IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

Student Performance Prediction Using Machine Learning

Hrutuja Umakant Aher
M.Sc. Computer Science
K.R.T Arts, A.M. Science & B.H. Commerce College, Nashik
hrutujaaher034@gmail.com

Abstract: *Student performance prediction is an important application of machine learning in the field of education. This study proposes a predictive system to analyze and forecast student academic outcomes using features such as previous marks, study time, and attendance. A dataset of 200 student records with realistic variations was used for training and testing the models.*

Machine learning algorithms including Random Forest and Decision Tree were applied to classify students into pass and fail categories. The data was preprocessed, relevant features were selected, and models were trained and evaluated using accuracy as the performance metric. The Random Forest model achieved an accuracy of 88.33%, while the Decision Tree model achieved 81.66%.

The results demonstrate that machine learning techniques can effectively identify at-risk students and support data-driven academic decision-making. This system can assist educators in providing early intervention, improving student performance, and enhancing overall educational outcomes. The proposed approach highlights the importance of predictive analytics in modern education systems and its potential to transform traditional evaluation methods..

Keywords: Machine Learning, Student Performance Prediction, Educational Data Mining, Random Forest, Decision Tree, Classification

I. INTRODUCTION

Education plays a vital role in shaping a student's future and overall development. However, many students face academic difficulties due to various factors such as poor study habits, lack of guidance, and irregular attendance. Traditional methods of evaluating student performance mainly rely on examinations and teacher observations, which are often unable to identify weak students at an early stage. As a result, timely intervention becomes difficult, leading to poor academic outcomes.

With the advancement of technology, educational institutions generate a large amount of student data, including academic records, attendance, and behavioral information. However, this data is not effectively utilized for predicting student performance. Machine Learning (ML) provides an efficient solution by analyzing such data and identifying patterns that can help predict future academic outcomes.

This research focuses on developing a machine learning-based system to predict student performance using features such as previous marks, study time, and attendance. By applying algorithms like Random Forest and Decision Tree, the system can identify students who are at risk and help educators take early corrective actions. This approach supports data-driven decision-making and aims to improve overall academic performance.

II. LITERATURE REVIEW

Student performance prediction has become an important research area in the field of Educational Data Mining (EDM). With the increasing availability of student data, researchers have applied various machine learning techniques to improve the accuracy of predicting academic outcomes. These studies help educators identify academically weak students at an early stage and take necessary actions to improve their performance.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429 Volume 6, Issue 3, May 2026 Impact Factor: 8.2

Several researchers have explored different machine learning algorithms such as Decision Tree, Naïve Bayes, Support Vector Machine, and K-Nearest Neighbors for predicting student performance. Among these, ensemble methods like Random Forest and Gradient Boosting have shown better accuracy due to their ability to handle complex data and reduce overfitting.

Recent studies highlight the importance of using multiple factors such as attendance, internal marks, study habits, and behavioral patterns. Some research also includes data from Learning Management Systems (LMS), such as assignment submissions and quiz performance, which further improves prediction accuracy.

Overall, the literature suggests that machine learning-based prediction systems are effective tools for improving academic decision-making. They enable early identification of at-risk students, reduce dropout rates, and support personalized learning.

III. METHODOLOGY

This research proposes a machine learning-based approach to predict student performance using academic and behavioral data. The methodology consists of data collection, preprocessing, feature selection, model training, and evaluation.

A. Dataset

A dataset of 200 student records was used in this study. The dataset includes important features such as:

- G1: First internal marks
- G2: Second internal marks
- Study Time
- Absences
- G3: Final marks

These features were selected because they have a strong influence on student performance.

B. Data Preprocessing

The dataset was preprocessed to ensure accuracy and consistency. This includes:

- Removing inconsistencies in data
- Converting values into suitable format
- Creating a target variable (Pass/Fail) based on final marks

Students scoring 10 or above in G3 were considered as Pass, and below 10 as Fail.

C. Feature Selection

The most relevant features were selected for prediction:

- Academic performance (G1, G2)
- Study behavior (study time)
- Attendance (absences)

These features help improve model accuracy.

D. Machine Learning Models

Two machine learning algorithms were used:

- Decision Tree • Random Forest

Decision Tree is simple and easy to interpret, while Random Forest is an ensemble method that improves accuracy by combining multiple trees.

E. Model Training and Testing

The dataset was divided into:

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

• 70% training data
• 30% testing data

The models were trained on the training data and evaluated on the testing data using accuracy as the performance metric.

IV. RESULTS

The performance of the proposed machine learning models was evaluated using the accuracy metric. The models were trained on 70% of the dataset and tested on the remaining 30%.

The experimental results show that the Random Forest algorithm achieved an accuracy of 88.33%, while the Decision Tree algorithm achieved an accuracy of 81.66%.

These results indicate that the Random Forest model performs better in predicting student performance due to its ensemble nature, which reduces overfitting and improves generalization. On the other hand, the Decision Tree model, although simple and easy to interpret, shows comparatively lower accuracy.

The comparison of both models is illustrated in Figure 1, which shows the accuracy of each algorithm.

