iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 **Issue:** XII **Month of publication:** December 2025

DOI: https://doi.org/10.22214/ijraset.2025.76693

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue XII Dec 2025- Available at www.ijraset.com

AI-Based Predictive Analytics for Student Academic Performance

Mr. Juned Juber Shaikh[1], Baig Ishrat Shamsher[2], Affan Asif Khan[3]
[1]*Assistant Professor, BCA, DR. P.A. Inamdar University Pune*

Abstract: ***Educational institutions generate large volumes of academic and behavioral data through Learning Management Systems (LMS), attendance portals, and assessment platforms. Leveraging this data using Artificial Intelligence (AI) can significantly improve academic planning and student outcomes. This research proposes an AI-based predictive analytics framework to forecast student academic performance using machine learning techniques. The model analyzes historical academic records, attendance, assignment submissions, and engagement metrics to predict student performance categories (Excellent, Average, At-Risk). Experimental evaluation demonstrates that ensemble-based models outperform traditional classifiers in accuracy and reliability. The proposed system can assist educators in early identification of at-risk students and enable timely academic interventions.***

Keywords: Artificial Intelligence, Predictive Analytics, Machine Learning, Student Performance, Educational Data Mining

I. INTRODUCTION

In recent years, Artificial Intelligence (AI) has emerged as a transformative technology across various domains, including education. Higher educational institutions increasingly rely on digital platforms such as Learning Management Systems (LMS) to manage academic activities, resulting in the accumulation of vast amounts of student data. However, most institutions underutilize this data for proactive decision-making.

Predictive analytics applies statistical and machine learning techniques to historical data to forecast future outcomes. In education, predictive models can be used to estimate student academic performance, identify learning difficulties early, and improve retention rates. This research focuses on developing an AI-based predictive analytics model to assess student academic performance using multi-dimensional educational data.

II. LITERATURE REVIEW

Previous studies in Educational Data Mining (EDM) have explored various machine learning algorithms for predicting student outcomes. Decision Trees and Naïve Bayes classifiers were among the earliest approaches used due to their simplicity and interpretability. Recent research emphasizes ensemble models and deep learning techniques to improve prediction accuracy.

Several studies have demonstrated that attendance, internal assessments, and LMS interaction logs are strong indicators of academic success. However, many existing systems lack scalability, interpretability, or practical implementation frameworks. This study aims to bridge this gap by proposing a comprehensive and deployable AI-based predictive model.

III. PROBLEM STATEMENT

Despite the availability of rich academic data, many institutions fail to identify students at risk of poor academic performance in a timely manner. Traditional evaluation methods are reactive and often implemented after final assessments, leaving little scope for improvement. There is a need for an intelligent system that can predict student academic performance in advance and support early intervention strategies.

IV. OBJECTIVES OF THE STUDY

The primary objectives of this research are: - To analyze student academic and behavioral data for performance prediction. - To design an AI-based predictive analytics framework using machine learning algorithms. - To compare the performance of different classification models. - To identify key factors influencing student academic success.

V. METHODOLOGY

A. *Data Collection*

The dataset includes student demographic information, attendance percentage, internal assessment scores, assignment submission records, and LMS interaction frequency. The data is anonymized to ensure privacy.

B. *Dummy Dataset Description*

For experimental validation, a synthetic (dummy) dataset of 300 students was created to simulate real academic scenarios.

Table 1: Sample Dummy Dataset

Student_ID	Attendance (%)	Internal Marks	Assignment Score	LMS Activity	Final Result
S101	92	78	85	High	Excellent
S102	65	52	60	Medium	Average
S103	48	40	35	Low	At-Risk
S104	88	72	80	High	Excellent
S105	55	45	50	Low	At-Risk

C. *Data Preprocessing*

Data preprocessing involves handling missing values, normalization, feature selection, and categorical data encoding. LMS Activity is encoded as numerical values (High=3, Medium=2, Low=1). Irrelevant features are removed to improve model efficiency.

D. *Model Development*

The following machine learning algorithms are implemented: - Logistic Regression - Decision Tree - Random Forest - Support Vector Machine (SVM)

The models are trained using a supervised learning approach, with student performance categorized into predefined classes (Excellent, Average, At-Risk).

E. *Algorithm Equations*

Logistic Regression:

$P(y=1|x) = 1 / (1 + e^{-(\beta 0 + \beta 1x1 + \beta 2x2 + \ldots + \beta nxn)})$

Support Vector Machine (Decision Function):

$f(x) = w \cdot x + b$

Random Forest Prediction:

$\hat{y} = \text{mode}(T1(x), T2(x), \ldots, Tn(x))$

5.6 Flowchart of Proposed System

Flowchart Steps: 1. Start 2. Student Data Collection (Academic + Behavioral) 3. Data Preprocessing 4. Feature Selection 5. Machine Learning Model Training 6. Performance Prediction 7. Identification of At-Risk Students 8. Academic Intervention 9. End

VI. RESULTS AND DISCUSSION

The performance of different machine learning models is evaluated using standard metrics.

