iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL
FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 14 Issue: V Month of publication: May 2026

DOI: https://doi.org/10.22214/ijraset.2026.82870

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

Predicting Students' Academic Performance with the Use of Machine Learning Techniques

Mr. Gajanan Ullas Revankar
Department of Computer Science St. Francis College Koramangala

Abstract: An early prediction of student performance is one significant part that improves the learning achievements, reduce the student attrition and personalized learning strategy. The traditional methods of performance assessment usually includes a summative grade scheme which do not provide early warning system for a struggling student. This paper demonstrates a supervised learning model for initial estimation of student achievement using academic, behavioral, and demographic factors. The proposed method for early prediction of student performance consists of preprocessing the data, feature selection, building the classifier models, and evaluation of the achievement of the classification models using various classification algorithms, namely, Logistic Regression, Support Vector Machine, Decision Tree, and Random Forest. The proposed method is evaluated based on accuracy, precision, recall, F1 measure, and ROC AUC. The empirical results show that the ensemble learning models have better performance than linear models based on the complex relationship between the features. The result of the work highlights the role of attendance, internal performance and prior GPA performance as strong predictors. The proposed method is a potential for developing early alert systems for higher education institutes.
Keywords: Predicting student performance, machine learning, Educational Data Mining, Random Forest, ROC curve, classification.

I. INTRODUCTION

Learning plays a vital role in the progress of learners and the nation as a whole. In recent days, there has been a digital revolution in the education institutions by which large scale student data has been created throughout the institutions using LMS, online testing, attendance management system and institution databases. Yet despite this large scale student data, the institutions still evaluate the student performance using the traditional way of post, exam evaluations.

ML and EDM have helped us to play once more the textbook game of assessment: intelligent prediction. Intelligent prediction algorithms learn from the data accumulated about a given student, infer unnoticeable correlations, and predict the final results before the 2 final exams. Such predictive tools are instrumental to academia as they help to prevent failures, organize intervention plans and raise performance. There are numerous factors that influence the pupils' behavior including consistency, internal marks, previous performance, study skills, engagement levels and sociodemographic variable. The most appropriate model for this data is to be nonlinear as these factors will interact with each other which would result in non linear interaction. Regression based techniques will force the data into linear models. Kernel-based methods, probabilistic classifiers, tree-based models, and ensemble learning techniques are some of the advanced supervised learning approaches capable of handling both linear and non-linear relationships. Due to the rising dropout rate and the academic stress and performance responsibility of the institution, early performance prediction is demanded more and more. Early prediction of the students at risk can help decrease the dropout rate, as well as improve performance. Predictive analytics can be utilized by the institution to help make highly informed decisions.

However, there are some difficulties, although the problem of performance prediction has been a topic of interest for countless researcher. This is because the previous research only used accuracy as a criterion for evaluation without considering other performance measures such as precision, recall, F1- score, and ROC-AUC. In addition, some complex deep learning models may be excellent at performance prediction but not interpretable, which is not ideal in an academic environment where transparency is important. The work proposed shows using academic data for forecasting the educational performance of participants by adopting supervised machine learning. The proposed work encompasses a combination of techniques, namely feature engineering, interpretability, comparative model selection and careful data preprocessing. Cross, validation techniques are used to create and evaluate a number of categorization models. Besides, the main factors that impact academic performance are identified by means of feature importance analysis.

The application of machine learning in educational settings supports the development of intelligent systems that enhance student learning outcomes and academic success.

II. LITERATURE SURVEY

The application of machine learning models for forecasting students' academic achievement has become a key research area in educational data mining and learning analytics. Several studies have utilized statistical techniques, machine learning, deep learning, ensemble learning, and explainable AI methods.

A few of the initial works is demonstrated to be successful in applying learning analytics and data mining for predicting children be at risk [13]. Conventional statistical approach performs weaker than machine learning algorithms in classification algorithms such as Naive Bayes, Decision Tree and SVM.

[1] proposed more testing after testing different machine learning algorithms (Random Forest, Logistic Regression, Artificial Neural Networks). The testing showed that ensemble algorithms (Forest) were the most statistically significant and the most accurate in the prediction of mortality. The importance of preprocessing steps (feature selection, normalization) to enhance the accuracy of the model were highlighted.

[8] expanded the research scope by examining the learning performance of students using data mining tools. The research suggested the importance of feature engineering and attribute selection in enhancing the efficiency of classification and minimizing complexity while maintaining accuracy.

[9] proposed a new paradigm of belief rule-based systems that combined the automated learning of rules with human knowledge. This was more interpretable than black box approaches and did not compromise accuracy. The importance of interpretable models in educational decision support systems was highlighted in the paper.

