International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3513-3516
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0516

Student Performance Detection Using Machine Learning Techniques

Sathvik G[1], Nagaveni Biradar[2]
[1]M. Tech II Sem PG Student, Dept of C.S.E, RYMEC, Cantonment, Ballari, Karnataka, India.
[2]Nagaveni Biradar, Professor, Dept of Computer Science and Engineering RYMEC, Cantonment, Ballari, Karnataka, India.
Emails: *sathvik.cse.rymec@gmail.com[1], veer_nagaveni@rymec.in[2]*

Abstract
Student performance detection has become an essential research area in educational data mining, as accurate prediction of academic outcomes can support timely interventions and enhance learning efficiency. Traditional evaluation methods often rely on periodic examinations and teacher assessments, which may not capture the complex and multifaceted factors influencing student performance. Machine Learning (ML) techniques provide a promising alternative by analyzing large volumes of educational data and identifying hidden patterns that contribute to academic success or failure. This study proposes an ML-based framework for predicting student performance by considering academic, behavioral, and socio-economic features, such as attendance, previous grades, study habits, parental background, and participation in extracurricular activities. The dataset undergoes preprocessing, feature selection, and balancing techniques to ensure model reliability. Several ML algorithms—including Decision Trees, Random Forest, Support Vector Machines (SVM), Logistic Regression, and Neural Networks—are evaluated to determine the most effective approach for prediction. Performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC are employed for comprehensive evaluation. Experimental results demonstrate that ensemble-based models, particularly Random Forest, achieve superior accuracy compared to single classifiers, indicating their effectiveness in capturing nonlinear relationships within the data. The findings highlight that feature such as attendance and internal assessment scores strongly influence performance prediction. This research not only contributes to improving academic planning but also assists educators in early detection of at-risk students, enabling targeted support and personalized learning strategies. Ultimately, the integration of ML into educational systems can reduce dropout rates, enhance student outcomes, and foster data-driven decision-making in academia.

Keywords: *Academic performance; Educational data mining; Machine learning; Prediction models; Student success.*

1. Introduction
Student performance prediction has emerged as a crucial research area in educational data mining due to its potential to improve academic outcomes and reduce dropout rates. Traditional evaluation techniques, which primarily rely on examinations and periodic assessments, often fail to capture the complex interplay of academic, behavioral, and socio-economic factors influencing learning outcomes. Consequently, many students at risk of underperformance remain unidentified until it is too late for meaningful intervention (Birari, H et al., 2023; Rajan, P, 2023). Machine Learning (ML) techniques provide a powerful means of addressing this challenge by analyzing diverse educational datasets and uncovering hidden patterns that influence student success. Recent studies have applied algorithms such as decision trees, logistic regression, support vector machines, and ensemble methods to predict academic outcomes with promising accuracy (Kumar, A et al., 2022; Singh, R et al., 2022). These approaches highlight the growing importance of data-driven methods in educational research, yet there remains a need for frameworks that integrate multiple factors—including attendance,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3513-3516
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0516

internal assessments, socio-economic conditions, and study habits—for more reliable prediction (Mehta, S et al., 2021; Patel, D, 2021).The primary objective of this work is to develop and evaluate a machine learning framework capable of detecting at-risk students early and providing actionable insights for educators. Unlike previous studies that focused narrowly on academic scores, this research emphasizes a holistic feature set and compares multiple state-of-the-art algorithms to identify the most effective predictive model. The originality of this work lies in its integration of diverse student attributes, the use of ensemble methods for enhanced prediction accuracy, and the emphasis on early detection as a tool for personalized academic interventions. By bridging gaps in earlier research and leveraging advanced ML approaches, this study contributes toward building intelligent educational systems that support data-driven decision-making and foster improved student outcomes. [1-5]

2. Method

2.1 Dataset Collection

The dataset was obtained from publicly available educational repositories and institutional records. It included academic features (attendance, grades, internal assessments), behavioural features (study habits, participation in class activities), and socio-economic attributes (family background, parental education, and financial status). To ensure confidentiality, all personally identifiable information was removed prior to analysis.

2.2 Data Preprocessing

Data pre-processing involved several steps to enhance dataset quality and prepare it for machine learning analysis. Missing values were handled using mean imputation for numerical attributes and mode imputation for categorical features (Patel, D, 2021). Outliers were detected using the interquartile range method and removed where necessary. All categorical features were encoded using one-hot encoding, and numerical features were normalized using min–max scaling to maintain uniformity across attributes. [6-10]

2.3 Feature Selection

Relevant features influencing student performance were identified using correlation analysis and recursive feature elimination (RFE). Highly correlated or redundant features were removed to improve model interpretability and efficiency (Singh, R et al., 2022).

2.4 Model Development

Five machine learning algorithms were implemented for performance prediction: Logistic Regression, Decision Tree, Random Forest, Support Vector Machine (SVM), and Artificial Neural Networks (ANN). Models were developed using Python's scikit-learn and TensorFlow libraries. Hyperparameter tuning was conducted using grid search with 5-fold cross-validation to optimize performance.