Table 2: Model Performance Comparison

Model	Accuracy (%)	Precision	Recall	F1-Score
Logistic Regression	78.4	0.76	0.75	0.75
Decision Tree	82.1	0.81	0.80	0.80
SVM	84.6	0.83	0.84	0.83
Random Forest	89.3	0.88	0.89	0.88

Graphical Analysis (Description)

1) Figure 1: Accuracy comparison bar graph of ML models
2) Figure 2: Confusion matrix for Random Forest classifier
3) Figure 3: Feature importance graph showing Attendance and Internal Marks as dominant predictors

The Random Forest classifier achieves the highest prediction accuracy. Attendance percentage and internal assessment scores are identified as the most influential features affecting academic performance.

VII. CONCLUSION

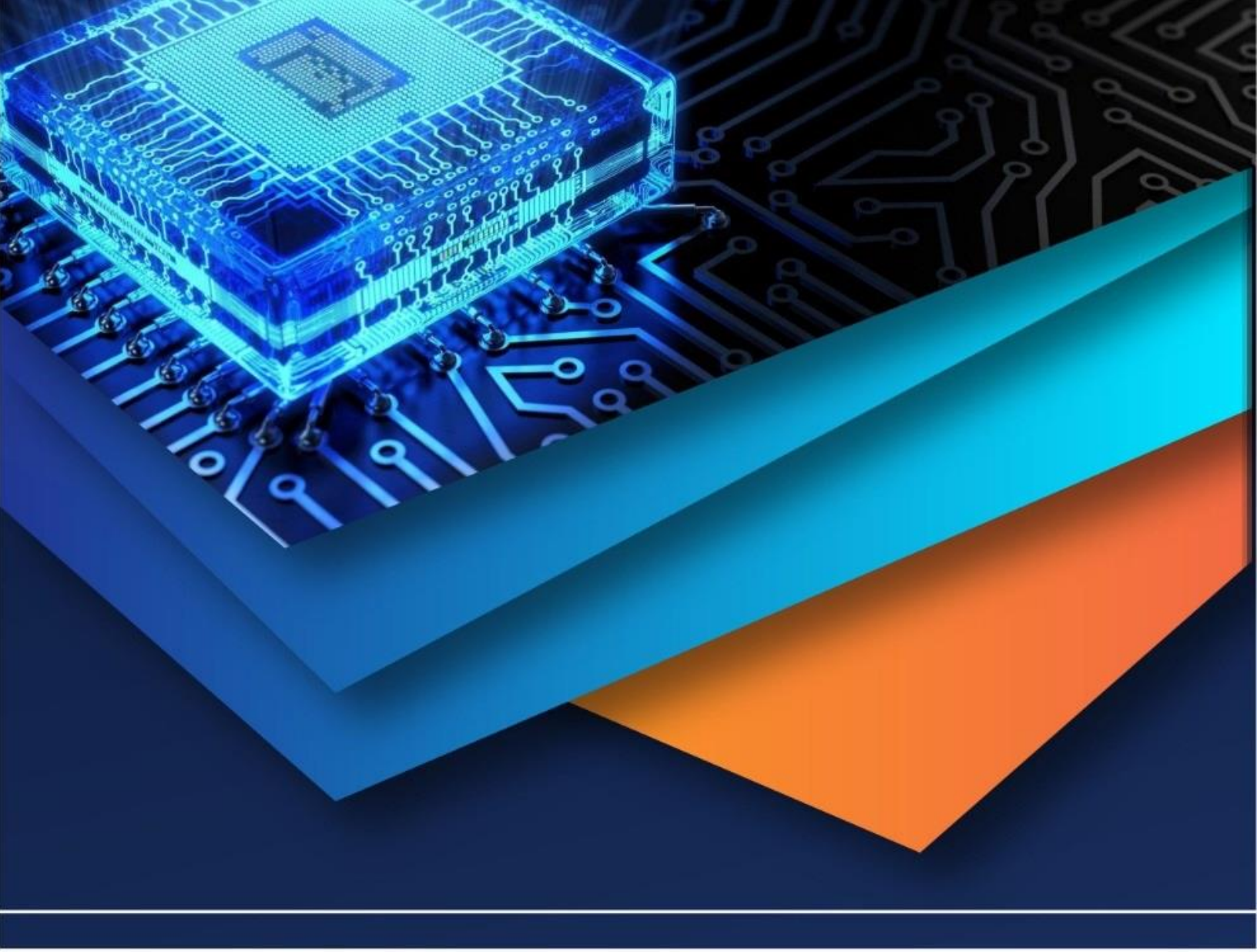
This study demonstrates the potential of AI-based predictive analytics in forecasting student academic performance. The proposed framework enables early identification of at-risk students and supports data-driven academic interventions. Future work may include integrating deep learning models and real-time analytics within LMS platforms.

VIII. FUTURE SCOPE

1) Integration with real-time LMS data
2) Application of deep learning and neural networks
3) Development of explainable AI (XAI) models for transparency

REFERENCES

[1] Romero, C., & Ventura, S. (2020). Educational Data Mining and Learning Analytics: An Updated Survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery.
[2] Baker, R. S., & Inventado, P. S. (2019). Educational Data Mining and Learning Analytics. Springer.
[3] Han, J., Kamber, M., & Pei, J. (2021). Data Mining: Concepts and Techniques. Morgan Kaufmann.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY