhancing Classification Performance: A Study on SMOTE and Ensemble Learning Techniques. *2024 International Conference on Future Technologies for Smart Society (ICFTSS)*, 63–68. https://doi.org/10.1109/ICFTSS61109.2024.10691339

Rajasekhar, C., & Fernandes, A. S. H. D. (2026). Student Performance Prediction Using Machine Learning and Deep Learning. In R. Shanthi, *Artificial Intelligence–Driven Intelligent Learning Systems for Language Proficiency, Student Success, and Well-Being in Higher Education* (2026th ed., pp. 275–297). RADemics Research Institute. https://doi.org/10.71443/9789349552401-12

Rakitin, I. V., Rakitina-Qureshi, E. V., Arabadzhieva-Kalcheva, N. A., Todorova, M. P., Penev, I. P., & Petrova, D. I. (2025). Predicting Student Placement Status Using Tree-Based Ensemble Classifiers: A Comparative Analysis of Random Forest and Gradient Boosting Models. *2025 International Conference Automatics and Informatics (ICAI)*, 53–58. https://doi.org/10.1109/ICAI67591.2025.11325082

Rawat, B., & Pant, H. (2025). Boosting vs. Traditional Algorithms for Student Performance Prediction. *2025 International Conference on Sustainable Communication Networks and Application (ICSCN)*, 1998–2003. https://doi.org/10.1109/ICSCN67106.2025.11308228

Rico-Juan, J. R., Cachero, C., & Macià, H. (2024). Study regarding the influence of a student's personality and an LMS usage profile on learning performance using machine learning techniques. *Applied Intelligence*, *54*(8), 6175–6197. https://doi.org/10.1007/s10489-024-05483-1

Rochmadi, T. (2025). Addressing Class Imbalance in Predicting Student Academic Outcomes: A Comparative Study of Resampling Techniques with Machine Learning Classifiers in Higher Education. *Artificial Intelligence in Learning*, *1*(3), 211–227. https://doi.org/10.63913/ail.v1i3.32

Sathvik & Nagaveni Biradar. (2025). Student Performance Detection Using Machine Learning Techniques. *International Research Journal on Advanced Engineering Hub (IRJAEH)*, *3*(09), 3513–3516. https://doi.org/10.47392/IRJAEH.2025.0516

Sharanya, B. (2024). An Efficient Ensemble Machine Learning Model for Cardiovascular Disease Prediction Using Digital Health Records. *International Journal for Research in Applied Science and Engineering Technology*, *12*(3), 2598–2609. https://doi.org/10.22214/ijraset.2024.58888

Sharma, Y. P., & Sasikala, K. (2026). Early Warning Systems for Student Dropout Risk Prediction. In R. Shanthi, *Artificial Intelligence–Driven Intelligent Learning Systems for Language Proficiency,*

Student Success, and Well-Being in Higher Education (2026th ed., pp. 298–321). RADemics Research Institute. https://doi.org/10.71443/9789349552401-13

Sujon, K. M., Hassan, R., Khairudin, A. R., Moi, S. H., Mohd Shafie, M. L., Saringat, Z., & Erianda, A. (2024). The Effects of Imbalanced Datasets on Machine Learning Algorithms in Predicting Student Performance. *JOIV : International Journal on Informatics Visualization*, *8*(3–2), 1599. https://doi.org/10.62527/joiv.8.3-2.2449

Teresa, & et. al. (2025). *Predictive Modelling of Student Outcomes Using Ensemble Regression and Classification Methods*. https://doi.org/10.5281/ZENODO.17599070

Tiwari, M., & Jain, N. (2024). STUDENT PERFORMANCE PREDICTION USING MACHINE LEARNING ALGORITHMS. *ShodhKosh: Journal of Visual and Performing Arts*, *5*(6). https://doi.org/10.29121/shodhkosh.v5.i6.2024.4552

Villar, A., & Andrade, C. R. V. (2024). Supervised machine learning algorithms for predicting student dropout and academic success: A comparative study. *Discover Artificial Intelligence*, *4*(1), 2. https://doi.org/10.1007/s44163-023-00079-z

Yu, L., Zhao, X., Huang, J., Hu, H., & Liu, B. (2023). Research on Machine Learning with Algorithms and Development. *Journal of Theory and Practice of Engineering Science*, *3*(12), 7–14. https://doi.org/10.53469/jtpes.2023.03(12).02

Yuningsih, L., Pradipta, G. A., Hermawan, D., Ayu, P. D. W., Hostiadi, D. P., & Huizen, R. R. (2023). IRS-BAG-Integrated Radius-SMOTE Algorithm with Bagging Ensemble Learning Model for Imbalanced Data Set Classification. *Emerging Science Journal*, *7*(5), 1501–1516. https://doi.org/10.28991/ESJ-2023-07-05-04