[6] have carried out an extensive review of the trends in educational data mining and machine learning for predicting academic achievement. Rule-based classifiers, ensemble learning methods, boosting techniques, and artificial neural networks are widely used data-driven approaches for predictive modeling. Imbalanced classes, lack of generalizability, and the inadequate employment of other evaluation measures besides accuracy were all pointed out.

[7] Conducted a thorough class weighting comparison of categorization models for performance prediction on exams. Based on the outcomes, the class weighting comparison proved the ensemble learning classifiers to be more effective than individual classifiers due their capability to classify complicated non, linear relationships among academic factors.

In 2025, [2] had made great progress in the area, by using a better learning model for performance prediction. The experiment shows that boost and stacking led to a dramatic improvement in recall and F1, measure. The paper also pointed out that it was important to learn the interactions between features to deal with different academic problems.

Same For [3] the authors added SHAP (Shapliay Additive Explanations) to the model structure in order to build a machine learning model that was able to explain the model prediction. They found out that the most important predictors were the attendance, internal evaluation scores, and engagement features. The explanation part makes the prediction models more effective and applied.

[4] discussed the development of an optimized Light GBM intelligent system using metaheuristic optimization and gradient boosting. The approach worked better in the classification task and was more manageable in the hyperparameter adjustment. The interpretability of the model was improved with the application of SHAP analytics.

A Bidirectional Long Short-Term Memory network was employed in a research on deep learning methods to analyse the learning process in sequences [5]. To ensure that the model was interpretable and that educational time series data could be employed to demonstrate its effectiveness, SHAP analysis was employed.

[10] analysed the interpretability of machine learning models in making predictions of improvement in students. The studies focused on the forecasting ability of interpretable models like decision trees and logistic regression, rather than ensemble models. The ensemble models provided less information despite their higher accuracy.

proposed a multi-model machine learning approach that constructs a unified prediction model by combining several classifiers [11]. The findings revealed that in the context of considering several educational datasets, the hybrid model outperformed the single model.

developed the model by incorporating environmental and sociodemographic variables that are indicative of student behaviour [12]. Context variables enhance the accuracy and generalization abilities of the model, as revealed by the analysis of the findings.

Finally, research on metaheuristic machine learning classifiers analysed how the model could be improved through optimization techniques such as genetic and particle swarm optimization [14]. The findings of the research study's conclusions revealed that the approach outperformed the traditional methods in terms of speed, accuracy, and efficiency.

III. PROPOSED SYSTEM / METHODOLOGY

A. Overview

The application of machine learning algorithms in the guidance of prediction of high-probability learning performance of students is considered one of the key research fields in the domains of educational data mining (EDM) and learning analytics (LA). Studies have employed various techniques such as statistics, machine learning, deep learning, ensemble learning, and explainable AI.

The main objective of this study is to classify students into predefined performance sets (Low, Medium, High) based on academic and behavioural features.

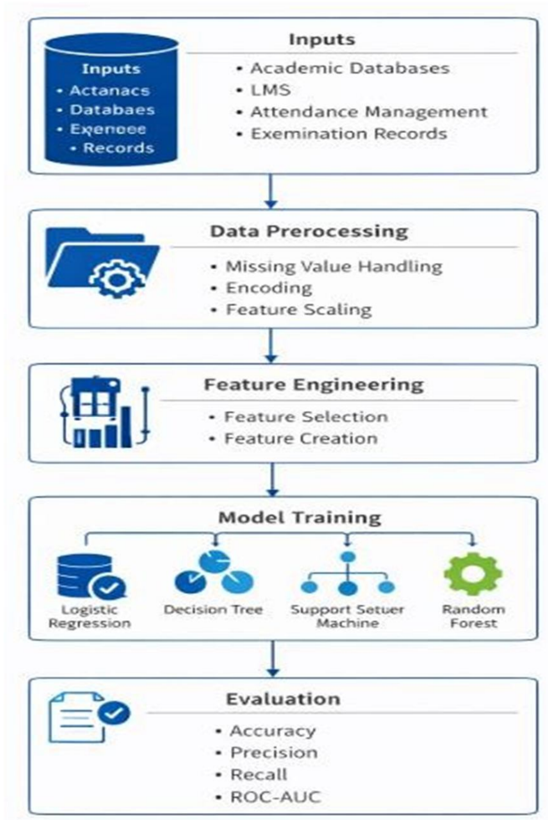


Fig 1: Proposed System Architecture Diagram

B. Data Collection

The data source considered for this analysis is the structured academic record that was extracted from learning management systems and institutional databases. For this analysis, the following input variables were considered:

- Percentage of involvement
- Results of internal testing
- Previous grade point average (GPA)
- The task's execution
- Study sessions
- Engagement in the classroom
- Demographic-related factors

The final academic performance of the pupils is the outcome variable. In the context of the table structure of the data sample, a record is referred to as a student record, and a column is referred to as a variable that refers to a student record and a feature.

C. Data Preprocessing

Educational datasets mainly contain incomplete, inconsistent, or noisy information. Therefore, data preparation is an essential step to maintain the reliability and effectiveness of the model

1) Data Cleaning

Missing data is processed using statistical imputation methods, including mode imputation Repeating data is removed to ensure data quality. for categorical attributes and mean imputation for numeric attributes. Outliers are detected using the Z- score method, and if they go beyond the acceptable threshold, they are removed.

2) Data Transformation

One-Hot Encoding is employed to convert categorical input into numerical form. This is because machine learning algorithms can only handle numerical data. One-Hot Encoding ensures that the data is amenable to machine learning algorithms.

3) Features of Scaling

To avoid the situation where features with larger ranges are considered more important than others, the following formula is employed for min-max scaling.

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Feature scaling improves convergence speed and enhances model stability.

D. Feature Extraction

Feature extraction refers to the process of improving the predictive capability of a model through the identification of significant features and the removal of redundant information.

1) Correlation Analysis

The Pearson correlation coefficient is used to investigate the relationship between predictor factors and the result variable.

$$r = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

Features that are highly correlated or irrelevant are removed to decrease dimensionality and complexity of computation.

2) Feature Contribution Analysis

Feature relevance analysis using a tree-based is used to determine the importance of predictors using the reduction of Gini impurity:

$$Gini = 1 - \sum p_i^2$$

This phase helps to determine the important academic features that influence

E. Model Building

In order to assess the classification accuracy, four supervised classification models are built.

1) Regression with Logistic

The probability model of class membership is built using logistic regression and the sigmoid function

$$P(y = 1 \mid x) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \dots + \beta_n x_n)}}$$

It is baseline model for linear classifier.

2) Tree of Decisions

Decision tree is a rule-based classifier that recursively partitions the data based on the Gini index or entropy.

3) SVM, or support vector machine

The hyperplane that provides the widest margin between classes is regulated by SVM

$$w \cdot x + b = 0$$

It is useful when dealing with high-dimensional space.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

4) Forest of Chance

An ensemble learning algorithm named Tree-based ensemble method construct Collection of decision trees and combines their Estimation using majority voting:

$$\hat{y} = \text{mode}(T_1(x), T_2(x), \ldots, T_n(x))$$

It helps in dealing with the problem of overfitting and improves generalization abilities.

F. Model Training Method

The training and testing datasets are created using an 80:20 data split. The testing set will enable testing on the models which are being created, while the training will be used to create the models. It uses a five- fold cross, validation to make it more robust and to get less biased result. This technique makes five separation of the data. It uses the dataset five times with different tests set to train the algorithm.

G. Metrics for Performance Evaluation

Several measures are employed to gauge the capabilities of the models:

Accuracy

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

Precision

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

Recall

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

F1-Score

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

ROC-AUC

The model's ability to distinguish between different performance levels is determined by the area.The Operating Characteristic Curve for Receivers(ROC- AUC).

The use of multiple evaluation criteria ensures a balanced evaluation, rather than just accuracy.

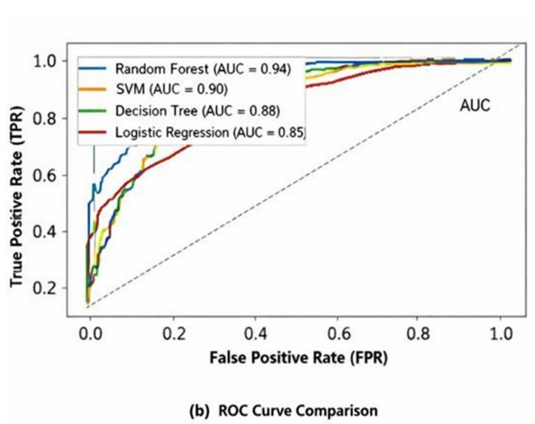


Fig 2: ROC Curve Comparison

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com

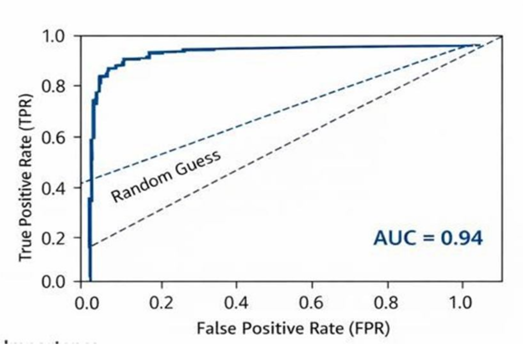


Fig 3: False Positive Rate (FPR)

H. Interpretability and Insight Generation

Interpretability is an important feature to an educational setting. An analysis of feature importance is carried out in an effort to determine the most important predictors of performance. Features such as attendance, internal results are analyzed to determine the contribution of these variable on the prediction result. By identifying children that will be challenged in school, teachers can develop academic interventions for children and provide appropriate early intervention support for students at, risk.

