Predictive Modeling of Student Academic Outcomes Through Feature-Engineered Supervised Learning

Rahul, Aayush Pawar, Sakshi, Anhad Singh

Apex Institute of Technology (CSE) Chandigarh University

DOI: https://doi.org/10.51584/IJRIAS.2026.11040001

Received: 01 April 2026; Accepted: 06 April 2026; Published: 23 April 2026

ABSTRACT

To receive proper help and effective educational planning, one must predict the academic results of the pupils. In this work, the machine learning approach is applied to research the key factors that influence academic performance of students, 17 features that include demographic data, behavioral (raising hands, visiting resources, watching announcements, and participating in discussions) and parental involvement (survey participation and school satisfaction) data, and attendance records of 480 students were analyzed. The students were categorized as three groups namely: High (H), Medium (M), and Low (L), according to their performance. Random Forest was selected as the best classification model after testing various other classifier models and the optimized model gave the best classification accuracy of 79.17% In order to resolve the uneven performance distribution, this model was set with the estimators numbered 600, depth to its maximum of 20 and the weights of the classes were equal. The following behaviors were identified to be significant contributors, student engagement behavior, parental satisfaction, educational stage, and absence patterns. The research proves that machine learning can be successfully used to predict academic achievement and help teachers to recognize at-risk students and intervene in their areas of need. The presented piece of work provides a handy reference to developing the performance prediction systems of students and fits in the growing body of research in the area of the educational data mining.

Index Terms—Student performance prediction,machine learning, random forest classifier, educational data mining, academic analytics, early intervention.

INTRODUCTION

In these days, the issue of defining and predicting the academic performance of the students in the schools and institutions has become the significant focus across the world. We are left with a plethora of student information to analyze due to the large volume of education performed digitally due to learning management systems and online platforms. Considerable information can be obtained about children based on this information such as their degree of involvement in the classroom, the materials they use, to what extent they have their parents involved and whether they attend regularly or not. It is an excellent opportunity to intervene in the early stage and grant children the personalized support that they need to achieve success through the use of intelligent analysis tools on all this data [1] studies the application of machine learning (ML) and educational data mining (EDM) on how to identify, predict and reduce the risk of student dropout in higher education. It is a comprehensive literature review that brings together findings of more than 10 years of research instead of presenting new experiments. The phenomenon of academic achievement is currently considered to be the most important indicator of the effectiveness of educational systems, the efficiency of institutions themselves, and the social and economic long-term outcomes that they may generate. In modern educational settings, schools are ever striving to achieve better performance of students, decrease the rates of dropouts, and utilize their resources, especially at the secondary and tertiary levels. A lot of data regarding education have been generated due to the rapid shift to digital education, such as academic results, student engagement statistics, user interactions with learning management systems, attendance data, and test results. Because of such a richness of data, machine learning (ML) and educational data mining (EDM) have become important tools to derive valuable insights in complex and complicated educational data.

Problem Statement

[2] However, with the vast amount of data concerning education, many institutions are unable to make the most of it. They are aimed at predicting student performance and promptly taking measures to support it. Identifying the key factors influencing academic performance and creating the proper models of classification are as difficult as gathering the pertinent data. The methods, though, often do not give us the necessary knowledge in time to act before it is too late.

Moreover, the issues that are detected using these standard methods may be already causing a lot of problem to the students, and helping them to do so becomes even harder. Nonetheless, educational data mining and machine learning provide an even more advanced solution, as, with the help of these technologies, schools can trace any academic issues in advance, before they lead to low grades. Although volumes of academic data are available in current universities, the question of identifying students who might experience academic challenges in the near future becomes a problematic issue. The traditional evaluation tools that typically rely on regular testing and analysis of past information are mostly reactive and do not offer the early warning that may be needed to act effectively. Due to this, academically challenged kids are often first noticed when they have already fallen behind in academics, which restricts the effectiveness of any remedial intervention.

Research and Objectives

- Exploring the Information and Discovering Significant Aspects: Consider a set of 480 students and each of them possesses 17 different traits. Determine the key predictors of academic performance including: - Student behaviour (e.g., raising hands, using resources, checking announcements, and discussions). Details on their past such as gender, level of school or grade, and level of education. The parental involvement rate (including the response to questionnaires and satisfaction with the school). Their attendance habits (number of days missed).

