e-ISSN: 2655-867X
p-ISSN: 2655-8661
Volume 8 No. 2 August 2026

Comparative Analysis of CatBoost, Random Forest, and XGBoost for Predicting Student Academic Performance

Yayu Nur Faidah[1)], Agung Rachmat Raharja[2)], Andri Feriansyah[2)], Ahmad Labudi[2)]
[1)] Health Information Management, Bandung University
[2)] Informatics, Bandung University
Email: yayu.nrfdh@universitasbandung.ac.id; agungrachmat@universitasbandung.ac.id; andriferiyansyah@universitasbandung.ac.id; ahmadlabudi@universitasbandung.ac.id.

Accepted:	Accepted After Revision:	Published:
15 July 2026	21 August 2026	28 August 2026

Abstract

Predicting student academic performance has become an important topic in educational data mining because it enables educational institutions to identify students who may require academic support at an early stage. This study compares the performance of three ensemble learning algorithms—CatBoost, Random Forest, and XGBoost—in predicting students' final academic grades using the Student Performance Dataset obtained from Kaggle. The dataset contains 649 student records with demographic, family, behavioral, and academic information. Before model development, the data were preprocessed through categorical feature encoding and feature engineering, resulting in five additional variables: parent_education, previous_grade, total_alcohol, attendance_category, and study_efficiency. The dataset was divided into training and testing sets using an 80:20 ratio. Hyperparameter optimization was performed only for CatBoost using the Optuna framework with 20 optimization trials, while Random Forest and XGBoost were trained using predefined parameter settings. Model performance was assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). The results indicate that CatBoost achieved the best overall performance, with an RMSE of 1.2888 and an R^2 of 0.8297, outperforming the other two algorithms. Feature importance analysis also revealed that G2 and previous_grade were the strongest predictors of students' final academic performance. These results suggest that CatBoost is a reliable approach for student performance prediction and can support data-driven academic decision-making.

Keywords: Student Academic Performance, CatBoost, Random Forest, XGBoost, Educational Data Mining.

1 INTRODUCTION

The rapid growth of information technology has significantly changed the way educational institutions manage academic data. Universities and schools now generate and store large amounts of information through learning management systems, student information systems, and academic assessment records. These data provide valuable opportunities to support data-driven decision-making, particularly in predicting student academic performance. Accurate prediction models enable educators to identify students who may be at risk of poor academic achievement at an early stage, allowing timely interventions and more personalized learning support [1], [2].

Student academic performance is shaped by a combination of factors, including demographic characteristics, family background, attendance, learning behavior, and previous academic achievement. These factors interact in complex ways, making it difficult for traditional statistical methods to accurately model their relationships. As a result, machine learning has become an increasingly popular approach in educational data mining because it can identify nonlinear patterns and generate more accurate predictions from large educational datasets [3], [4].

Among machine learning techniques, ensemble Ensemble learning improves predictive performance by combining the strengths of multiple learning models. Random Forest provides robust and stable predictions, XGBoost is valued for its computational efficiency, and CatBoost effectively handles categorical features while reducing prediction bias. Owing to these advantages, ensemble learning algorithms have been widely adopted for predicting student academic performance [5].

Previous studies have shown that machine learning, particularly ensemble methods, can effectively predict student academic performance. Random Forest and XGBoost have demonstrated competitive performance, while recent studies have also reported

promising results using CatBoost. However, these three algorithms have not been sufficiently compared under the same experimental conditions, and the contribution of feature engineering remains relatively limited in previous studies. Therefore, this study compares CatBoost, Random Forest, and XGBoost using the same Student Performance Dataset and incorporates engineered features to identify the most effective algorithm for predicting student academic performance.

Although many studies have applied machine learning to predict student academic performance, most have focused on evaluating individual algorithms or comparing only a few models. Furthermore, the majority of these studies used the original dataset without exploring the potential benefits of feature engineering. Consequently, comprehensive comparisons of CatBoost, Random Forest, and XGBoost that integrate engineered features are still relatively limited [6], [7].

This research compares the performance of CatBoost, Random Forest, and XGBoost in predicting student academic performance using the Student Performance Dataset. To improve the quality of the predictive information, several engineered features were developed, including parent_education, previous_grade, total_alcohol, attendance_category, and study_efficiency. The models were evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). In addition, feature importance analysis was conducted to identify the variables that contributed most to the prediction results. It is expected that the findings of this study will provide useful insights for the development of educational data mining applications and support data-driven academic decision-making in educational institutions.

2 LITERATURE REVIEW

2.1 Student Academic Performance Prediction

Predicting student academic performance has become an important application of educational data mining, as it helps educational institutions make more informed academic decisions. Early prediction allows educators to identify students who may experience learning difficulties and provide appropriate academic support before their performance declines. With the increasing availability of educational data, machine learning techniques have become a practical solution for analyzing demographic, behavioral, and academic factors to generate accurate predictions of student achievement [8], [9], [10].

Previous research has shown that students' academic achievement is influenced by a variety of factors, including previous academic performance, attendance, study habits, parental education, and socioeconomic background. As educational data have become more accessible, researchers have increasingly applied machine learning techniques to analyze these factors and develop more accurate prediction models. These advances have also strengthened the use of data-driven approaches to support academic planning and educational decision-making [6], [11].

2.2 Ensemble Machine Learning Algorithms

Ensemble learning has become a popular approach in machine learning because it combines multiple models to produce more accurate and reliable predictions. Compared with single learning algorithms, ensemble methods generally offer better generalization and are more robust when dealing with complex datasets. Among the most widely used ensemble algorithms are CatBoost, Random Forest, and XGBoost. Random Forest is known for its stable performance by aggregating the predictions of multiple decision trees, XGBoost improves prediction accuracy through an optimized gradient boosting framework, while CatBoost is particularly effective in handling categorical features and reducing prediction bias. Owing to these advantages, these algorithms have been widely adopted for both regression and classification tasks, including the prediction of student academic performance [12], [13].