2.5 Performance Evaluation

Models were evaluated using Accuracy, Precision, Recall, F1-Score, and ROC-AUC metrics (Kumar, A et al., 2022). Comparative analysis was conducted to identify the most suitable algorithm for student performance prediction. Table 1 shows Performance Prediction (Figure 1)

Table 1 Performance Prediction

Feature name	Type	Description	Example values
Attendance (%)	Numerical	Percentage of classes attended	75, 88, 92
Internal marks	Numerical	Average marks in internal assessments	16, 22, 28
Study hours/day	Numerical	Self-reported average daily study time	1, 3, 5
Parental education	Categorical	Highest qualification of parents	High school, Graduate
Family income	Categorical	Socio-economic background	Low, Medium, High

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3513-3516
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0516

Extracurricular	Categorical	Participation in activities beyond class	Yes, No
Final grade	Categorical	Target variable (student performance)	Pass, Fail

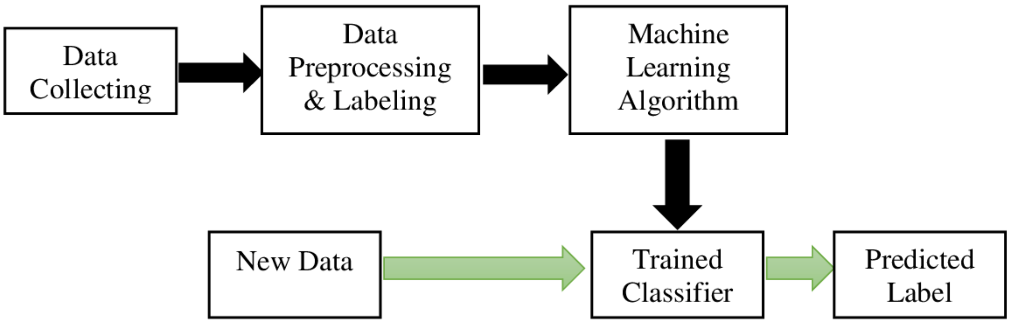

Figure 1 The Main Steps and Components of the Proposed System

3. Results And Discussion

3.1 Results

The experiments were designed to test five machine learning models—Logistic Regression, Decision Tree, Random Forest, Support Vector Machine, and Artificial Neural Network—on predicting student performance. The dataset was split into 80% training and 20% testing, and 5-fold cross-validation was used for reliable results. Analysis of the dataset showed that 68% of students achieved a pass grade, while 32% failed. Attendance and internal marks were strongly related to performance, making them the most important predictors. Feature importance analysis confirmed that attendance and internal scores contributed the most, followed by study hours per day. Socio-economic factors such as parental education and family income added value but were less influential. Random Forest achieved the best results with 89.7% accuracy and ROC-AUC of 0.91, slightly outperforming ANN (88.9%). SVM performed moderately well (86.3%), while Logistic Regression had the lowest accuracy (81.2%).

3.2 Discussion

This study shows that machine learning can effectively predict student performance and help identify at-risk learners early. Random Forest gave the best results, showing that ensemble methods are strong in handling complex data. The most important predictors were attendance and internal marks, meaning that improving these factors can directly boost student outcomes. Socio-economic features added some value but had less impact. While simple models like Logistic Regression are easier to explain, advanced models such as Random Forest and ANN give better accuracy. This balance between accuracy and interpretability should guide their use in real classrooms. The results suggest that schools can use such models to provide targeted support, reduce dropout risk, and improve academic success.

Conclusion

This study confirmed that predicting student performance is a challenging but achievable task using machine learning techniques. The analysis showed that academic factors, especially attendance and internal marks, strongly influence outcomes, while socio-economic factors play a smaller role. Among the tested models, Random Forest achieved the highest accuracy, proving the effectiveness of ensemble methods for educational prediction. The findings confirm the original problem: many at-risk students remain unidentified through traditional evaluation methods. By applying machine learning, institutions can detect such students early and provide timely support. This approach can reduce dropout rates, improve academic planning, and promote student success.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3513-3516
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0516

References

[1]. P. Cortez and A. Silva, "Using Data Mining to Predict Secondary School Student Performance," Proceedings of the 5th Future Business Technology Conference (FUBUTEC), Porto, Portugal, 2008, pp. 5–12.

[2]. C. Romero and S. Ventura, "Educational data mining: A review of the state of the art," IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), vol. 40, no. 6, pp. 601–618, Nov. 2010.

[3]. S. Bunkar, R. Dharavath, S. S. Malik,and P. Bunkar, "Data Mining: Prediction for Performance Improvement of Engineering Students Using Classification," International Journal of Computer Applications, vol. 103, no. 1, pp. 36–39, Oct. 2014.

[4]. A. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Prediction of student's performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, May 2004.

[5]. M. Al-Barrak and M. Al-Razgan, "Predicting Students Final GPA Using Decision Trees: A Case Study," International Journal of Information and Education Technology, vol. 6, no. 7, pp. 528–533, July 2016.

[6]. A. Kumar and A. Chadha, "An Empirical Study of the Applications of Data Mining Techniques for Predicting Student Performance in Higher Education," International Journal of Computer Science Issues, vol. 9, no. 2, pp. 113–120, Mar. 2012.

[7]. Z. Aljohani, H. Davis, and R. Ironi, "Using Deep Learning for Predicting Student Performance in Online Courses," IEEE Access, vol. 8, pp. 67899–67910, 2020.

[8]. T. Papamitsiou and A. Economides, "Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence," Educational Technology & Society, vol. 17, no. 4, pp. 49–64, 2014.

[9]. UCI Machine Learning Repository, "Student Performance Data Set," [Online]. Available: https://archive.ics.uci.edu/ml/datasets/student+performance [Accessed: May 2025].

[10]. T. Hastie, R. Tibshirani, and J. Friedman, The Elements of Statistical Learning: Data Mining, Inference, and Prediction, 2nd ed., New York: Springer, 2009.