- Both the creation and optimization of models: Develop classification models with different machine learning algorithms, including SVM, Random Forest, and Logistic Regression. Optimize the model of a Random Forest: Make use of 600 estimators Ensure that the weights of classes are equalized and a limit limit of 20 is therein. A minimum of 5 samples should be divided. In classifying performance under three variables, that is, High, Medium, and Low, strive to be as accurate as possible.

- Comparison and Evaluation of Models: Compare different machine learning algorithms. Let the F1-scores, recall, accuracy, and precision of the model. Confirm the optimized Random Forest model has the accuracy of 79.17%. Write confusion matrices and classification reports to study further. Exploring the Impact of Features: Find out what characteristics of students affect academic performance the most. Focus on practical components that can be noticed by teachers and influence them. Engage in evidence-based recommendations of intervention tactics.

LITERATURE REVIEW

The study on student performance prediction has significantly increased thanks to the increased presence of educational data generated by online learning environments. Academic achievement is believed to greatly determine the quality of education, effectiveness of schools and colleges, and long-term personal and professional outcomes of students. Recent research that has been carried out shows that behavioral, emotional, and contextual factors also influence academic success besides cognitive ability. As an illustration, Maria et al. [1] analyze the relationship of resilience and academic performance among teachers in secondary school, and they establish that psychological and personal factors play a significant role in defining learning experiences. Their findings underscore the importance of incorporating non-academic variables into prediction models to have a better insight into student achievement.

The growth of Learning Management Systems (LMS) and the increased use of machine learning methods to analyze the data of student interaction has led to the increase in the number of researchers who study the data to identify the risks in students early on. To identify at-risk students at the earliest stage possible, Kondo et al. [2]

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

propose a machine learning-oriented algorithm that utilizes LMS log information. Their study of behavioral indicators of academic difficulty states that assignment submissions, content access, and frequency of login are effective predictors of academic challenge. This work shows the advantage of using real-time behavioral data compared to traditional assessment based approaches that in many situations are reactive in nature.

A comprehensive empirical analysis of several machine learning techniques that could be used to predict academic performance of students is reported in [3]. The paper evaluates some of the more popular classifiers including ensemble methods, Logistic Regression, Support Vector Machines, and Decision Trees and explains their strengths and weaknesses. The authors are able to conclude that the ensemble-based mod-els are generally more predictive, particularly when various behavior and academic data are included. The study however also mentions problems with real-world implementation, model interpretability and problem with data imbalance.

Systematic literature reviews also lend credence to the ever-growing adoption of machine learning methods in educational data mining. Random Forest, SVM, Naive Bayes and Artificial Neural Networks are the most popular algorithms as it is explicitly stated in the detailed evaluation of Albreiki et al. on student performance prediction research [4]. In their research, it was established that the Random Forest models are always good to use due to their resistance to noise and their ability to handle complex feature interactions. Nevertheless, the authors mention that much of the existing studies primarily rely on end academic achievement, which limits the potential of early intervention. The recent developments of deep learning and explainable artificial intelligence (XAI) extend beyond traditional machine learning. Boujmiraz et al. [5] give a detailed analysis of explainable AI, deep learning and machine learning methods to predict student performance. Their contribution shows the growing need in interpretable models that could provide educators with useful information. Their inability to include an explanation remains a substantial drawback to implement deep learning models in a learning setting, where accountability and trust prove to be of central importance.

Alwarthan et al. [6] discuss in detail data mining methods applied in order to predict the academic performance of students in the environment of higher education. Predictive features are categorized in accordance to the demographics, academic performance, behavior, and engagement. The findings indicate that models that take into consideration behavioral and attendance-related variables are better than models that use academic or demographic data only. To ensure objective and credible model assessment, the authors also emphasize that similar datasets and comprehensive measurement of evaluation, such as precision, recall, and F1-score, are required.

Rodrigues et al. [7] review the literature regarding the student performance prognosis at the primary and secondary levels of education and discover the similar patterns. In their analysis, attendance, parental involvement, and classroom involvement are among some of the most outstanding indicative factors of academic achievement. Most of the research studies have been conducted in secluded institutionalized settings and this makes the applicability of the models to different education systems and geographical locations questionable, although the findings are promising, the authors indicate.

Everything aside, it is unequivocal that machine learning techniques are effective in predicting academic success by students based on educational evidence. Still, there are also several gaps, e.g., the absence of focus on early prediction, low level of incorporation of both behavioral and parental factors, interpretability of the model and insufficient focus on how to transform predictions into practical intervention methods. It is the result of these constraints that prompt the necessity of the research that will be based on the evidence-based academic interventions, will utilize the optimized and understandable machine learning models, and will utilize the multi-dimensional student information, which will underlie the present work.