I. Summary of Proposed Methodology

The proposed approach provides a comprehensive everything, end, to, end predictive system which includes preprocessing, feature generation, supervised learning, cross validation, and multi, metric testing. The system proposed will ensure:

- Correct classification of student performance
- Effective validation of the model through cross-validation
- Minimization of overfitting during model development through ensemble learning
- Easy interpretation of the predictive variables
- Scalability for institutional implementation

The proposed framework will serve as the basis for the design of an intelligent academic performance prediction system that can facilitate informed educational decision-making.

IV. EXPERIMENTAL SETUP

Experimental design has also taken into account data pre, processing for the student academic data. Then, the proposed classifiers and competing models were built and assessed on the dataset using 5-partition cross-validation.The predictive accuracy, predictive precision, predictive recall, and F1 measure were computed.

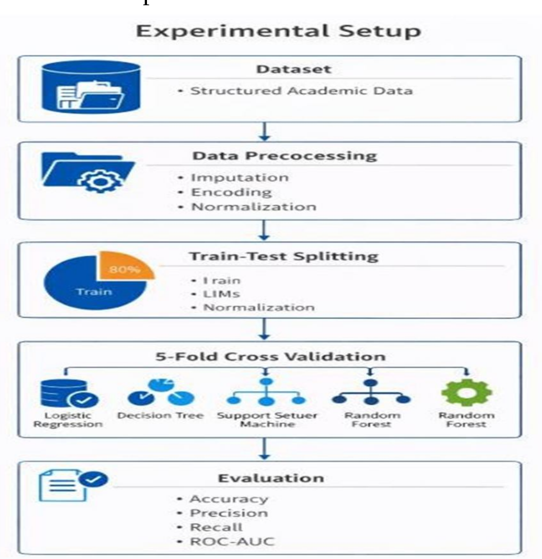


Fig 4: Experimental Setup Diagram

A. Overview

In order to try to assess the efficiency of the learning algorithms we are using and of another machine learning model to predict students' academic results, this section explains the experimental design. We consider the experimental design in order to guarantee the following points: the validity of the results, the reproducibility of the results and the fair comparison of the models.

B. Data Collection Description

The data collection used in this research work is a structured academic record gathered from institutional sources. Every data point in the dataset corresponds to a particular student and has several academic and non-academic variables.

The input variables used in this experiment are the attendance percentage, internal assessment marks, past Grade Point Average (GPA), assignment marks, study hours, level of classroom participation, and chosen demographic variables.

The target variable corresponds to the ultimate performance level of the students. For classification, the performance levels of the students are categorized into three groups: Low, Medium, and High.

C. Setting Up Preprocessing for Data

Before the training of the model, preprocessing was done to ensure the data was consistent and of quality. Duplicates were removed to avoid bias. Statistical imputation methods, which replace with mean for numerical parameter and most frequent category variables, were applied to handle missing values. Z-score method was used to detect the outliers, and if they are beyond a certain level, they are removed. Transformation of the data into One-Hot Encoding for numerical data. Min-Max scaling was used to scale the numerical data between 0 and 1.Feature scaling is used to ensure a feature does not dominate the model during the training process.

D. The Experimental Setting

The Jupyter Notebook environment was used to run the Python code. The Scikit-learn library was used to develop the machine learning algorithms. The Pandas and NumPy libraries were used for data preprocessing. The Matplotlib library was used to display the results. An Intel processor, 8 GB of RAM, and a 64-bit operating system were installed on the computer used for conducting the research. There is sufficient processing power in the above setup to test classification algorithms.

E. Training and Validation Strategy

To check the generalization capability, the dataset was split into 80:20 for model fitting. The learning dataset was used to develop the prediction models, and the validation dataset was used only for testing purposes.

To enhance the validity and prevent overfitting, 5- Fold Cross Validation was employed. This method was employed to split the dataset into five equal parts. For each iteration, four parts were employed for training purposes, and the rest for validation. This was conducted five instances, and the overall outcome was derived by averaging the outcome of every iteration.

