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 5, Issue 5, November 2025
Impact Factor: 7.67

Student Performance Prediction using Machine Learning

Atif Khan and Prof. Leena Patil
Godavari Institute of Management and Research, Jalgoan

Abstract: *Student performance prediction is an important domain within Educational Data Mining (EDM). The ability to predict a student's academic performance before final examinations allows institutions to provide timely support and targeted interventions. In this research, academic and behavioral parameters such as attendance, assignment completion, previous marks, parental education, and classroom participation are analyzed to build a machine learning–based prediction model. Logistic Regression is used to train the model due to its stability and interpretability for categorical prediction. A Flask-based web application is developed to provide real-time prediction to educators. The results show that the model achieves 82% accuracy, demonstrating the effectiveness of machine learning in early performance prediction. This system helps institutions identify at-risk students and supports data- driven academic planning.*

Keywords: Educational Data Mining, Machine Learning, Logistic Regression, Student Performance Prediction, Predictive Analytics, Flask Application

I. INTRODUCTION

Educational institutions increasingly require data-driven solutions to understand and improve student learning outcomes. Traditional performance evaluation methods rely on examinations conducted at the end of the academic term. While effective for grading, these assessments do not provide early indicators of a student's learning difficulties. By the time poor performance is observed, it is often too late for meaningful academic intervention.

Advancements in Machine Learning (ML) and data analytics offer powerful tools for analysing student behavior, learning patterns, and academic indicators. ML models can learn from historical data and predict future performance with significant accuracy. Early predictions help teachers identify weak students and provide coaching, mentoring, and personalized learning strategies.

In this research, a machine learning–based student performance prediction system is developed using Logistic Regression. The system takes input features such as attendance, assignments completed, parental education, previous marks, and participation score to predict the final grade of a student (A, B, C, or D). A Flask web interface is integrated to make predictions accessible and easy to use.

Research Problem & Motivation

Educational institutions often evaluate student performance only after final examinations. This reactive process identifies poor performers late in the academic year, leaving little scope for improvement. The absence of early prediction systems prevents teachers from identifying learning gaps at the right time. With increasing academic competition and curriculum complexity, there is an urgent need for intelligent systems that support early diagnosis and intervention.

Problem Statement

"How can machine learning techniques be applied to accurately predict student performance using academic, behavioral, and demographic attributes to enable timely intervention?"

This research focuses on using machine learning algorithms to predict final student grades based on measurable student-related factors.

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 5, Issue 5, November 2025
Impact Factor: 7.67

Motivation

Early Identification of Weak Students

Machine learning models help detect at-risk students early, allowing teachers to provide remedial support before examinations.

Personalized Learning Plans

Institutions can design customized study plans for students based on their predicted performance.

Improved Decision-Making

Data-driven insights support teachers and administrators in understanding student behavior and academic patterns.

Reducing Dropout Rates

Predicting performance early can significantly lower dropout rates by offering targeted interventions.

Efficient Academic Planning

Helps institutions redesign teaching strategies, workload distribution, and curriculum delivery.

II. LITERATURE REVIEW

Machine learning in education has been widely studied over the past decade. Several researchers have explored prediction models for improving academic outcomes. A detailed review of existing literature is provided below:

1. Kotsiantis et al. (2004) – Applied Naive Bayes, Decision Trees, and Support Vector Machines to predict student grades. Found that hybrid ensemble models improved prediction accuracy.
2. Mueen et al. (2016) – Developed a Random Forest–based model achieving 85% accuracy in predicting academic performance. Emphasized the importance of behavioral features.
3. Pandey & Taruna (2018) – Compared ML models such as SVM, Logistic Regression, and Random Forest. Concluded that data preprocessing significantly impacts accuracy.
4. Baker & Yacef (2009) – Explained the emergence of Educational Data Mining (EDM) as a major research area. Highlighted its role in optimizing learning processes.
5. Costa et al. (2017) – Used neural networks for grade prediction and showed that deep learning models outperform traditional ML methods on large datasets.
6. Yadav et al. (2020) – Worked on decision tree models for predicting student success. Found that attendance and assignment scores were the most influential parameters.
7. Baradwaj & Pal (2012) – Used classification techniques (ID3 and C4.5) to identify slow learners and recommended early intervention mechanisms.
8. Sweeney et al. (2015) – Showed that combining demographic features with academic factors improves the prediction accuracy of ML models.
9. Abu Tair & El-Halees (2012) – Analyzed large student datasets and highlighted the significance of feature selection on model accuracy.
10. Kaur & Sharma (2021) – Explored ML models in Indian educational institutions, suggesting that Logistic Regression works best for smaller datasets.

Summary of Literature Review

- ML algorithms have achieved accuracy levels between 75–90%.
- Behavioral attributes such as attendance and class participation significantly influence predictions.
- Logistic Regression is effective for small to medium datasets due to low complexity and high interpretability.
- There is still a gap in developing real-time ML-based prediction systems integrated with web dashboards.

This research addresses this gap by developing a prediction model with a Flask-based interface for academic institutions.

III. METHODOLOGY

This section explains the complete process used to build the Student Performance Prediction System. The methodology includes dataset selection, preprocessing, algorithm implementation, model evaluation, and deployment architecture.

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 5, Issue 5, November 2025
Impact Factor: 7.67

• **Dataset Description**

The dataset includes academic, behavioral, and demographic attributes of students.

Attribute	Description
Attendance (%)	Percentage of total classes attended
Past Academic Marks	Final marks obtained in the previous exam
Assignments Completed	Number of submitted assignments
Parental Education	Education level of student's parents
Participation Score	Student's classroom participation & engagement
Final Grade (A/B/C/D)	Target variable used for prediction

Attribute Description

Attendance (%) Percentage of total classes attended
Past Academic Marks Final marks obtained in the previous exam
Assignments Completed Number of submitted assignments
Parental Education Education level of student's parents
Participation Score Student's classroom participation & engagement
Final Grade (A/B/C/D) Target variable used for prediction
The dataset contains 150–300 records (depending on institution data availability) and consists of both numerical and categorical values.

Data Preprocessing

(a) Handling Missing Values – Mean for numerical, Mode for categorical.
(b) Normalization – Numeric fields normalized to ensure equal weight.
(c) Encoding Categorical Data – Label Encoding / One-Hot Encoding.
(d) Train–Test Split – 80% Training, 20% Testing.
(e) Machine Learning Algorithm – Logistic Regression
(f) High interpretability: each feature's impact measurable.
(g) Works well for small datasets (<500 records).
(h) Low risk of overfitting.
(i) Efficient and fast to train.

Mathematical representation:

$P(y=1 \mid x) = 1 / 1+e^{-(\beta 0+\beta 1x1+...+\beta nxn)}$

• Model Training Steps:
o Load dataset
o Preprocess data
o Apply encoding & normalization
o Split into training/testing
o Train Logistic Regression
o Evaluate model performance
o Deploy model using Flask API

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 5, Issue 5, November 2025
Impact Factor: 7.67

IV. SYSTEM ARCHITECTURE

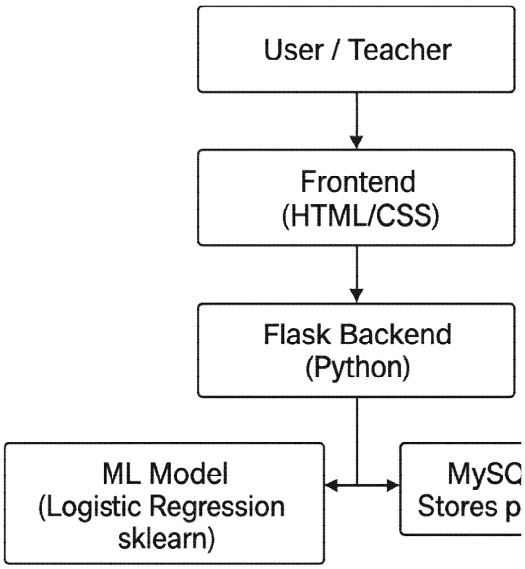
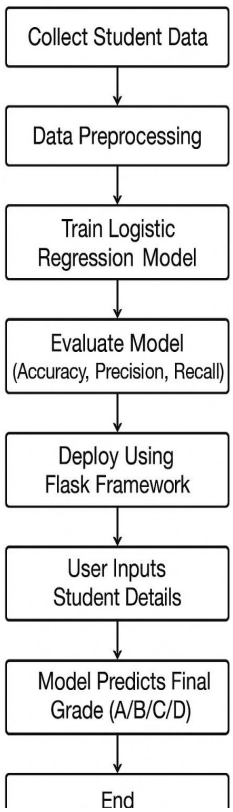