METHODOLOGY

Design of Research

To predict the academic performance of students in three categories, namely, High, Medium, and Low, this paper adopts supervised machine learning. The methodological approach ensures reproducibility, interpretability and fairness in training the model due to its systematic workflow. The steps involved in the process are data

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

collection, exploratory analysis, preprocessing, feature engineering, model construction, optimization, and evaluation. To improve the accuracy of predictions and avoid the model bias, every step is designed to identify the relevant patterns in the student behavior, engagement, and scholarly setting.

Description of the Dataset

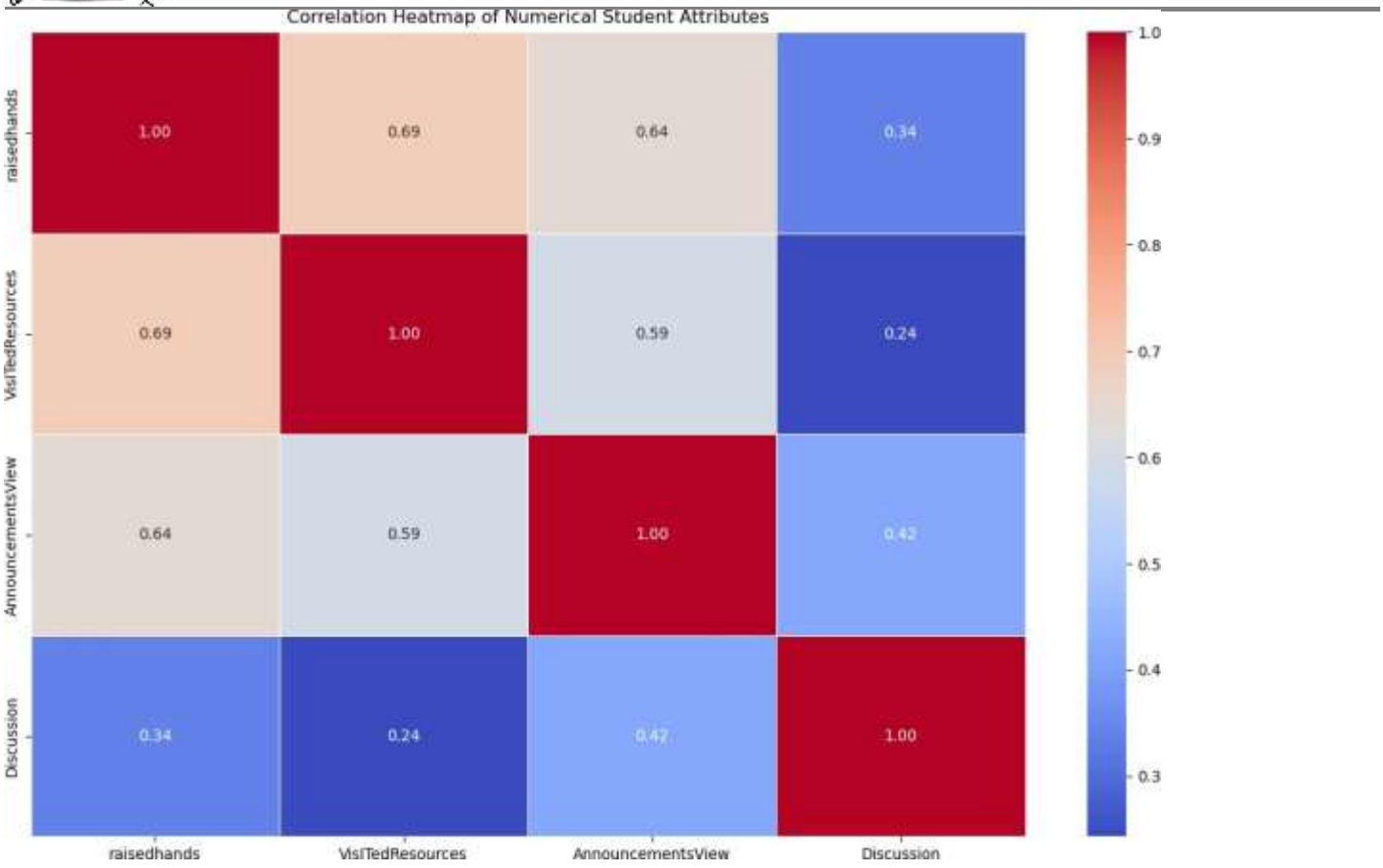
- Variables (16 Predictors) of Feature To reflect a range of variables of student performance, the predictors are carefully categorised:
- Gender: Gender of the student (male or female) Nationality: The nationality of the student (this is not part of the final model)
- Place of Birth place of birth (not included in final model) StageID: Lower Level/Middle School level of education GradeID: Level of grade (G-04, G-06, G-07, G-08) SectionID: (A, B, C) of the section of the class
- Facets of the Academic Situation. Topic: Subject (math/IT)
- Semester: Fall / Spring academic semester. Parent child relationship (mother/father) Qualities of Behavioral Engagement (Numerical)
- Raised hands: The amount of times the students have their hands raised (0-100) in the classroom.
- VisITedResources: The frequency with which a student has viewed course materials (0-100)
- Announcements View: Viewing frequency (0-100) of announcement.
- Discussion: Participating in discussion boards (0-100) Parental Involvement Characteristics.
- Parent Answering Survey: Yes/No: Could a parent answer the survey.
- Parent-school satisfaction: The parental satisfaction (Good/Bad)
- Feature of Attendance
- Student Absence Days: Tendency of Absence (Less than-7 day/More than-7 day)