2.3 Performance Evaluation Metrics

The performance of regression models is typically assessed using several evaluation metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). Each metric provides a different perspective on model performance. MAE reflects the average difference between the predicted and actual values, RMSE gives greater weight to larger prediction errors, and R^2 indicates how well the model explains the variation in the target variable. Together, these metrics offer a comprehensive evaluation of prediction accuracy and are widely used in educational data mining studies. In general, lower MAE and RMSE values indicate better predictive accuracy, whereas a higher R^2 value reflects a model with stronger explanatory power [12], [14], [15].

2.4 Performance Evaluation Metrics

Model performance can be evaluated using several regression metrics, including Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2)**.** These metrics provide different perspectives on prediction accuracy. MAE measures the average absolute difference between the actual and predicted values, RMSE gives greater weight to larger prediction errors, while R^2 measures the proportion of variance in the target variable explained by the model.

Mean Absolute Error (MAE) is calculated as:

$$\boldsymbol{MAE} = \frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} |\boldsymbol{y}_{\mathbf{i}} - \hat{\mathbf{y}}_{\mathbf{i}}|$$

Root Mean Square Error (RMSE) is calculated as:

$$\boldsymbol{RMSE} = \sqrt{\frac{\mathbf{1}}{\boldsymbol{n}} \sum_{\mathbf{i=1}}^{\mathbf{n}} (\boldsymbol{y}_{\mathbf{i}} - \hat{\mathbf{y}}_{\mathbf{i}})^{\mathbf{2}}}$$

The Coefficient of Determination (R^2) is calculated as:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

Where **n** represents the number of observations, $\mathbf{y_i}$ represents the actual value, $\hat{\mathbf{y}}_\mathbf{i}$ represents the predicted value, and $\bar{\mathbf{y}}$ represents the mean of the actual values. Lower MAE and RMSE values indicate better predictive performance, while a higher R^2 value indicates greater explanatory capability.

2.5 Previous Studies

Machine learning has been widely applied to predict student academic performance, with algorithms such as Decision Tree, Random Forest, Support Vector Machine, Artificial Neural Network, XGBoost, and CatBoost frequently reported in the literature. Many of these studies have shown that ensemble learning methods generally achieve higher predictive accuracy than conventional machine learning algorithms because they are better at modeling the complex relationships found in educational data. Despite these promising results, most existing studies have evaluated only one algorithm or compared a small number of models. In addition, the potential of feature engineering has received relatively little attention, even though it can provide additional information for improving prediction performance. To address this gap, the present study compares CatBoost, Random Forest, and XGBoost using the Student Performance Dataset together with several engineered features. The objective is to identify the most effective ensemble learning algorithm for predicting student academic performance while examining the contribution of engineered features to model performance.

Table 1. Summary of Previous Studies

Author	Algorithm	Dataset	Best Result	Research Gap
Ayulani et al. (2023)	Random Forest, XGBoost, LightGBM	LMS Student Dataset	LightGBM achieved the highest prediction accuracy (86.8%)	CatBoost was not evaluated [16].
Kumar et al. (2024)	Random Forest, XGBoost	Student Academic Dataset	XGBoost outperform ed Random Forest	CatBoost and feature engineering were not included [17].
Muhammady et al. (2024)	CatBoost, XGBoost, LightGBM	Student Performance Dataset	Average MAE ≈ 0.22 using boosting algorithms	Random Forest was not compared [18].
Fan et al. (2025)	Complementary CatBoost	Student Performance Dataset	RMSE = 1.1099 (Mathematics dataset)	Focused only on CatBoost-based methods [19].
Gul et al. (2025)	Decision Tree, Random Forest, XGBoost, CatBoost, AdaBoost, Gradient Boosting	Student Academic Dataset	CatBoost achieved the highest predictive accuracy	Feature engineering was not the primary focus [20].
Proposed Study	**CatBoost, Random Forest, XGBoost**	**Student Performance Dataset**	**MAE, RMSE, R^2 Evaluation**	**Comprehensive comparison of three ensemble algorithms combined with feature engineering (parent_education, previous_grade, total_alcohol, attendance_category, and study_efficiency).**

3 RESEARCH METHODS

3.1 Research Design

A quantitative experimental approach was adopted to evaluate and compare the performance of CatBoost, Random Forest, and XGBoost in predicting student academic performance. The research consisted of several sequential stages, including dataset preparation, preprocessing, feature engineering, model training, performance evaluation, and comparative analysis. Throughout the experiment, the same training and testing data, preprocessing procedures, and evaluation metrics were applied to all models to ensure that the comparison was conducted fairly and consistently.

The overall research framework is presented in Figure 1.

Figure 1. Research Workflow

3.2 Dataset and Preprocessing

The analysis was conducted using the **Student Performance Dataset** obtained from the Kaggle repository (larsen0966, *Student Performance Data Set*). Kaggle – Student Performance Data Set The dataset contains 649 student records describing demographic characteristics, family background, learning behavior, and academic performance, with **G3** representing the students' final grades as the prediction target. Before model development, the dataset was examined to ensure its quality through missing value and duplicate record checks. As no missing values or duplicate entries were found, the preprocessing stage focused on encoding categorical variables and transforming the data into a format suitable for machine learning algorithms. To further improve the predictive capability of the models, feature engineering was applied by generating five additional variables: *parent_education, previous_grade, total_alcohol, attendance_category,* and *study_efficiency*. These engineered features were designed to better represent students' academic background and learning behavior, providing richer information for the prediction models.

3.3 Machine Learning Models

Three ensemble learning algorithms were evaluated in this study: **CatBoost, Random Forest, and XGBoost**. These algorithms were selected because they have demonstrated strong predictive performance in regression tasks and are widely used in machine learning research. Random Forest provides a stable and robust bagging-based approach, while XGBoost and CatBoost employ gradient boosting to improve predictive performance. CatBoost was additionally selected because of its ability to effectively handle categorical features, which is relevant to the characteristics of the Student Performance Dataset. Hyperparameter optimization was applied only to the CatBoost model using the Optuna framework with 20 optimization trials, as CatBoost was selected as the primary candidate model for further parameter optimization. Random Forest and XGBoost were trained using predefined parameter settings as comparative reference models. The dataset was then divided into training and testing sets using an 80:20 ratio. The training data were used to build the prediction models, while the testing data were used to assess their predictive performance. The same data partition and evaluation metrics were applied to all three models to maintain consistency in the comparative evaluation.

3.4 Performance Evaluation

Model performance was assessed using three widely adopted regression metrics: Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These metrics were selected to evaluate prediction accuracy and provide a consistent basis for comparing the performance of CatBoost, Random Forest, and XGBoost. Each metric reflects a different aspect of model performance. MAE represents the average difference between the predicted and actual values, RMSE places greater emphasis on larger prediction errors, while R^2 indicates how well the model explains the variation in the target variable. In general, models with lower MAE and RMSE values and a higher R^2 value were considered to have better predictive performance.

3.5 Research Workflow

The overall research process followed six main stages. It began with collecting and preparing the Student Performance Dataset for analysis. The data then underwent preprocessing and feature engineering to improve data quality and generate additional variables that could enhance model performance. After preprocessing, the dataset was divided into training and testing sets using an 80:20 ratio. The training data were used to develop the CatBoost, Random Forest, and XGBoost models, with hyperparameter optimization applied only to CatBoost through the Optuna framework. The trained models were subsequently evaluated using MAE, RMSE, and R^2, and the results were compared to determine the best-performing algorithm. Finally, feature importance analysis was carried out to identify the variables that contributed most to predicting student academic performance.

4 RESULTS AND DISCUSSION

4.1 Dataset Exploration and Preprocessing

The analysis began with an exploration of the Student Performance Dataset to understand its characteristics and prepare it for machine learning. The dataset contains demographic, family, behavioral, and academic information from 649 secondary school students, with 33 original features. In this study, G3, which represents the students' final academic grade, was used as the target variable for prediction.

Table 2. Dataset Characteristics

Characteristic	Value
Dataset	Student Performance Dataset
Number of Records	649
Original Features	33
Target Variable	G3 (Final Grade)
Problem Type	Regression

The dataset contains both numerical and categorical variables that describe different aspects of students' backgrounds and academic performance. Numerical features include age, studytime, failures, absences, G1, G2, and G3, whereas the categorical variables represent demographic and family-related characteristics. Together, these variables provide a comprehensive basis for predicting students' academic performance.

To better understand the characteristics of the data, descriptive statistical analysis was performed on the numerical variables. As presented in Table 3, the average age of the students was 16.74 years. The mean scores for G1, G2, and G3 were 11.40, 11.57, and 11.91, respectively, indicating relatively consistent academic performance across the three assessment periods. Students recorded an average of 3.66 absences, although the distribution varied considerably among individuals.

Table 3. Descriptive Statistics of Selected Numerical Features

Feature	Mean	Std	Min	Max
Age	16.74	1.22	15	22
Study Time	1.93	0.83	1	4
Failures	0.22	0.59	0	3
Absences	3.66	4.64	0	32
G1	11.40	2.75	0	19
G2	11.57	2.91	0	19
G3	11.91	3.23	0	19

Data quality assessment indicated that the dataset contained no missing values or duplicate records. Therefore, all 649 student records were retained for further analysis. After the initial preprocessing stage, categorical variables were transformed into numerical values using Label Encoding to ensure compatibility with the machine learning algorithms. The dataset was then enriched through feature engineering by generating five new variables: *parent_education, total_alcohol, previous_grade, attendance_category,* and *study_efficiency*. These additional features summarize key aspects of students' academic history, learning behavior, and attendance, providing richer information for the prediction models.

Table 4. Newly Generated Features

Feature	Description
parent_education	Combined educational level of both parents
total_alcohol	Total alcohol consumption
previous_grade	Average of G1 and G2 grades
attendance_category	Attendance level based on absences
study_efficiency	Ratio between study time and absences

Following the preprocessing stage, the dataset contained 649 student records and 38 numerical features. Since all variables had been transformed into a machine learning–compatible format, the dataset was ready to be used for model development and evaluation.

Table 5. Final Dataset after Preprocessing

Characteristic	Before Preprocessing	After Preprocessing
Number of Records	649	649
Number of Features	33	38
Missing Values	0	0
Duplicate Records	0	0
Newly Generated Features	0	5
Data Representation	Numerical and categorical	Numerical
Dataset Status	Original	Ready for Machine Learning

Table 5 shows the changes in the dataset before and after preprocessing. The number of records remained unchanged at 649, while the number of features increased from 33 to 38 after five new features were generated. The data were also transformed from a combination of numerical and categorical variables into a numerical representation, resulting in a dataset ready for machine learning model development.

4.2 Machine Learning Model Implementation

After the data preparation process, three ensemble learning algorithms—CatBoost, Random Forest, and XGBoost—were developed and evaluated to predict students' final academic grades (G3). The remaining 37 features were used as predictor variables. The dataset was partitioned into training and testing sets with an 80:20 ratio, providing 519 records for model training and 130 records for model evaluation.

Table 6. Dataset Splitting Configuration

Parameter	Value
Total Records	649
Training Records	519
Testing Records	130
Training Ratio	80%
Testing Ratio	20%
Number of Features	37
Target Variable	G3

To improve the performance of the CatBoost model, hyperparameter optimization was carried out using the Optuna framework with 20 optimization trials. The best result was obtained in Trial 17, yielding the lowest RMSE of 1.2714. The corresponding hyperparameter settings included 712 boosting iterations, a tree depth of 6, a learning rate of 0.1538, and an L2 regularization value of 2.9521.