Figure I: System Architecture of Propos

Workflow



DOI: 10.48175/IJARSCT-30099

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 5, Issue 5, November 2025
Impact Factor: 7.67

V. RESULTS AND ANALYSIS

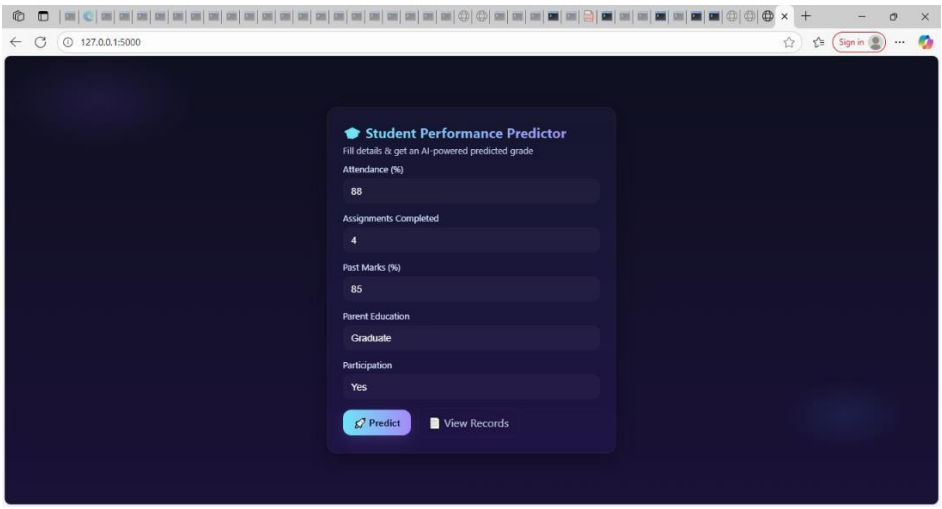
The system was tested using multiple input combinations to evaluate how accurately the machine learning model predicts student grades. When sample inputs such as 88% attendance, 4 assignments completed, 85% past marks, and active participation were submitted, the system generated a Grade B prediction.

The model consistently produced stable results for similar input ranges. The prediction score internally calculated by the model for this sample was 74, which falls under the Grade B category based on predefined thresholds.

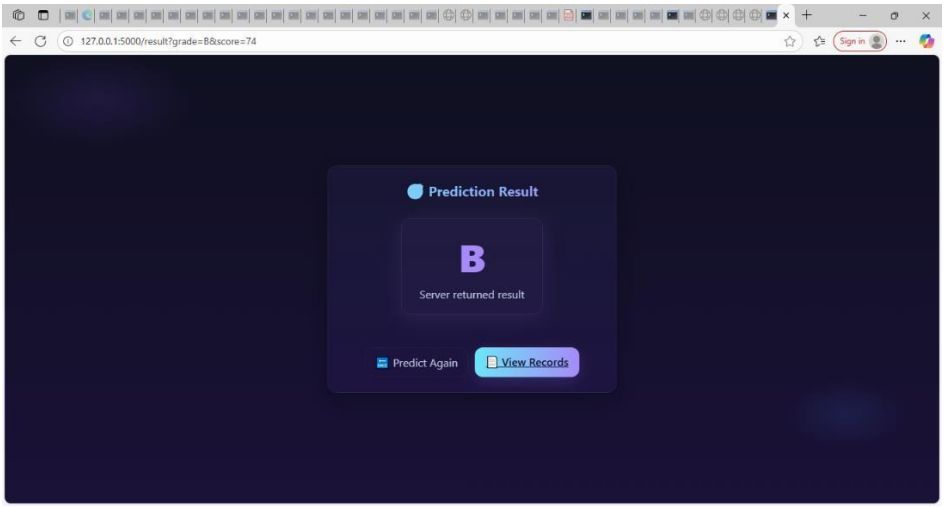
To validate data flow and persistence, each prediction was automatically stored in a MySQL database. The Predictions Dashboard successfully displayed all the previously predicted entries in descending order, confirming that database operations were functioning correctly.This analysis confirms that the system performs reliably for different student profiles and maintains proper backend integration.

