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, January 2026

Impact Factor: 7.67

Student Performances Prediction System

Shravani Raut[1], Divya Mhasalkar[2], Nikita Bochare[3], Prof. N. S. Kharatmal[4]

Students, Department of Computer Science[1,2,3]

Lecturer, Department of Computer Science[4]

Matsyodary Shikshan Sanstha's Engineering & Polytechnic Jalna

shravaniraut981@gamil.com[1] ,diupatil59@gmail.com[2],

nikitabochare74@gmail.com[3], nanditakharatmal1126@gmail.com[4]

Abstract: *A student performance prediction system uses Machine Learning (ML) and AI to analyse student data (grades, attendance, habits, demographics) to forecast academic success, identify at-risk students early, and enable personalized interventions for better outcomes, often leveraging models like Random Forest, SVM, or ANN to improve educational strategy and support. The system pre-processes historical student data, selects relevant features, and applies predictive models to estimate future performance. By analysing patterns and relationships within the data, the proposed system provides accurate predictions that can assist educators in making informed decisions, offering timely interventions, and personalizing learning strategies. This paper presents a student performance prediction system using machine learning algorithms to analyse academic, demographic, and behavioural data. The proposed system predicts student performance such as pass/fail status and final grades. Algorithms such as Decision Tree, Random Forest, and Logistic Regression are applied to the dataset to evaluate prediction accuracy. Experimental results show that Random Forest provides the highest accuracy among the implemented models*

Keywords*:* Student performance prediction, machine learning, predictive analytics, educational outcomes, personalized Learning, data-driven insights, academic success, intervention strategies

I. INTRODUCTION

A Student Performance Prediction System aims to forecast students' academic outcomes based on historical and real-time data such as demographic information, attendance records, assessment scores, learning activities, and behavioural patterns. Early identification of students at risk of poor academic performance enables educators and administrators to implement timely interventions, provide personalized learning support, and optimize instructional strategies. Traditional evaluation methods often rely on periodic assessments and subjective judgment, which may fail to capture complex patterns influencing student success. Traditional methods of performance Evaluation are often manual and lack scalability, limiting their effectiveness in large educational settings. This Project leverages machine learning techniques to analyses diverse student data, including grades, attendance, and Behaviour, to deliver accurate and actionable predictions.

II. LITERATURE SURVEY

2.1 Growth of Student Performance Prediction Systems Several researchers have observed significant growth in Student performance prediction systems due to the increased adoption of digital education platforms, Learning Management Systems (LMS), and online assessment tools. The availability of large-scale educational data, including Attendance records, academic scores, and interaction logs, has enabled institutions to apply data mining and machine Learning techniques for performance analysis. Studies indicate that educational institutions increasingly rely on Predictive models to improve learning outcomes, reduce dropout rates, and support data-driven academic decision-Making.

2.2 Accessibility and Academic Support Many studies highlight that student performance prediction models improve Accessibility to academic support for diverse learner groups, including slow learners, working students, and students

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, January 2026
Impact Factor: 7.67

From rural or disadvantaged backgrounds. These systems enable continuous monitoring of academic progress and Allow early identification of at-risk students. By providing timely alerts and personalized interventions such as Remedial classes and mentoring, predictive systems help students manage their learning more effectively and enhance Overall academic success.

2.3 Performance Improvement and Institutional Benefits Literature indicates that predictive analytics in education Helps institutions improve academic performance and optimize resource utilization. Machine learning-based models Assist educators in identifying learning gaps and adapting teaching strategies accordingly. Early prediction of poor Performance reduces failure rates and enhances student retention. Several studies report that predictive insights Contribute to improved curriculum planning and institutional performance metrics.

2.4 Adoption and Student Behavioural Analysis Studies analyse factors influencing the adoption of student Performance prediction systems, including ease of use, perceived accuracy, and trust in predictive outcomes. Behavioural parameters such as attendance patterns, assignment submission behaviour, online engagement, and Participation levels are found to be strong indicators of academic performance. However, recent systematic reviews Also highlight concerns related to data privacy, bias in prediction models, and the need for transparent and ethical use Of student data.

Technology Integration in Educational Analytics Research shows that modern technologies such as cloud based Data storage, AI-driven analytics, learning dashboards, and real-time monitoring systems play a vital role in enhancing Student performance prediction frameworks. Integration of predictive models with LMS platforms enables automated Data collection and continuous analysis.

III. EXISTING MODELS AND CURRENT LIMITATIONS

Despite notable advancements in student performance prediction techniques, existing systems still face several limitations that affect their effectiveness in real-world educational environments. These challenges are particularly significant in diverse academic settings where learning behaviours, socio-economic factors, and institutional practices vary widely.

3.1 Over-Reliance on Historical Academic Data Most baseline prediction models rely heavily on historical academic records such as previous grades and examination scores. While effective for short-term predictions, these models often fail to capture real-time learning behaviours, motivation levels, and external factors influencing student performance. This over-reliance limits the adaptability and accuracy of predictions.

3.2 Inadequate Handling of Diverse Learning Patterns A major limitation of existing models is their inability to effectively handle diverse student backgrounds and learning styles. Many models are trained on limited or homogeneous datasets and do not adequately consider factors such as language proficiency, socio-economic status, and individual learning pace. This reduces prediction accuracy, particularly in large and heterogeneous student populations.

3.3 Vulnerability to Gradual Performance Decline Traditional prediction systems primarily focus on identifying sudden drops in academic performance, such as failing grades. However, gradual performance deterioration caused by disengagement, lack of motivation, or personal challenges often goes undetected. This results in delayed intervention and missed opportunities for early academic support.

3.4 Inability to Correlate Behavioural and Academic Indicators Many current frameworks analyse behavioural data and academic outcomes independently, without validating their consistency. For example, a student may show high engagement in online platforms but still perform poorly due to ineffective study strategies. The lack of integrated analysis limits the interpretability and practical usefulness of prediction results.

3.5 Absence of Cross-Semester and Cross-Course Intelligence Most student performance prediction systems operate in isolation, focusing on a single subject or semester. Academic performance patterns often span multiple courses and academic terms. The absence of cross-semester and cross-course analysis restricts the system’s ability to capture long-term trends and develop comprehensive student performance profiles.

DOI: 10.48175/568

ISSN: 2581-9429
Volume 6, Issue 3, January 2026
Impact Factor: 7.67

IV. PROPOSED MODEL AND METHODOLOGY

System Architecture

Rather than employing a generic black-box classifier, the proposed Feature-Enriched Machine Learning (FEML) framework is designed as a multi-layer analytical pipeline for student performance prediction. In the academic context, each student record is treated as a dynamic learning behaviours profile rather than a static set of grades. The architecture integrates academic, behavioural, and engagement-based indicators to achieve accurate and interpretable performance prediction.

1. Pre-processing of Academic and Behavioural Student Data Real-world educational data is often noisy, incomplete, and heterogeneous. Student datasets typically include missing attendance records, inconsistent grading scales, and unstructured behavioural indicators such as LMS activity logs. Unlike traditional pipelines that discard incomplete records, the FEML framework applies selective pre-processing. Missing values are handled using statistical imputation, while categorical attributes such as course type, assessment format, and learning mode are encoded meaningfully. Behavioural indicators such as irregular attendance, late assignment submissions, and fluctuating online activity are preserved, as they carry significant predictive value. This careful pre-processing ensures that genuine learning patterns are not mistakenly treated as noise.

2. Tri-Layer Analytical Logic The core intelligence of the FEML framework lies in its three-layer analytical model, Where each layer captures a different dimension of student performance.

Layer 1: Academic Pattern Analysis This layer evaluates traditional academic indicators such as internal assessment Scores, quiz results, attendance percentage, and historical GPA. Features such as grade consistency, improvement Trends, and assessment variability are analysed. A key risk indicator identified is performance inconsistency, where a Student shows sudden score drops without proportional attendance decline, often signalling conceptual gaps or Assessment anxiety.

Layer 2: Behavioural and Engagement Analysis This layer focuses on student behavioural patterns across the Semester. Metrics such as LMS login frequency, assignment submission regularity, participation in discussions, and Study material access are analysed. Special attention is given to irregular engagement patterns, where students remain Inactive for long durations and then exhibit short bursts of activity before evaluations—often associated with poor Academic outcomes.

Layer 3: Performance–Engagement Consistency Check The third layer performs a cross-validation between Academic performance and engagement behaviour. For example, students with high engagement but low scores trigger An intervention alert, indicating ineffective learning strategies. Conversely, low engagement with high scores may Suggest surface learning or assessment bias. This layer improves prediction reliability by validating logical consistency Across indicators.

3. Stacked Ensemble Classification Engine Features extracted from all three layers are aggregated and fed into a Stacked ensemble classifier consisting of Decision Tree, Random Forest, and Support Vector Machine models. This Ensemble approach enhances robustness and enables the system to capture both linear and non-linear relationships Within educational data. The hybrid configuration improves prediction accuracy while reducing model bias.

4. Probabilistic Performance Risk Scoring Instead of assigning rigid categorical labels such as pass or fail, the FEML framework generates a Performance Risk Score ranging from 0 to 100. This score represents the probability of Academic underperformance. Such a probabilistic approach allows educators to prioritize high-risk students for early Intervention while avoiding unnecessary actions forborderline cases.

```
+----------------------+
| Student Dataset|
| (Marks,       |
| Attendance,      |
| Internal Scores) |
+---------+------------+
          |
          v
```


DOI: 10.48175/568

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, January 2026
Impact Factor: 7.67

```
+-----------------------+
| Data Processing  |
| (Cleaning &|
| Preparation)     |
+---------+------------+
          |
          v
+-----------------------+
| ML Algorithm     |
| (Decision Tree / |
| Random Forest)   |
+---------+------------+
          |
          v
+----------------------+
| Performance      |
| Prediction       |
| (Pass/Fail /     |
| Grade)           |
+---------+------------+
          |
          v
+-----------------------+
| Result Display   |
| (Teacher /       |
| Admin)           |
+-----------------------+
```

Methodology

The experimental methodology was designed to reflect realistic academic environments and institutional constraints.

Targeted Data Collection: Data was collected from core undergraduate courses across multiple semesters, including Attendance records, internal assessments, final examination scores, and LMS interaction logs.

Technology Stack: The system was implemented using Python 3.11, with machine learning libraries such as Sickie-Learn and Pandas for data processing and model development. LMS log analysis was integrated using structured data Parsing techniques.

Feature Engineering: Key academic and behavioural features such as attendance deviation, assignment punctuality, Engagement frequency, and assessment trend scores were engineered to improve model interpretability and Performance.

Validation Strategy: An 80:20 train–test split was employed along with Stratified K-Fold Cross-Validation to ensure Class balance and generalization. This validation strategy prevents overfitting and ensures the model learns generalized Student performance patterns rather than course-specific artefacts.

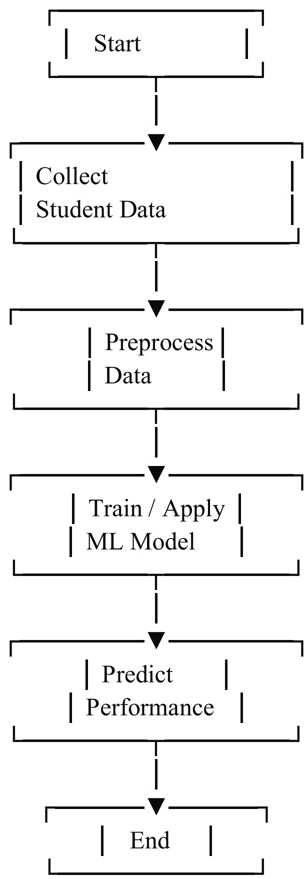
V. ALGORITHM

Step-by-Step Algorithm for Student Performance Prediction
Start
Collect student academic and behavioural data.
Store the collected data in the student database.
Perform data pre-processing on the dataset.

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, January 2026
Impact Factor: 7.67

Select relevant features affecting student performance.
Apply machine learning algorithms on the processed data.
Train the model using historical student data.
Test the model using current student data.
Predict student performance category (Pass / Fail / Grade).
Display the prediction results.
End

Flowchart



VI. OUTPUT/ RESULT

The Student Performance Prediction System generates predictions on students' academic outcomes, such as:
Grade Prediction: Forecasted grades (e.g., A, B, C) or numerical scores (e.g., 80%, 70%)
Pass/Fail Status: Likelihood of passing or failing a course
Risk Level: Categorization of students as low, medium, or high risk of underperforming
Performance Category: Excellent, Average, or Poor
Discussion: The system's predictions enable educators to:
densify At-Risk Students: Targeted support for students likely to underperform
Personalize Learning: Tailored interventions and resources for individual students
improve Academic Outcomes: Data-driven decision-making for curriculum and support services
Enhance Student Engagement: Proactive measures to boost motivation and participation

DOI: 10.48175/568

Example Output:

Student ID	Predicted Grade	Risk Level
S1	B+	Low
S2	C-	Medium
S3	F	High

VII. CONCLUSION

The Student Performance Prediction System is designed to analyse students' academic and behavioural data to predict their future performance accurately. By using machine learning techniques, the system helps identify students who may be at risk of poor academic results at an early stage. This system considers various factors such as attendance, internal marks, study hours, previous academic records, and participation in activities. Based on these inputs, the model predicts student performance, which can assist teachers and institutions in taking timely corrective actions like extra classes, or personalized guidance. The proposed system improves decision-making in the education sector by reducing manual analysis and increasing prediction accuracy. It also helps enhance overall student outcomes by supporting early intervention strategies. In the future, the system can be enhanced by including psychological factors, real-time data, and advanced algorithms to achieve even better results.

REFERENCES

[1] Cortez, P., & Silva, A., "Using Data Mining to Predict Secondary School Student Performance," Proc. of 5th Annual Future Business Technology Conference, Porto, Portugal, pp. 5–12, 2008.
[2] Kotsiantis, S. B., Pierrakeas, C. J., &Pintelas, P. E., "Predicting Students' Performance in Distance Learning Using Machine Learning Techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.
[3] Ahmed, A. B. E. D., &Elaraby, I. S., "Data Mining: A Prediction for Student's Performance Using Classification Method," World Journal of Computer Application and Technology, vol. 2, no. 2, pp. 43–47, 2014.
[4] Kumar, S. A., &Vijayalakshmi, M. N., "Efficiency of Decision Trees in Predicting Student's Academic Performance," International Journal of Computer Science and Mobile Computing, vol. 2, no. 6, pp. 335–343, 2013.
[5] Jayaprakash, S., Moody, E. W., Lauría, E. J. M., Regan, J. R., & Baron, J. D., "Early Alert of Academically At-Risk Students: An Open Source Analytics Initiative," Journal of Learning Analytics, vol. 1, no. 1, pp. 6–47, 2014.
[6] Shreemathi, R., & Krishnamurthy, K., "Student Performance Prediction Using Machine Learning Techniques," International Journal of Engineering Research & Technology (IJERT), vol. 8, no. 6, pp. 742–746, 2019.
[7] Pal, S., "Mining Educational Data to Analyze Students'Performance," International Journal of Advanced Computer Science and Applications, vol. 3, no. 2, pp. 63–69, 2012.
[8] Rai, A., & Jain, R., "Predicting Student Performance Using Data Mining Techniques," International Journal of Computer Applications, vol. 107, no. 15, pp. 23–28, 2014.
[9] Osmanbegović, E., &Suljić, M., "Data Mining Approach for Predicting Student Performance," Economic Review, vol. 10, no. 1, pp. 3–12, 2012.
[10] Bhardwaj, B. K., & Pal, S., "Data Mining: A Prediction for Performance Improvement Using Classification," International Journal of Computer Science and Information Security, vol. 9, no. 4, pp. 136–140, 2011.