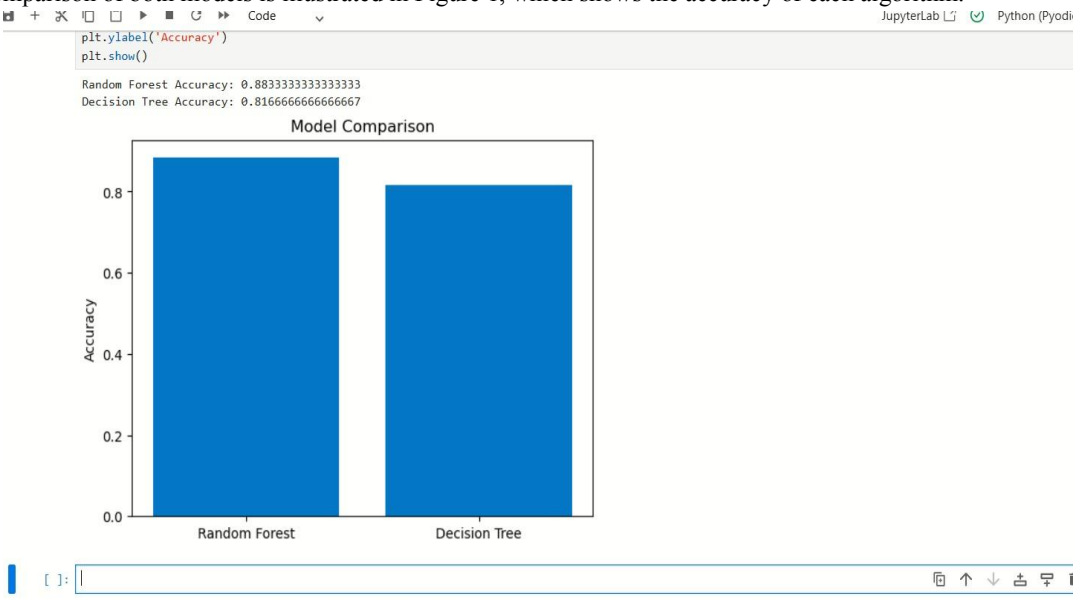

Figure 1: Model Accuracy Comparison

V. DISCUSSION

The results obtained from the experiments clearly demonstrate that machine learning techniques can effectively predict student performance. Among the selected features, previous academic scores (G1 and G2) were found to have the most significant impact on prediction of accuracy. Study time and attendance also contributed to the model but to a lesser extent.

The Random Forest algorithm outperformed the Decision Tree model due to its ensemble approach, which combines multiple decision trees and reduces the chances of overfitting. This allows the model to generalize better on unseen data. In contrast, the Decision Tree model is more prone to overfitting, especially when dealing with smaller or less complex datasets.

The findings of this study highlight the importance of using appropriate machine learning techniques and relevant features for accurate prediction. The results also support the idea that predictive analytics can play a key role in improving educational outcomes.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

VI. CONCLUSION

This research presents a machine learning-based system for predicting student performance using academic and behavioral data. The study demonstrates that machine learning models, particularly Random Forest, can achieve high accuracy in predicting whether a student will pass or fail.

The system helps in identifying at-risk students at an early stage, allowing educators to take timely corrective actions. This approach improves academic planning, enhances student performance, and supports data-driven decision-making in educational institutions.

Overall, the research highlights the potential of machine learning in transforming traditional education systems into smarter and more efficient systems.

VII. FUTURE SCOPE

The proposed system can be further enhanced in several ways to improve its performance and real-world applicability. Firstly, the model can be integrated into a web-based or mobile application to provide real-time prediction of student performance. This would allow teachers and administrators to easily access predictions and take timely actions.

Secondly, the system can relate to Learning Management Systems (LMS) to collect real-time data such as assignment submissions, quiz results, and student engagement. This will help improve prediction accuracy by using more detailed and dynamic data.

In addition, advanced machine learning techniques such as deep learning and artificial intelligence can be applied to handle larger and more complex datasets. Incorporating features like student behavior, psychological factors, and extracurricular activities can also provide deeper insights into student performance.

Finally, the system can be implemented in real educational institutions to evaluate its effectiveness in improving academic outcomes and reducing dropout rates.

VIII. LIMITATIONS

Despite the effectiveness of the proposed system, there are certain limitations in this study. The dataset used in this research is limited in size and is artificially generated, which may not fully represent real-world student behavior. As a result, the model's performance may vary when applied to actual institutional data.

Another limitation is the restricted number of features used for prediction. Important factors such as psychological aspects, learning styles, and external influences are not included, which may affect prediction of accuracy.

Additionally, the system has not been deployed in a real-time environment, and therefore its practical usability has not been tested. Issues such as data privacy, security, and ethical concerns are also not addressed in this study.

Lastly, the study focuses only on basic machine learning models and does not include a detailed comparison with advanced algorithms, which could further improve performance.

REFERENCES

[1] A. John and V. Vijendra, "An analysis of machine learning algorithms for predicting student performance," International Journal of Advanced Research in Science, Communication and Technology (IJARSCT), vol. 4, no. 2, pp. 657–660, 2024.

[2] A. Rodrigues and P. Silva, "Deep learning-based student GPA prediction: A longitudinal study," Journal of Educational Data Science, vol. 5, no. 1, pp. 55–70, 2024.

[3] S. Ventura and E. García, "Recent advances in educational data mining and learning analytics," Journal of Learning Analytics, vol. 10, no. 2, pp. 41–62, 2023.

[4] L. Poovarasi and N. Ramya, "Student performance prediction using data mining algorithms," International Journal of Advanced Research in Computer and Communication Engineering (IJARCCE), vol. 11, no. 6, pp. 267–272, 2022.

[5] R. Duche, R. Tipnis, and S. Singh, "Academic performance prediction of engineering students," Current Opinion in Reviews and Reports (COJRR), vol. 3, no. 3, 2021.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 3, May 2026
Impact Factor: 8.2

[6] P. Cortez and A. Silva, "Using data mining to predict secondary school student performance," in Proceedings of the 5th Future Business Technology Conference, 2008.
[7] T. Hastie, R. Toshigami, and J. Friedman, The Elements of Statistical Learning, Springer, 2009.