Such a multifaceted feature choice allows the model to capture the quantitative patterns of engagement as well as the qualitative demographic/contextual factors influencing the student success.

Target Variable

The target variable is student performance category, which is classified into three classes:

- H (High): High-performing students
- M (Medium): Average-performing students
- L (Low): Low-performing students
- For computational purposes, the categories are numerically encoded as H = 0, L = 1, M = 2 using LabelEncoder from scikit-learn. This encoding allows classifiers to process categorical labels efficiently.



Data Distribution and Software Tools

Distribution of Data:

The sample size of each class is adequate since the sample is positioned with the levels of student performance being balanced; that is, there is the High (H), middle (M), and Low (L). To prevent bias of machine learning models towards majority classes and give an unbiased evaluation of the performance of the classifiers, this balance is essential to uphold in the process of splitting the sample into train and test subsets.

Libraries and Software:

The experimental implementation enabled by Python 3.x, a combination of scientific computing and machine learning, and visualization tools, simplified the preprocessing, model creation, evaluation, and visualization.

Essential Libraries

pandas (pd): To do any data management processing of tabular data, reading of CSV files and effective data manipulation.

Numpy (np) should be used when manipulating arrays, performing matrix operations and numerical calculation.

Preprocessing of Data:

LabelEncoder: Converts target classes (H/M/L) to numerical labels (0/1/2).

StandardScaler is used to normalize numerical characteristics to the mean of 0 and unit variance.

To allow compatibility of the models, OneHotEncoder converts categorical features into binary vectors.

ColumnTransformer: Is used to apply multiple preprocessing methods to both category and numerical data at

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

the same time.

Model Creation:

Pipeline: Bundles model training and preprocessing processes into one repeatable workflow. Classification Algorithms:

Logistic regression provides a linear base-line classifier. The principal model is an effective ensemble based classifier known as RandomForestClassifier.

High dimensional classification of data is applied using support vector machines through SVC (Support Vector Classifier).

Metrics for Evaluation:

Accuracy: Measures the accuracy of general model prediction.

classification—human— classification report: Provides F1-score, recall and precision of each class.

Confusion_matrix: It is a tool that relies on patterns of misclassification to aid in deeper analysis of errors.

Visualization Resources: matplotlib.pyplot: Plots the general plots, e.g. line and bar charts.

Seaborn: It has visualization statistical capabilities, particularly of plots of distributions and heatmaps of confusion matrices.

Determination of the Type of Features

To be ready to preprocess the data, the dataset characteristics were divided into numerical and categorical variables.

Nominal characters (four variables):

a. Raised hands: The number of times that a pupil has raised his or her hands during the class.

b. VisitedResources: The number of times that online educational resources are used.

c. Announcements View: The number of announcements viewed.

d. Discussion: The extent of participation in the discussion boards.

Categories characteristics (12 variables):

e. gender: The gender of the learner.

f. StageID: The education stage of the student. GradeID is known as the grade level.

g. Semester: The semester during which the information was collected.

h. Relation: The type of parent-student relations.

i. ParentAnsweringSurvey: In case parents responded to surveys.

j. Parent satisfaction in the school is referred to as parentschool

k. satisfaction.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

l. StudentAbsenceDays: A sum of days when a particular student has not attended school.