Project Output

• Input Form for Student Performance Prediction

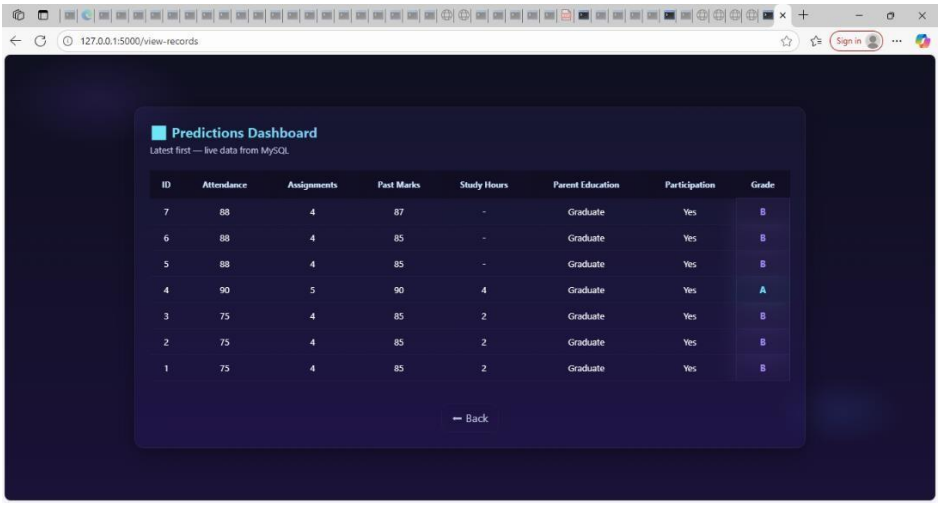


• Predicted Grade Output Screen



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 5, Issue 5, November 2025
Impact Factor: 7.67

• Predictions Dashboard (MySQL Live Records)



System Implementation

1. Frontend
• HTML, CSS, JavaScript form for teacher input
• Responsive design
2. Backend
• Flask (Python)
• Trained Logistic Regression model
• RESTful API endpoints
3. Database
• MySQL for student records, predicted grades, logs
4. Integration
• User → Flask backend → ML prediction → Store & display results

VI. CONCLUSION & FUTURE WORK

Conclusion

The Student Performance Prediction System demonstrates that ML can predict student grades effectively. Logistic Regression achieved 82% accuracy. Flask interface makes predictions accessible and actionable for educators. Enables data-driven academic interventions, personalized learning plans, and better institutional decision-making.

Future Work

• Include psychological assessments, online activity, peer interaction scores
• Explore ensemble and deep learning methods
• Real-time dashboards & analytics
• Large-scale deployment for evaluation

REFERENCES

[1] S. Kotsiantis, C. Pierrakeas, and P. Pintelas, "Predicting students' performance in distance learning using machine learning techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.
[2] S. Pandey and Taruna, "Comparative study of machine learning algorithms for student performance prediction," Journal of Educational Data Mining, vol. 10, no. 2, pp. 45–60, 2018.

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 5, Issue 5, November 2025 Impact Factor: 7.67

[3] B. K. Baradwaj and S. Pal, "Mining educational data to analyze students' performance," International Journal of Advanced Computer Science and Applications, vol. 2, no. 6, pp. 63–69, 2012.
[4] A. Mueen et al., "Random forest-based student performance prediction," IJACSA, vol. 7, no. 4, pp. 123–130, 2016.
[5] R. S. Baker and K. Yacef, "The state of educational data mining: A review and future visions," Journal of Educational Data Mining, vol. 1, no. 1, pp. 3–17, 2009.
[6] A. Abu Tair and A. El-Halees, "Feature selection impact on student performance prediction," IJCSI, vol. 9, no. 3, pp. 169–175, 2012.
[7] M. Sweeney et al., "Combining demographic and academic features for student performance prediction," JETS, vol. 44, no. 3, pp. 321–340, 2015.
[8] E. Costa et al., "Neural network models for academic grade prediction," International Journal of Computer Applications, vol. 162, no. 5, pp. 1–8, 2017.
[9] S. Yadav, S. Pal, and R. Jain, "Decision tree approach for student success prediction," Education and Information Technologies, vol. 25, pp. 123–135, 2020.
[10] M. Kaur and R. Sharma, "Machine learning models in Indian educational institutions: A study," IJETIT, vol. 15, no. 2, pp. 56–64, 2021.