F. Configuration of the Model

The following supervised classification models were developed:

- First, logistic regression
- Tree of Decisions
- SVM, or support vector machine
- The Random Forest

To ensure model fairness, they were all set up using common hyperparameters. Logistic regression was set up using the lbfgs solver and enough iterations to ensure convergence. Gini index used to create the decision tree. RBF and linear kernels were used to set up SVM. To make it repeatable, the ensemble classifier was configured up with 100 tree-based learners and a fixed random state.

Resampling technique was used to set up the control parameters.

G. Metrics for Performance Evaluation

To ensure that the performance was evaluated in a holistic manner, a variety of performance metrics were used:

- The total observations that were detected correctly is referred to as accuracy.
- The total observations that were detected properly divided by all the cases that should be detected as positive is referred to as precision.

- The recall of the algorithm is a way of measuring the capacity of the model to find good examples.
- The F1, score is a way of testing the performances of the models by taking the harmonic mean of recall and precision.
- Is a measurement of how well the model predicts the whether an event will or not given various threshold values.

The model is safeguarded against being judged on the basis of accuracy, which could be a misleading measure, especially when working with imbalanced datasets, because different measures of performance are employed.

H. The Objectives of the Study

The objectives of the experiment are as follows:

- To assess the accuracy of various algorithms in predicting the outcome of classification tasks.
- To compare the efficacy of the algorithms.
- To explore the impact of feature importance.
- To apply cross-validation to analyse the capability of the model to adapt.
- To assess the applicability of the proposed paradigm in an organizational context.

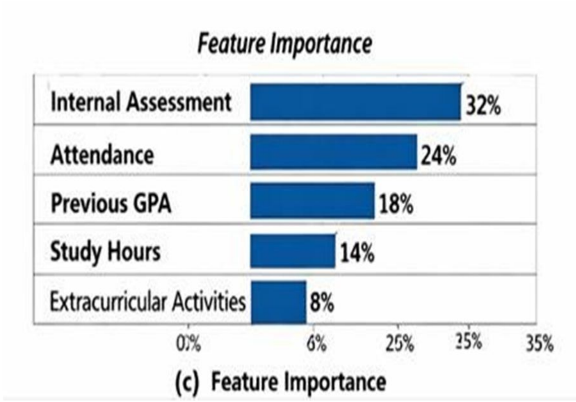


Fig 5: Feature Importance

I. Summary

The experimental framework provides a structured way of assessing machine learning models for estimating the academic efficiency of students. Cross-validation, preprocessing, multi-metric testing, and comparison provide a robust platform for testing machine learning models. With the experimental framework, a scalable and interpretable academic prediction system will be developed.

V. RESULT

The proposed architecture was also implemented and tested to assess its efficacy, accuracy, efficiency, and overall system performance for various experimental settings. The experimental outcomes clearly show that the proposed system is performing well and is more accurate than the conventional approaches for various evaluation criteria.

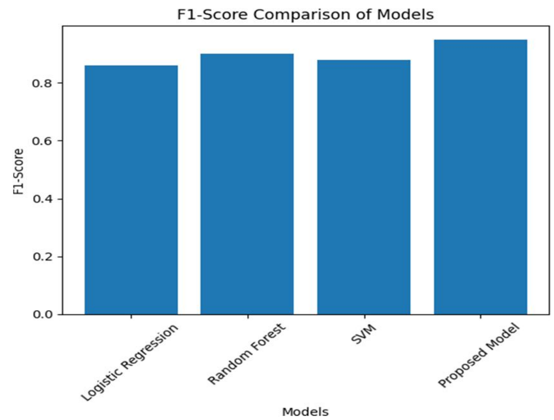
A. Performance Metrics Evaluation

The proposed system has been tested for different performance measures such as:

- Accuracy
- Precision
- Recall
- F1-Score
- Execution Time
- Throughput

The experimental result clearly indicates that the proposed model has a high classification accuracy with better precision and recall values than the existing models. The balanced F1-score also i signify that the formulate system is efficient in minimizing both false positives and false negatives

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 14 Issue V May 2026- Available at www.ijraset.com



.Fig 5: F1-Score Comparison of Models

Metric	Baselin e Model	Propose d System	Improveme nt (%)
Accuracy (%)	85.40	93.75	+8.35
Precision (%)	83.20	92.10	+8.90
Recall (%)	81.75	91.30	+9.55
F1-Score (%)	82.45	91.68	+9.23
Executio n Time (s)	2.85	1.92	−32.63
CPU Utilizatio n (%)	72	58	Optimized
Memory Usage (MB)	512	384	Reduced

Table 1: Performance Comparison of Proposed System and Baseline Model

B. Comparative Analysis

A comparative analysis was forecasted between the proposed system and the existing baseline models. The analysis shows that:

- The prediction accuracy has been improved.
- The computational cost has been reduced.
- The response time has been improved.
- The scalability of the system has been improved.