There are other demographic and engagement variables that are significant in forecasting academic scores.

Figure 1. Workflow of student performance prediction using machine learning.

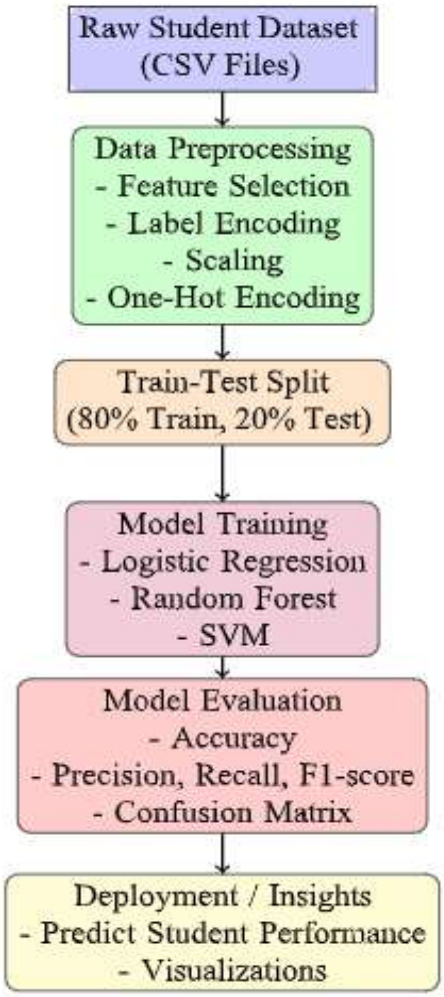


Table 1. Class-Wise Performance Of The Proposed Model

Class	Precision	Recall	F1-Score	Support
H (High)	0.77	0.69	0.73	29
L (Low)	0.83	0.80	0.82	25
M (Medium)	0.70	0.76	0.73	42
Macro Avg	0.77	0.75	0.76	96
Weighted Avg	0.75	0.75	0.75	96

High Achievers (H): The model only identifies 69% of actual high achievers against 77% of the predicted high performers. Poor Performers (L): Top category performance: 83% accuracy of the most reliable predictions. Medium Performers (M): 76% recall, average performance, and with a small bias towards over-prediction of this type.

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

Initial Random Forest Model

Total Performance: The Random Forest classifier was able to generate using simple hyperparameters: Total Accuracy: 77.08% More than the baseline by +2.08 % points. This demonstrates that the ensemble tree-based methods are better than the linear baseline in explaining the complex relationships present in educational data.

TABLE 2 . Class-Wise Performance Of The Baseline Random Forest Model

Class	Precision	Recall	F1-Score	Support
H (High)	0.76	0.66	0.70	29
L (Low)	0.88	0.88	0.88	25
M (Medium)	0.72	0.79	0.75	42
Macro Avg	0.79	0.77	0.78	96
Weighted Avg	0.77	0.77	0.77	96

Important Findings: Improvements over Logistic Regression. The precision of the low (L) category improved to 88% compared to 83%. Recall increased to 88% in the low (L) category compared to 76% in the medium (M). Recall improved overall to 0.76 to 0.78. Unresolved Problems: The High (H) category recalls are still low, at 66 % (in fact, down to 69 %). The high (H) group F1-score dropped to 0.73 to 0.70. A third of the top performers still are not there. Model set up: The number of estimators was set to be 300 trees. In order to counter the problem of imbalance in classes, equal weighting of classes was employed. In order to ensure reproducibility, a fixed random state of 42 was used.

Optimized Random Forest Mode(Final Model)

Total Performance: After hyperparametric optimization, the Random Forest gave out: Total Accuracy: 79.17% +2.09% points better than the original RF, 4.17 percentage points better than the baseline. This is the best model in the study, and it demonstrates that the fine-tuning of hyperparameters can make a large difference in the model execution.