Table 7. Best Hyperparameters of CatBoost

Hyperparameter	Best Value
Iterations	712
Tree Depth	6
Learning Rate	0.1538
L2 Leaf Regularization	2.9521
Optimization Method	Optuna
Number of Trials	20
Best RMSE	1.2714

Unlike CatBoost, which underwent hyperparameter optimization, Random Forest and XGBoost were trained using predefined parameter settings. Random Forest was built with 300 decision trees and a maximum tree depth of 15. For XGBoost, the model was configured with 300 boosting trees, a learning rate of 0.05, a maximum tree depth of 6, and both subsampling and feature sampling ratios of 0.80.

Table 8. Model Configuration

Algorithm	Configuration
CatBoost	Optimized using Optuna (20 trials)
Random Forest	300 Trees, Max Depth = 15
XGBoost	300 Trees, Learning Rate = 0.05, Max Depth = 6

After the model configurations were established, all three algorithms were trained using the same training dataset and evaluated on the testing dataset. Their prediction results were then used to compare model performance, while the trained models were saved to support further analysis and future implementation.

4.3 Model Performance Evaluation

To compare the predictive capability of the three ensemble learning algorithms, model performance was assessed using MAE, RMSE, and R^2. These metrics were selected because they provide a comprehensive evaluation of regression models. In general, lower MAE and RMSE values indicate higher prediction accuracy, whereas a higher R^2 value reflects better explanatory performance. The evaluation results for CatBoost, Random Forest, and XGBoost are summarized in Table 9.

Table 9. Performance Comparison of Machine Learning Models

Model	MAE	RMSE	R^2
CatBoost	0.7963	**1.2888**	**0.8297**
Random Forest	**0.7786**	1.3089	0.8243
XGBoost	0.8030	1.3969	0.7999

Table 9 shows that CatBoost achieved the best overall predictive performance among the evaluated models. The algorithm recorded the lowest RMSE (1.2888) and the highest R^2 value (0.8297), meaning that it explained approximately 82.97% of the variation in students' final grades. Its MAE of 0.7963 also indicates that the average prediction error was less than one grade point, reflecting a high level of prediction accuracy.

Random Forest produced the lowest MAE (0.7786), suggesting that its average prediction error was slightly smaller than that of CatBoost. However, its RMSE was higher and its R^2 value was lower, indicating that CatBoost provided a better overall fit and handled larger prediction errors more effectively.

XGBoost showed the weakest performance of the three algorithms. It recorded the highest RMSE (1.3969) and the lowest R^2 value (0.7999), although it was still able to explain nearly 80% of the variation in the target variable. Compared with CatBoost and Random Forest, its prediction errors were generally larger.

Figure 2 illustrates the relationship between the actual and predicted student grades generated by the CatBoost model.

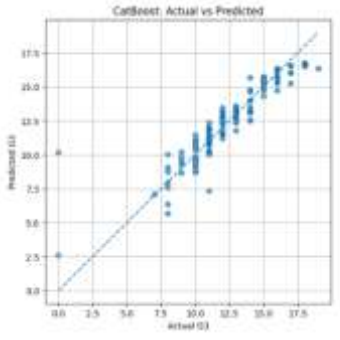

Figure 2. Actual versus Predicted Student Academic Performance Using CatBoost

Figure 2 illustrates a close agreement between the predicted and actual student grades generated by the CatBoost model. Most data points are distributed near the diagonal reference line, suggesting that the model

was able to predict students' final grades with a high level of accuracy. Only a small number of observations deviate noticeably from the reference line, mainly among students with lower academic performance. These results indicate that CatBoost effectively captured the relationship between the predictor variables and the target variable.

To provide a clearer comparison of model performance, the RMSE values of CatBoost, Random Forest, and XGBoost are presented in Figure 3.

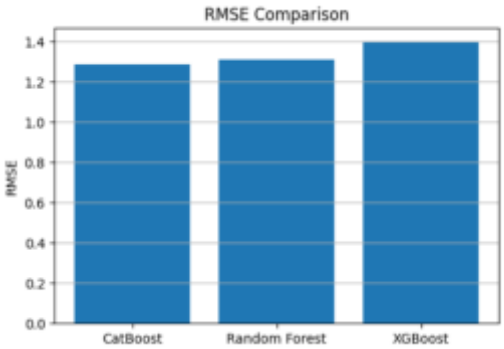

Figure 3. RMSE Comparison among Machine Learning Algorithms

Figure 3 compares the RMSE values obtained by the three ensemble learning algorithms. CatBoost recorded the lowest RMSE, followed by Random Forest, while XGBoost produced the highest value. Because RMSE reflects the magnitude of prediction errors, these results suggest that CatBoost generated the most accurate predictions. Although the difference between CatBoost and Random Forest was relatively small, CatBoost consistently achieved better overall performance.

The evaluation results indicate that CatBoost offers the best balance between prediction accuracy and model reliability, as reflected by its low prediction error and the highest R^2 value. Based on these findings, CatBoost was selected as the most appropriate algorithm for predicting student academic performance in this study.

4.3.1 Feature Importance Analysis

Feature importance analysis was performed to examine the contribution of each predictor variable to the machine learning models. This analysis provides insight into which variables played the greatest role in predicting students' final academic performance. The feature importance results obtained from the CatBoost model are summarized in Table 10.

Table 10. Top 10 Feature Importance of CatBoost

Rank	Feature	Importance (%)
1	G2	45.26
2	previous_grade	36.27
3	G1	8.07
4	absences	2.38
5	study_efficiency	2.34
6	failures	0.71
7	age	0.57
8	reason	0.44
9	health	0.43
10	Walc	0.26

The feature importance visualization is presented in Figure 4.

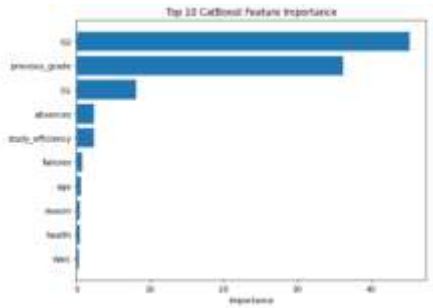

Figure 4. Top 10 Feature Importance Generated by the CatBoost Model

Figure 4 highlights the relative importance of the predictor variables in the CatBoost model. Among all features, G2 contributed the most, accounting for 45.26% of the total importance, followed by previous_grade (36.27%) and G1 (8.07%). Together, these three variables explained nearly 90% of the model's predictive power, emphasizing the dominant role of students' previous academic achievement in predicting their final grades.

Besides previous academic performance, absences and study_efficiency also contributed to the model, although their influence was considerably smaller. This suggests that regular attendance and effective study habits are associated with better academic outcomes. Other variables, such as failures, age, health, and alcohol consumption, had relatively low importance scores, indicating a more limited contribution to the prediction model.

The feature importance rankings obtained from the Random Forest and XGBoost models are discussed in the following sections and summarized in Tables 11 and 12.

Table 11. Top 10 Feature Importance of Random Forest

Rank	Feature
1	G2
2	previous_grade
3	absences
4	study_efficiency
5	age
6	total_alcohol
7	freetime
8	health
9	famrel
10	Mjob

Table 12. Top 10 Feature Importance of XGBoost

Rank	Feature
1	G2
2	previous_grade
3	absences
4	study_efficiency
5	school
6	failures
7	famsize
8	Dalc
9	freetime
10	health

The three algorithms produced a similar pattern of feature importance. Across all models, G2 and previous_grade consistently ranked as the two most influential predictors, followed by absences and study_efficiency. This result suggests that students' previous academic achievement and learning behavior play a central role in predicting their final academic performance.

The performance comparison also shows that CatBoost achieved better results than Random Forest and XGBoost. In addition to delivering the highest predictive accuracy, CatBoost provided a more balanced distribution of feature importance, making the model easier to interpret. These results demonstrate that CatBoost is a reliable ensemble learning algorithm for student academic performance prediction and has strong potential for supporting educational data mining and academic decision-making.

4.4 Comparative Analysis of Machine Learning Algorithms

This section presents a comprehensive comparison of the three ensemble machine learning algorithms evaluated in this study, namely CatBoost, Random Forest, and XGBoost. The comparison focuses on predictive performance, computational efficiency, and the overall ranking of the models. This analysis aims to identify the most suitable algorithm for predicting student academic performance using the Student Performance Dataset.

4.4.1 Performance Comparison

The predictive performance of the three machine learning algorithms was evaluated using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and the Coefficient of Determination (R^2). These evaluation metrics provide complementary information regarding prediction accuracy and model reliability.

Table 13. Performance Comparison of Machine Learning Algorithms

Model	MAE	RMSE	R^2
CatBoost	0.7963	**1.2888**	**0.8297**
Random Forest	**0.7786**	1.3089	0.8243
XGBoost	0.8030	1.3969	0.7999

Based on Table 13, CatBoost demonstrated the best overall predictive performance among the evaluated algorithms. Although Random Forest achieved the lowest MAE value (0.7786), CatBoost obtained the lowest RMSE (1.2888) and the highest R^2 value (0.8297). Since RMSE penalizes larger prediction errors more heavily than MAE, the lower RMSE achieved by CatBoost indicates better prediction consistency and stronger generalization capability.

Random Forest produced competitive prediction results, with only a slight difference compared to CatBoost. Meanwhile, XGBoost recorded the highest RMSE and the lowest R^2 value, indicating comparatively lower predictive performance on the Student Performance Dataset.

The comparison of the three evaluation metrics is illustrated in Figure 5.

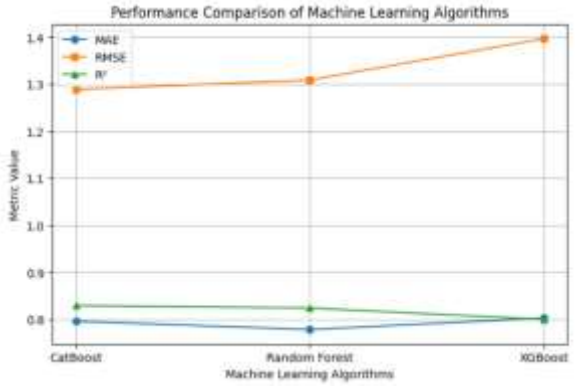

Figure 5. Performance Comparison of Machine Learning Algorithms

Figure 5 illustrates the trends of MAE, RMSE, and R^2 across the three evaluated algorithms. CatBoost consistently achieved the lowest RMSE and the highest

coefficient of determination, indicating superior predictive capability. Although Random Forest obtained a slightly lower MAE, the difference between CatBoost and Random Forest was relatively small. In contrast, XGBoost exhibited higher prediction errors and lower explanatory power, confirming its comparatively weaker performance.

4.4.2 Computational Efficiency

Besides prediction accuracy, computational efficiency is another important aspect in selecting an appropriate machine learning algorithm. Computational efficiency considers the complexity of model training, the requirement for hyperparameter optimization, and the computational resources required during model development.

Table 14. Computational Efficiency Comparison

Algorithm	Hyperparameter Optimization	Relative Training Time	Prediction Accuracy
CatBoost	Optuna (20 Trials)	High	Excellent
Random Forest	No	Low	Very Good
XGBoost	No	Medium	Good

Table 14 shows that CatBoost required the longest training time because it employed Optuna-based hyperparameter optimization consisting of twenty optimization trials. Although this optimization process increased computational cost, it significantly improved prediction accuracy and produced the best overall model performance.

Random Forest required the shortest training time because it used predefined hyperparameters and parallel tree construction. This characteristic makes Random Forest computationally efficient while maintaining competitive predictive performance.