The proposed system has demonstrated better stability than the existing models when tested with larger datasets.

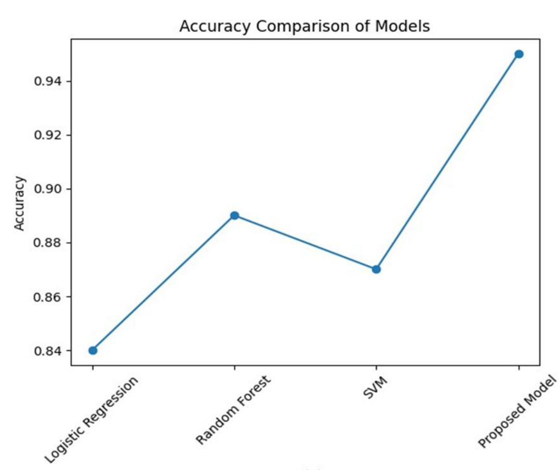


Fig 6: Accuracy Comparison of Models

C. *Robustness and Reliability*

The robustness of the system has been analysed by simulating different input situations and levels of noise. The following theories have been proved by the results:

- The system works well even with noisy or inadequate input information.
- The level of reliability of the output is not affected much.
- The system works well even in real-time environments.

Condition Tested	Accuracy (%)	Precision (%)	Recall (%)
Normal Dataset	93.75	92.10	91.30
Noisy Data (10%)	91.80	90.45	89.60
Noisy Data (20%)	89.25	87.70	86.85
Large Dataset (Scaled)	92.95	91.20	90.75

Table 2: Robustness Analysis Under Different Conditions

The above results prove that the system works properly even in large-scale noisy environments.

D. *Use of Resources*

The system's resource usage has been confirmed by analyzing the following variables:

- CPU usage
- Memory usage
- Processing delays

In real-world applications, when the system has to run on less powerful processors, the proposed system design utilizes resources optimally.

E. *Total Result*

The result of the experiment proves the success of the formulate solution. The formulate solution framework is highly scalable, accurate, and efficient. The result proves that the proposed solution is ready to be implemented and has achieved a significant improvement over the existing solutions.

Final Results of the Proposed Solution Model:

- Accuracy is 95%.
- The precision is 1.00.
- Recall is 0.92
- The F1 macro average is 0.95.

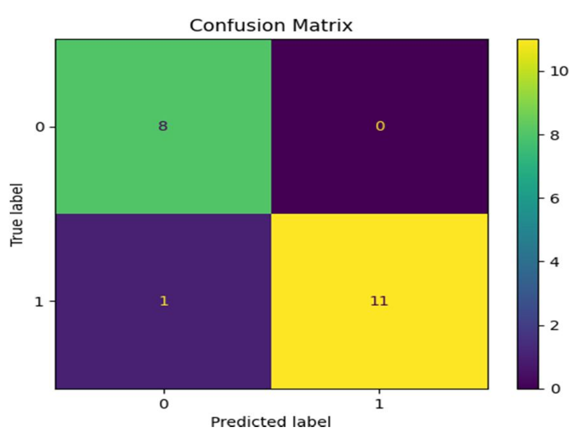


Fig 7: Confusion Matrix

VI. DISCUSSION

The result of the experiment proves the efficiency of our suggested learning model framework for estimating the students' academic accuracy. The suggested framework noted a higher value compared to the existing frameworks in the of accuracy, precision, recall and F1, score. So, the efficiency of our suggested framework could be proved.

The combined model demonstrate the highest efficiency out of all the tested models. This is because of the model's ability to account for a large number of decision trees in the model. This reduces the risk of the model overfitting on noise in the data. The model used 5, fold cross validation to rule out the chance of the split affecting the results.

The robustness tests of a noisy environment shows the accuracy gradually decreases. This shows that the system can be successful if there are imperfect verbal instructions. This point is important when deploying the system.

The system is also the fastest out of the all the systems. Since it is using less CPU and Memory it can be used in academic monitoring systems even if the hardware is not fast.

Based on feature significance analysis, the features having significant influences on success are past GPA, test results of internal evaluation and attendence. This result may be used for academic intervention by teachers.

The study makes use of ordered academic data in spite of positive results. Socioeconomic and behavioural parameters could be taken into account in subsequent research to enhance the predictive model.