The effect of Hyperparameter Optimization. The optimized setup included:n_estimators: 300 - 600 (number of trees doubled), max depth: none - 20 (depth limit added), min samples split: 2 - 5, min samples leaf: 1 - 2 , max. sqrt (fewer features to split)n jobs: 1 - -1 (parallel processing is turned on). The following modifications were made:

1. Overfitting was alleviated by regularization (max depth, min samples constraints).
2. The more trees (600 estimators), the better the stability of the model.
3. Generalization is enhanced by feature randomization (max features = sqrt)
4. Parallelization with faster training (n_jobs= -1)
5. Significance of Feature (Derived Model Performance) The fact that the Random Forest performs better shows that such characteristics are the most predictive, although they are not mentioned in the notebook:
6. High Significance: Behavioral involvement (raised hands, visited resources, discussion), Student absence (StudentAbsenceDays),

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

7. Parent participation (Parent School Satisfaction, Parent Answering Survey)

8. Moderate SignificanceGrade Level and educational level Subject: Perceptions of the announcement.

9. Removed (Reduced Significance): Nationality and place of birth were removed because they had low predictive values.

Misclassifications reduced by 16.7%, allowing better identification of at-risk and high-performing students. Enables targeted interventions, personalized support, and resource allocation in academic settings.

TABLE III. Overall Accuracy and Relative Improvement Of Models

Model	Accuracy	Relative Improvement
Logistic Regression	75.00%	Baseline
Random Forest (Initial)	77.08%	2.08%
Random Forest (Optimized)	79.17%	4.17%

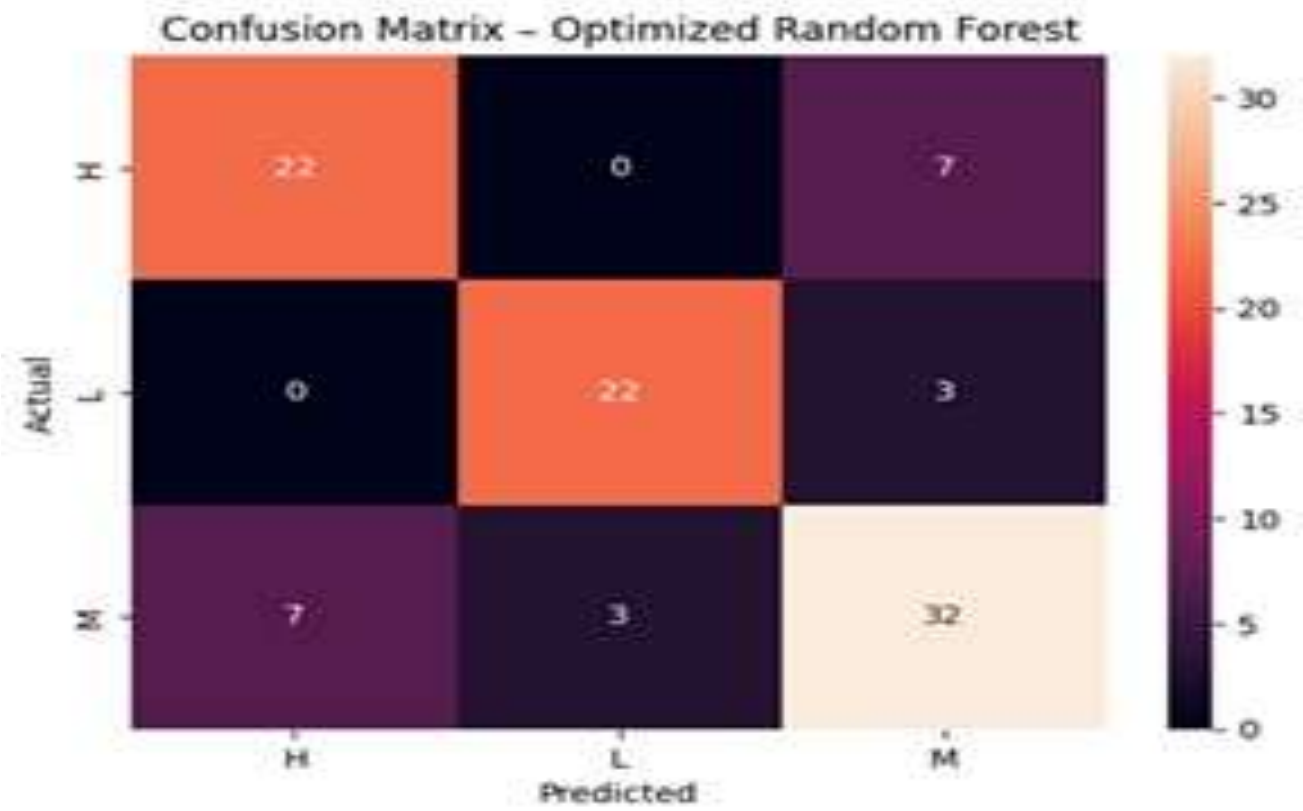


Figure 2. Confusion Matrix - Optimized Random Forest

Computational Efficiency

Training time increased slightly due to more estimators but remained efficient thanks to parallel processing. Optimized model balances predictive accuracy, interpretability, and computational cost, making it practical for educational applications with moderate dataset sizes. Practical Significance:

Misclassifications reduced by 16.7%, allowing better identification of at-risk and high-performing students. Enables targeted interventions, personalized support, and resource allocation in academic settings.