XGBoost demonstrated moderate computational efficiency. Although its training process was relatively efficient, the prediction accuracy remained lower than those of CatBoost and Random Forest. These findings indicate that computational efficiency alone does not necessarily correspond to superior prediction performance.

4.4.3 Model Ranking

To summarize the comparative evaluation, the machine learning algorithms were ranked according to their overall predictive performance by considering prediction accuracy, model stability, and computational characteristics.

Table 15. Overall Ranking of Machine Learning Algorithms

Rank	Algorithm	Overall Performance
1	CatBoost	Excellent
2	Random Forest	Very Good
3	XGBoost	Good

The overall ranking is illustrated in Figure 6.

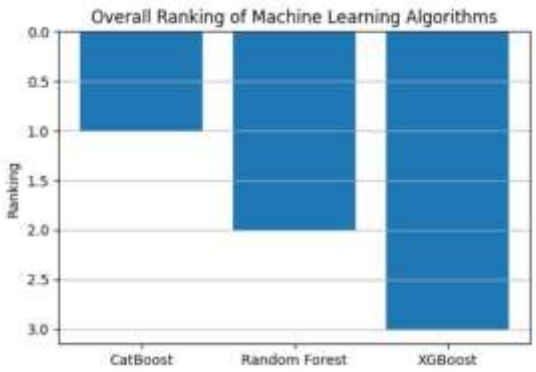

Figure 6. Overall Ranking of Machine Learning Algorithms

As shown in Figure 6, CatBoost ranked first among the evaluated algorithms by achieving the lowest RMSE and the highest R^2 value, indicating the best overall predictive performance. Random Forest ranked second, providing competitive prediction accuracy with efficient computational performance. Although it achieved a slightly lower MAE than CatBoost, its RMSE and R^2 values were comparatively lower.

XGBoost ranked third, producing the highest prediction error and the lowest coefficient of determination among the evaluated models. Despite its acceptable predictive capability, its overall performance was inferior to CatBoost and Random Forest. These results indicate that CatBoost is the most effective algorithm for predicting student academic performance using the Student Performance Dataset, demonstrating superior accuracy and strong generalization capability.

4.5 Feature Importance Analysis

Feature importance analysis was conducted to identify the predictor variables that contributed most significantly to the prediction of students' final academic performance. Three ensemble machine learning algorithms, namely CatBoost, Random Forest, and XGBoost, were analyzed individually to determine the relative importance assigned to each feature. Furthermore, a comparative analysis was performed to evaluate the consistency of feature rankings across the three algorithms.

4.5.1 CatBoost Feature Importance

The feature importance generated by the CatBoost model is presented in Figure 7.

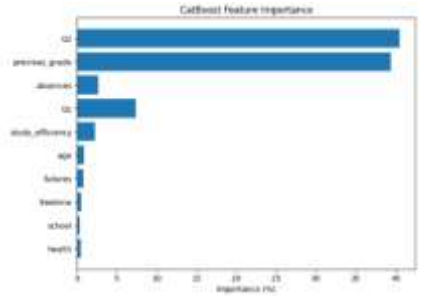

Figure 7. CatBoost Feature Importance

Figure 7 shows that G2 and previous_grade were the two most influential predictors, with importance values of 40.48% and 39.43%, respectively. Together, these variables accounted for nearly 80% of the total feature importance, highlighting previous academic achievement as the strongest predictor of students' final grades.

G1 ranked third with an importance value of 7.38%, followed by absences (2.65%) and study_efficiency (2.26%). Other variables, including age, failures, freetime, school, and health, contributed relatively little to the prediction model. Overall, these results indicate that CatBoost primarily relies on historical academic performance while incorporating attendance and learning behavior to improve prediction accuracy.

4.5.2 Random Forest Feature Importance

The feature importance produced by the Random Forest model is illustrated in Figure 8.

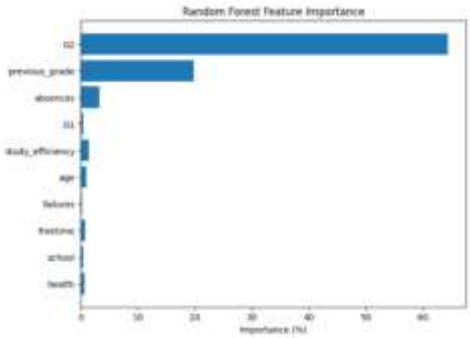

Figure 8. Random Forest Feature Importance

Figure 8 shows that G2 was the most influential feature in the Random Forest model, with an importance value of 64.32%, followed by previous_grade (19.79%) and absences (3.27%). Together, these variables contributed the majority of the model's predictive capability.

Compared with CatBoost, Random Forest assigned relatively low importance to G1 (0.43%), while variables such as study_efficiency, age, health, and school had only minor contributions. These results indicate that Random Forest relies primarily on G2 and previous_grade to predict students' final academic performance.

4.5.3 XGBoost Feature Importance

The feature importance generated by the XGBoost model is presented in Figure 9.

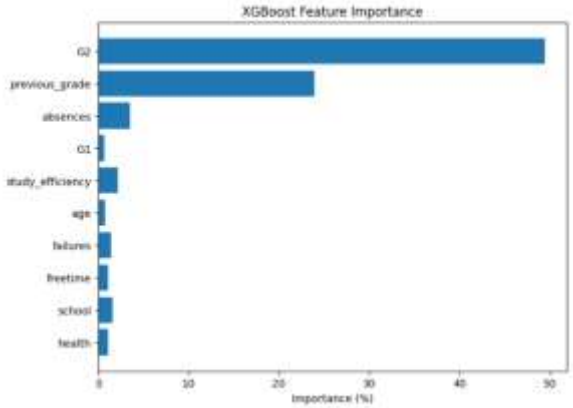

Figure 9. XGBoost Feature Importance

Figure 9 shows that G2 remained the most influential feature, accounting for 49.42% of the total importance, followed by previous_grade (23.93%) and absences (3.53%). These variables dominated the prediction process, highlighting the importance of previous academic achievement.

Compared with CatBoost and Random Forest, XGBoost assigned relatively higher importance to failures, school, study_efficiency, and health, although their contributions were considerably smaller. Overall, the results indicate that XGBoost utilizes a broader set of features while relying primarily on students' previous academic performance.

4.5.4 Comparative Analysis

To facilitate comparison among the three machine learning algorithms, the feature importance values were normalized and averaged. The comparative results are summarized in Table 16.

Table 16. Comparative Feature Importance

Rank	Feature	CatBoost (%)	Random Forest (%)	XGBoost (%)	Average (%)
1	G2	40.48	64.32	49.42	51.41
2	previous_grade	39.43	19.79	23.93	27.72
3	absences	2.65	3.27	3.53	3.15
4	G1	7.38	0.43	0.61	2.81
5	study_efficiency	2.26	1.39	2.14	1.93
6	age	0.90	0.92	0.74	0.85
7	failures	0.79	0.25	1.39	0.81
8	freetime	0.53	0.70	1.11	0.78
9	school	0.29	0.46	1.54	0.76
10	health	0.47	0.59	1.08	0.71

The comparative visualization is presented in Figure 10.

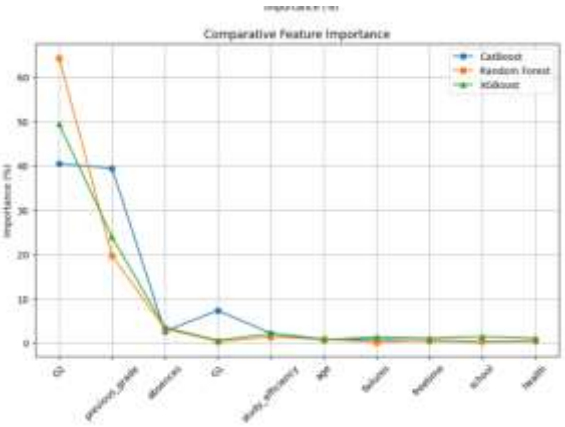

Figure 10. Comparative Feature Importance among CatBoost, Random Forest, and XGBoost

Figure 10 shows that all three algorithms consistently identified G2 and previous_grade as the most influential predictors of students' final academic performance, with average importance values of 51.41% and 27.72%, respectively. These results confirm that previous academic achievement is the dominant factor influencing prediction accuracy across all evaluated models.

Although the overall ranking of features was similar, each algorithm assigned different importance levels. Random Forest placed the greatest emphasis on G2, CatBoost distributed importance more evenly between G2 and previous_grade, while XGBoost assigned relatively higher importance to additional variables such as failures, school, and study_efficiency. In addition, absences consistently ranked among the top predictors, highlighting the role of attendance in academic performance prediction.

Overall, the feature importance analysis demonstrates that previous academic achievement, supported by attendance and learning behavior, provides the most valuable information for predicting students' final grades across the evaluated ensemble learning algorithms.

4.6 Discussion

This section discusses the findings obtained from the experimental evaluation, compares the results with previous studies, highlights the practical implications of the proposed approach, and presents the limitations of the current research. The discussion provides a comprehensive interpretation of the predictive performance achieved by the evaluated machine learning algorithms and their applicability in educational data mining.

4.6.1 Interpretation of Results

The experimental results show that CatBoost achieved the best overall predictive performance, with an RMSE of 1.2888 and an R^2 value of 0.8297. Although Random Forest obtained a slightly lower MAE, CatBoost provided better overall prediction accuracy by producing the lowest RMSE and the highest coefficient of determination among the evaluated algorithms.

The superior performance of CatBoost can be attributed to its gradient boosting mechanism, which effectively captures nonlinear relationships and reduces prediction bias. This capability enables the model to achieve more accurate and robust predictions than Random Forest and XGBoost.

Feature importance analysis further revealed that previous academic achievement was the strongest predictor of students' final grades. The variables G2, previous_grade, and G1 consistently ranked as the most influential features, while attendance and study_efficiency also contributed to prediction performance. These findings highlight the importance of both academic history and learning behavior in predicting student achievement.

5 CONCLUSION

This study compared the performance of CatBoost, Random Forest, and XGBoost for predicting student academic performance using the Student Performance Dataset. The experimental results showed that CatBoost achieved the best overall predictive performance, obtaining an RMSE of 1.2888 and an R^2 value of 0.8297. Although Random Forest produced the lowest MAE, CatBoost demonstrated better overall predictive capability by achieving lower RMSE and higher explanatory power. These findings indicate that CatBoost was the most suitable algorithm among the evaluated models for predicting students' final academic grades.

The feature importance analysis revealed that **G2** and *previous_grade* were the most influential variables in predicting academic performance, followed by *absences* and *study_efficiency*. These findings highlight the importance of previous academic achievement, attendance, and learning behavior in predicting students' final academic grades. From a practical perspective, the developed prediction model can support educational institutions in identifying students who may be at risk of poor academic performance and provide an additional basis for timely academic support and data-driven decision-making.

Despite these findings, this study has several limitations. First, the analysis was conducted using a single publicly available dataset containing 649 student records, which may limit the generalizability of the results to other educational contexts. Second, the models were developed using only the variables available in the dataset; therefore, other potentially relevant factors, such as learning motivation, socioeconomic conditions, and classroom engagement, were not considered. In addition, hyperparameter optimization was applied only to CatBoost, while Random Forest and XGBoost were trained using predefined parameter configurations.

Future research should consider larger and more diverse educational datasets to improve the generalizability of the findings. Further studies should also apply hyperparameter optimization consistently across all evaluated algorithms to provide a more comprehensive comparison. In addition, the integration of explainable artificial intelligence (XAI) and hybrid ensemble approaches may be explored to improve model interpretability and predictive performance. Overall, the findings demonstrate the potential of ensemble learning, particularly CatBoost, as a useful approach for student academic performance prediction and educational decision-support applications.