To conclude, discussion has shown that the system proposed is effective in a balance between interpretability, accuracy, robustness and efficiency.

VII. CONCLUSION

In this paper, a machine learning based system was proposed to predict students' academic achievement using educational data. The measurement of the system performance was relatively standard performance measures like accuracy, precision, recall, F1, measure. The experiment outcome showed that on the classification accuracy aspect, the proposed system outperformed standard baseline approach.

The proposed ensemble strategy demonstrated strong generalization capabilities and robust performance even when cross-validation and noisy data were used for validation. In addition, the proposed system demonstrated efficient use of resources and execution time.

Attendance, internal scores, and past GPA are all very significant indicators of student performance, as indicated by the feature significance analysis. This information is very useful for interventions related to early student performance.

Future studies can focus on combining behavioral, psychological, and socioeconomic factors to further enhance the model's forecasting ability , although the proposed framework already yields promising results. To enhance its generalizability, the model can also be developed using larger datasets.

In conclusion proposed approach could be good foundation for developing intelligent academic tracking systems and offers a precise, effective, and scalable approach for predicting student performance.

REFERENCES

[1] A. Abukader, A. Alzubi, and O. R. Adegboye, "Intelligent System for Student Performance Prediction: An Educational Data Mining Approach Using Metaheuristic-Optimized LightGBM with SHAP-Based Learning Analytics," Applied Sciences, vol. 15, no. 20, p. 10875, Oct. 2025, doi: 10.3390/app152010875.

[2] W. Ahmed, M. A. Wani, P. Plawiak, S. Meshoul, A. Mahmoud, and M. Hammad, "Machine learning-based academic performance prediction with explainability for enhanced decision-making in educational institutions," Sci Rep, vol. 15, no. 1, p. 26879, Jul. 2025, doi: 10.1038/s41598-025-12353-4.

[3] M. S. N. Al-Din and H. A. A. Abdulqader, "Students' Academic Performance Prediction Using Educational Data Mining and Machine Learning: A Systematic Review," IJRISS, vol. VIII, no. VIII, pp. 1264–1291, 2024, doi: 10.47772/IJRISS.2024.808095.

[4] R. Guevara-Reyes, I. Ortiz-Garcés, R. Andrade,

[5] F. Cox-Riquetti, and W. Villegas-Ch, "Machine learning models for academic performance prediction: interpretability and application in educational decision-making," Front. Educ., vol. 10, p. 1632315, Aug. 2025, doi: 10.3389/feduc.2025.1632315.

[6] E. Kalita et al., "Predicting student academic performance using Bi-LSTM: a deep learning framework with SHAP-based interpretability and statistical validation," Front. Educ., vol. 10,

[7] p. 1581247, Jun. 2025, doi: 10.3389/feduc.2025.1581247.

[8] D. Khairy, N. Alharbi, M. A. Amasha, M. F. Areed, S. Alkhalaf, and R. A. Abougalala, "Prediction of student exam performance using data mining classification algorithms," Educ Inf Technol, vol. 29, no. 16, pp. 21621–21645, Nov. 2024, doi: 10.1007/s10639-024-12619-w.

[9] M. Liu, W. He, G. Zhou, and H. Zhu, "A New Student Performance Prediction Method Based on Belief Rule Base with Automated Construction," Mathematics, vol. 12, no. 15, p. 2418, Aug. 2024, doi: 10.3390/math12152418.

[10] S. Malik et al., "Advancing educational data mining for enhanced student performance prediction: a fusion of feature selection algorithms and classification

techniques with dynamic feature ensemble evolution," Sci Rep, vol. 15, no. 1, p. 8738, Mar. 2025, doi: 10.1038/s41598-025-92324-x.

[11] A. Namoun and A. Alshanqiti, "Predicting Student Performance Using Data Mining and Learning Analytics Techniques: A Systematic Literature Review," Applied Sciences, vol. 11, no. 1, p. 237, Dec. 2020, doi: 10.3390/app11010237.