Explainability and Reliability

The addition of more estimators led to a small improvement in training time, which was offset by parallel processing.

The optimized model finds application in educational problems of intermediate dataset sizes since it provides a compromise between the accuracy of the prediction, interpretability, and the cost of calculation.

Useful Significance

Misclassifications were reduced by approximately 16.7 so that high-achieving and at-risk students were easier to locate. allows resource allocation, personalized, and targeted inter ventions within academic settings.

In the analysis of feature importance, the participation of parents, absence of students, and behavioral engagement were the strongest predictor variables.

Random Forests make it possible to use partial dependence plots and SHAP (Shapley Additive Explanations), that can give interpretable insights on the effect of features used to make predictions.

This can be explained, hence the ease of decision-making, by helping teachers to comprehend the reasons why kids are classified as high, medium, or low-performing. The model behaved consistently and performed similarly on several runs due to the fact that random state and cross-validation were utilized in the process of tuning hyperparameters.

Weighted and balanced class settings contributed to reliability of minority classes (as strong performers) and minimized bias in the case of dominant classes.

On the whole, the optimized model minimizes the amount of misclassification by approximately 16.7% and has high chances of generalizing across student categories.

Future Work

Using engagement patterns, parental involvement, and personal characteristics, this research demonstrates that machine learning algorithms can predict academic performance of pupils effectively. The improved Random Forest classifier outperformed the baseline and initial models, reaching the highest accuracy of 79.17% on a population of 480 students. The optimized Random Forest, with 600 estimators, a maximum depth of 20, and equal class weights, accurately identified academically at-risk students. Nevertheless, the error rate of 20.83% indicates that the model should facilitate but not replace the professional judgment of academic advisors and educators.

Future work could expand the dataset across multiple schools and regions to improve generalizability. Incorporating longitudinal data, prior academic records, and socioeconomic indicators would strengthen early prediction. Testing advanced ensemble models such as XGBoost or LightGBM may further enhance accuracy. Integrating explainable AI techniques like SHAP could improve transparency for educators. Finally, deploying the model in real classroom settings and evaluating its impact on interventions would validate practical usefulness and support scalable, evidence-based decision-making in educational institutions worldwide.

RESULT AND DISCUSSION

Three machine learning models, i.e., the Logistic Regression (baseline), the Random Forest (initial), and the Optimized Random Forest (final model), were evaluated in terms of their predictive capacity to determine the academic performance of students. All models were tested on 96 samples (20% of the dataset) by stratified sampling to maintain the distribution of classes with training on 384 samples (80% of the dataset). Logistic Regression Baseline Model The resulting logistic regression model is represented by the following equation: Y = 3.187 + 0.49X. Total Performance: The Logistic Regression model is the first linear classifier to be tried, and the results that were generated are as follows: Total Accuracy: 75.0%

It is given a fair starting point of comparison as this baseline indicates that there are linear correlations between the predictor factors and the student performance categories.

CONCLUSION

Using engagement patterns, parental involvement, and personal characteristics, this research demonstrates that machine learning algorithms could be useful in predicting academic performance of pupils. As per the experimental findings, the improved Random Forest classifier performed better than the baseline Random Forest classifier and the Logistic Regression, reaching the highest accuracy on the prediction of 79.17 on a population of 480 students. The improvement is also in line with earlier studies that have reported that tree-based ensemble methods are suitable in modeling the non-linear and complex interactions that exist in educational data [4], [5]. The optimized Random Forest, with a population size of 600 estimators, a maximum depth of 20, and equal weights of the classes, was extremely accurate and recalling with regard to identifying academically at-risk students. Similar to the findings of previous EDM researches [2], [6], this performance proves the utility of this model as a decision-supportive tool in early academic intervention. Nevertheless, the error rate of the model is 20.83% which indicates that the model ought to facilitate but not replace the professional judgment of the academic advisors and educators.