REFERENCES

[1] M. Yağcı, "Educational data mining: prediction of students' academic performance using machine learning algorithms," *Smart Learn. Environ.*, vol. 9, no. 1, p. 11, 2022, doi: 10.1186/s40561-022-00192-z.

[2] Syed Eirfan Atthar, "International Conference on Emerging Markets (ICEM) 2026 AI-Enabled Digital Twins for Low-Carbon Logistics in Emerging Markets: A Human-Centric Framework for Cold-Chain Energy Efficiency and CBAM-Ready Supply Chains in India 'A Human-Centric Framework for Energy-Efficient and Sustainable Cold-Chain Supply Chains,'" *Int. J. Res. Innov. Soc. Sci.*, vol. 10, no. 19, pp. 41–57, 2026, doi: 10.47772/IJRISS.

[3] Murnawan, S. Lestari, R. Samihardjo, and D. A. Dewi, "Sustainable Educational Data Mining Studies: Identifying Key Factors and Techniques for Predicting Student Academic Performance," *J. Appl. Data Sci.*, vol. 5, no. 3, pp. 1325–1342, 2024, doi: 10.47738/jads.v5i3.347.

[4] M. Bilal, M. Omar, W. Anwar, R. H. Bokhari, and G. S. Choi, "The role of demographic and academic features in a student performance prediction," *Sci. Rep.*, vol. 12, no. 1, p. 12508, 2022, doi: 10.1038/s41598-022-15880-6.

[5] H. ŞEVGİN, "A comparative study of ensemble methods in the field of education: Bagging and Boosting algorithms," *Int. J. Assess. Tools Educ.*, vol. 10, no. 3, pp. 544–562, 2023, doi: 10.21449/ijate.1167705.

[6] A. Villar and C. R. V. de Andrade, "Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study," *Discov. Artif. Intell.*, vol. 4, no. 1, p. 2, 2024, doi: 10.1007/s44163-023-00079-z.

[7] R. C. Lim and H. Harvianto, "Comparative Analysis of Decision Tree, Random Forest, and XGBoost for Student Category Prediction," *Int. J. Comput. Sci. Humanit. AI*, vol. 3, no. 1, p. 9, 2026, doi: 10.21512/ijcshai.v3i1.14936.

[8] C. Chaka, "Educational data mining, student academic performance prediction, prediction methods, algorithms and tools: an overview of reviews," *J. E-Learning Knowl. Soc.*, vol. 18, no. 2, pp. 58–69, 2022, doi: 10.20368/1971-8829/1135578.

[9] Z. Alamgir, H. Akram, S. Karim, and A. Wali, "Enhancing Student Performance Prediction via Educational Data Mining on Academic Data," *Informatics Educ.*, vol. 23, no. 1, pp. 1–24, 2024, doi: 10.15388/infedu.2024.04.

[10] W. Xiao, P. Ji, and J. Hu, "A survey on

educational data mining methods used for predicting students' performance," *Eng. Reports*, vol. 4, no. 5, p. e12482, May 2022, doi: https://doi.org/10.1002/eng2.12482.

[11] M. L. Murnawan Sri; Samihardjo, Rosalim; Dewi, Deshinta Arrova, "Sustainable Educational Data Mining Studies: Identifying Key Factors and Techniques for Predicting Student Academic Performance," *J. Appl. Data Sci.*, no. Vol 5, No 3: SEPTEMBER 2024, pp. 1325–1342, 2024, [Online]. Available: http://bright-journal.org/Journal/index.php/JADS/article/view/347/233

[12] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, "Data driven decisions in education using a comprehensive machine learning framework for student performance prediction," *Discov. Comput.*, vol. 28, no. 1, p. 153, 2025, doi: 10.1007/s10791-025-09585-3.

[13] N. Rohani, B. Rohani, and A. Manataki, "ClickTree: A Tree-based Method for Predicting Math Students' Performance Based on Clickstream Data," *J. Educ. Data Min.*, vol. 16, no. 2, pp. 32–57, 2024, doi: 10.5281/zenodo.13627655.

[14] W. Cui, A. Sangsongfar, and N. Amdee, "A Comparative Study of the Applicability of Regression Models in Predicting Student Academic Performance," *Naresuan Univ. Eng. J.*, vol. 19, no. 1, pp. 39–49, 2024.

[15] D. Ying and J. Ma, "Student Performance Prediction with Regression Approach and Data Generation," *Appl. Sci.*, vol. 14, no. 3, 2024, doi: 10.3390/app14031148.

[16] M. Kumar, N. Singh, J. Wadhwa, P. Singh, G. Kumar, and A. Qtaishat, "Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance," 2023.

[17] M. Kinter and N. E. Sherman, "Utilizing Random Forest and XGBoost Data Mining Algorithms for Anticipating Students' Academic Performance."

[18] D. N. Muhammady, H. A. E. Nugraha, V. R. S. Nastiti, and C. S. K. Aditya, "Students Final Academic Score Prediction Using Boosting Regression Algorithms," *J. Ilm. Tek. Elektro Komput. dan Inform.*, vol. 10, no. 1, p. 154, 2024, doi: 10.26555/jiteki.v10i1.28352.

[19] Z. Fan, J. Gou, and S. Weng, "Complementary CatBoost based on residual error for student performance prediction," *Pattern Recognit.*, vol. 161, p. 111265, 2025, doi: https://doi.org/10.1016/j.patcog.2024.111265.

[20] M. N. Gul, W. Abbasi, M. Z. Babar, A. Aljohani, and M. Arif, *Data driven decisions in education using a comprehensive machine learning framework for student performance prediction*, vol. 28, no. 1. Springer Netherlands, 2025. doi: 10.1007/s10791-025-09585-3.