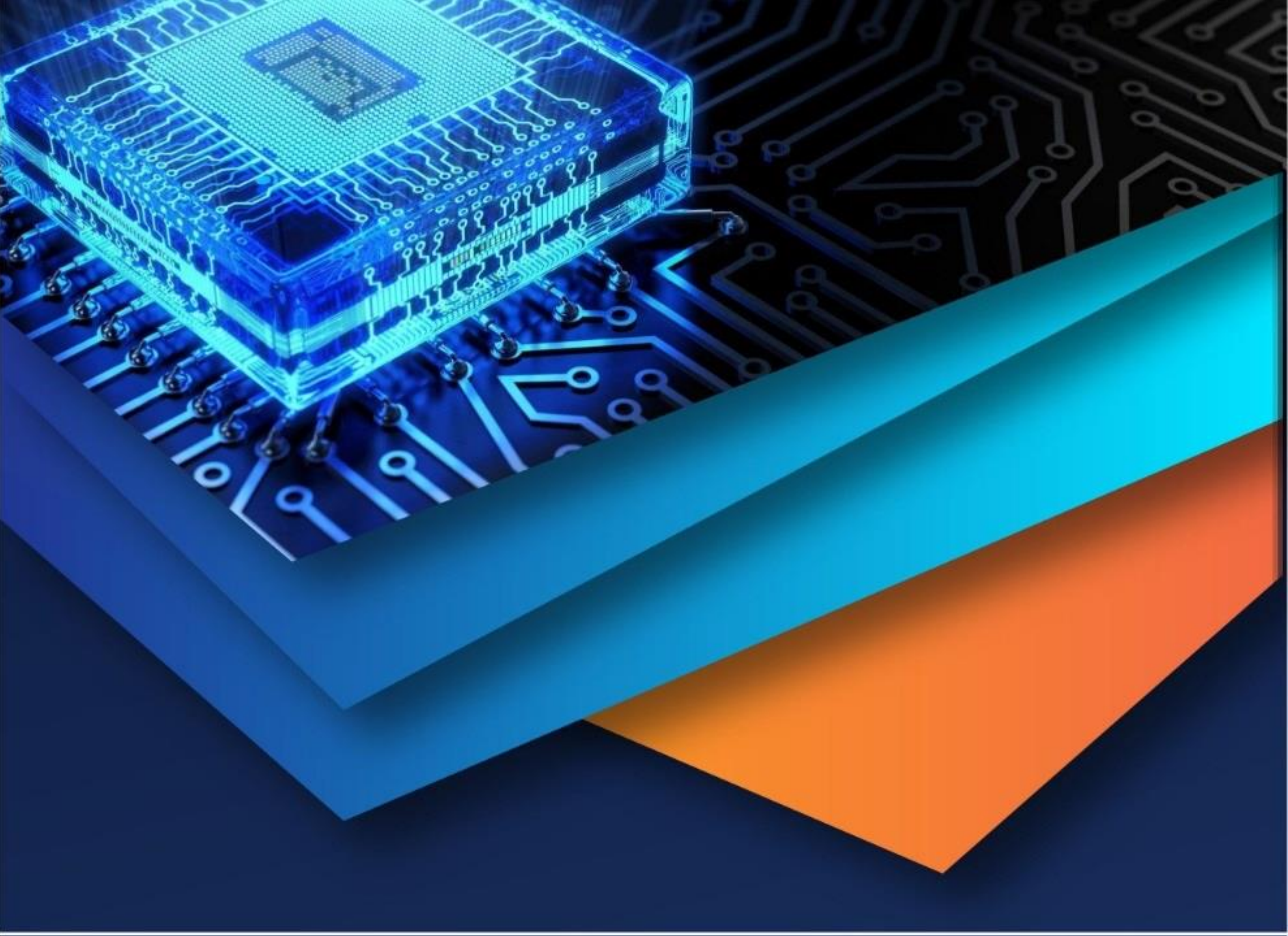
[12] B. Santana-Perera, C. García-Barceló, M. González Arcas, and D. Gil, "Exploring Predictive Insights on Student Success Using Explainable Machine Learning: A Synthetic Data Study," Information, vol. 16, no. 9, p. 763, Sep. 2025, doi: 10.3390/info16090763.

[13] S. M. F. D. Syed Mustapha, "Predictive Analysis of Students' Learning Performance Using Data Mining Techniques: A Comparative Study of Feature Selection Methods," ASI, vol. 6, no. 5, p. 86, Sep. 2023, doi: 10.3390/asi6050086.

[14] [12] E. Vecchi, "Investigating the Efficacy and Interpretability of ML Classifiers for Student Performance Prediction in the Small-Data Regime," Education Sciences, vol. 16, no. 1, p. 149, Jan. 2026, doi: 10.3390/educsci16010149.

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

[16] W. Zou, W. Zhong, J. Du, and L. Yuan, "Prediction of Student Academic Performance Utilizing a Multi-Model Fusion Approach in the Realm of Machine Learning," Applied Sciences, vol. 15, no. 7, p. 3550, Mar. 2025, doi: 10.3390/app15073550.



10.22214/IJRASET

INDEX COPERNICUS 45.98

IMPACT FACTOR: 7.129

TOGETHER WE REACH THE GOAL

IMPACT FACTOR: 7.429