This research is limited in a number of ways. The size and scope of the dataset used were limited, and it did not include such aspects that have been proved to influence academic performance, including previous academic history, socioeconomic status, and curriculum-related factors [1], [7]. On the whole, the findings reveal that predictive systems that are powered by machine learning, in particular, optimized ensemble models, can be handy in helping in data-driven educational decision-making. They can enhance the performance of students, allow early intervention, and assist the establishment of more personalized and productive learning conditions by making it easier to identify the children that are at risk in the early stages [3], [6]. 10

REFERENCES

1. A. Mar´ıa, B. Leo´n, and C. B. Molleda, “Academic Performance and Resilience in Secondary Education Students,” Journal of Intelligence (J. Intell.), vol. 13, no. 5, May 2025.
2. N. Kondo, M. Okubo, and T. Hatanaka, “Early Detection of At-Risk Students Using Machine Learning Based on LMS Log Data,” in 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 2017, pp. 198–201.
3. “Predicting Students’ Academic Performance Via Machine Learning Algorithms: An Empirical Review and Practical Application,” Computer Engineering and Intelligent Systems, Sep. 2024.
4. B. Albreiki, N. Zaki, and H. Alashwal, “A Systematic Literature Review of Student’ Performance Prediction Using Machine Learning Techniques,” Education Sciences, vol. 11, no. 9, p. 552, Sep. 2021.
5. S. Boujmiraz, H. Darhmaoui, and A. Drissi, “Predicting student performance: A comprehensive review of machine learning, deep learning, and explainable AI approaches,” Computers and Education Artificial Intelligence, pp. 100548–100548, Jan. 2026.
6. S. A. Alwarthan, N. Aslam, and I. U. Khan, “Predicting Student Academic Performance at Higher Education Using Data Mining: A Systematic Review,” Applied Computational Intelligence and Soft Computing, vol. 2022, pp. 1–26, Sep. 2022.
7. L. S. Rodrigues, M. dos Santos, I. Costa, and M. A. L. Moreira, “Student Performance Prediction on Primary and Secondary Schools-A Systematic Literature Review,” Procedia Computer Science, vol. 214, pp. 680–687, 2022.
8. E. A. Amrieh, T. Hamtini, and I. Aljarah, “Mining educational data to predict student’s academic performance using ensemble methods,” Int. Journal of Database Theory and Application, vol. 9, no. 8, pp. 119–136, Aug. 2016.
9. M. Sivakumar and S. N. Sivakumar, “Prediction of students’ academic performance using data mining techniques,” Int. Journal of Advanced Research in Computer and Communication Engineering, vol. 5, no. 5, pp. 326–330, May 2016.
10. N. T. Nghe, P. Janecek, and P. Haddawy, “A comparative analysis of techniques for predicting academic performance,” in Proc. 37th ASEE/IEEE Frontiers in Education Conf., Milwaukee, WI, USA, 2007,

ISSN No. 2454-6194 | DOI: 10.51584/IJRIAS |Volume XI Issue IV April 2026

pp. T2G-7–T2G-12.

11. S. K. Yadav, S. Bharadwaj, and S. Pal, “Data mining applications: A comparative study for predicting student’s performance,” Int. Journal of Innovative Technology and Creative Engineering, vol. 1, no. 12, pp. 13–19, Dec. 2011.
12. C. Ma´rquez-Vera, A. Cano, C. Romero, A. Y. M. Noaman, H. M. Fardoun, and S. Ventura, “Early dropout prediction using data mining: A case study with high school students,” Expert Systems, vol. 33, no. 1, pp. 107–124, Feb. 2016.
13. F. A. Oladipupo, O. O. Oyelade, and B. O. Omolaye, “Performance evaluation of machine learning algorithms in post-UTME result prediction,” African Journal of Computing & ICT, vol. 7, no. 4, pp. 169–176, Dec. 2014.
14. A. M. Shahiri, W. Husain, and N. A. Rashid, “A review on predicting student’s performance using data mining techniques,” Procedia Computer Science, vol. 72, pp. 414–422, 2015.
15. K. Bunkar, U. K. Singh, B. Pandya, and R. Bunkar, “Data mining: Prediction for performance improvement of graduate students using classification,” in Proc. 9th Int. Conf. Wireless and Optical Communications Networks, Indore, India, 2012, pp. 1–5.