International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 13 Issue: 10 | Oct 2026 www.irjet.net p-ISSN: 2395-0072

STUDENT PERFORMANCE PREDICTION USING STUDENT DATA

Pooja Kawde[1], Anjali Shendge[2], Aniket Payagavhan[3], PROF. S. Chechare[4], Dr.S.N. Kakarwal[5]

[1,2,3] (Students, Department of Computer Engineering), International Center Of Excellence in Enggineering & Management

[45](Professor, Dept. of Computer Engineering), International Center Of Excellence in Enggineering & Management

---

Abstract - *Student academic performance prediction is an important application of machine learning in the education sector. This paper presents a Student Performance Prediction System that analyzes student data to estimate academic outcomes based on factors such as attendance, assignment scores, previous examination results, study habits, and classroom engagement. The main objective of the system is to recognize students who may face academic difficulties and provide timely support. The proposed system collects student information, preprocesses the dataset, and applies different machine learning algorithms to identify patterns related to academic achievement. The trained models are evaluated using suitable performance metrics to select an appropriate prediction model. The system helps teachers monitor student progress, provide personalized guidance, and make data-driven decisions to improve learning outcomes and academic success.*

Key Words: Student Performance Prediction, Machine Learning, Educational Data Mining, Academic Analysis, Predictive Modelling, Learning Analytics

1. INTRODUCTION

Education is an essential part of individual growth and career development. Student academic performance helps teachers and educational institutions understand learning progress and identify areas that require improvement. However, students may face academic challenges due to irregular attendance, insufficient study time, poor assessment results, and lack of participation in learning activities. Identifying these difficulties at an early stage can help teachers provide proper guidance and academic support.

With the advancement of technology and digital education platforms, a large amount of student-related information is available for analysis. This information includes attendance records, previous examination marks, assignment performance, study hours, classroom participation, and academic history. By using machine learning and data mining techniques, these records can be analyzed to predict student performance more effectively.

The Student Performance Prediction System uses historical student data to identify patterns that influence academic achievement. It predicts expected performance levels and helps identify students who may require additional attention. This approach supports teachers in making informed decisions, providing timely assistance, and improving the overall quality of education.



Fig 1: System Architecture

1. Educational System Layer

At the top level, the **Educational System** forms the operational environment. It includes:

- Traditional classroom learning
- E-learning or online learning platforms
- Hybrid learning environments

The educational system acts as the primary source of data and interaction between teachers and students.

1) Roles in the System

- **Teachers**: Responsible for planning, designing, and maintaining course structure and content.
- **Students**: Engage with the system by communicating, participating, and using academic resources.

2. Data Collection Layer

The system collects multiple types of student-related data from the educational environment:

2) a) Course Information

Includes:

- Course materials
- Syllabus structure
- Assignments and quizzes
- Teaching strategies
- Course difficulty level

3) b) Interaction Information

Includes:

- Student-teacher communication
- Forum discussions
- Online activity logs
- Attendance records
- Participation frequency

4) c) Academic Information

Includes:

- Previous grades
- Internal assessment scores
- Examination results
- GPA
- Academic history

5) d) Student Usage Data

Includes:

- Login frequency
- Time spent on learning platforms
- Resource access patterns
- Submission timelines
- Clickstream data

These datasets collectively provide a multidimensional view of student behavior and academic engagement.

3. Data Mining Layer

All collected data is processed using **Data Mining Tools**, which may include:

- Clustering algorithms
- Classification techniques
- Regression models
- Machine learning algorithms
- Deep learning approaches The purpose of this layer is to:

- Identify hidden patterns
- Detect at-risk students
- Discover behavioral trends
- Extract meaningful features from raw data

This stage transforms raw educational data into structured and analyzable knowledge.

4. Prediction Model Layer

The processed data is fed into a **Prediction Model**, which may include:

- Logistic Regression
- Decision Trees
- Random Forest
- Support Vector Machines
- Neural Networks
- Deep Learning models The prediction model aims to:

- Predict student academic performance
- Forecast pass/fail outcomes
- Estimate grade ranges
- Identify dropout risk
- Measure probability of academic success

5. Result and Feedback Layer

The final output of the system is the **Result**, which provides:

- Performance prediction
- Risk classification (low, medium, high risk)
- Personalized recommendations
- Early warning alerts

6) Impact of Results:

- **For Teachers**:
  - Improve course design
  - Modify teaching strategies
  - Provide targeted interventions
  - Enhance overall educational quality
- **For Students**:
  - Reduce failure rate
  - Receive personalized academic guidance
  - Improve study strategies
  - Increase academic success probability

II. LITERATURE SURVEY

Predicting student academic performance has become a prominent research area in educational data mining (EDM) due to the increasing availability of student-related data generated by learning management systems (LMS),

online platforms, and traditional classrooms. The objective of these systems is to leverage historical and real-time data to identify at-risk students, enhance instructional strategies, and improve overall educational outcomes.

Early studies focused primarily on the use of traditional statistical techniques to explore relationships between student demographic attributes and academic success. Romero and Ventura (2007) pioneered the use of educational data mining to extract patterns from educational datasets, demonstrating that student behavior, when analyzed correctly, can forecast future outcomes. Their work laid the foundation for linking academic data with predictive models that assist educators in decision-making processes.
In the last decade, researchers have transitioned from simple statistical methods such as linear regression and correlation analysis to more sophisticated machine learning approaches. For example, Kotsiantis et al. (2004) compared classification algorithms such as Decision Trees, Naive Bayes, and Support Vector Machines in predicting student grades, finding that ensemble methods often outperform traditional classifiers. These studies highlight the benefit of machine learning in handling complex and non-linear relationships within educational datasets.

Recent literature reflects a noticeable shift toward deep learning and data-driven approaches. Deep neural networks, such as Long Short-Term Memory (LSTM) and Convolutional Neural Networks (CNN), have been applied to sequential and behavioral data to capture temporal engagement patterns that traditional techniques cannot extract. Al-Balawi et al. (2018) successfully utilized recurrent neural networks to model student activity logs in online courses, showing significant improvement in prediction accuracy compared to baseline models.

Several researchers have also emphasized feature engineering as a critical step in performance prediction. Student interaction logs, resource access patterns, time-on-task measurements, and forum participation metrics have been shown to contribute strongly to predictive accuracy. For example, Yang and Liu (2019) demonstrated that combining academic scores with interaction data from LMS platforms significantly enhances the model's ability to detect early signs of student disengagement.

III. Proposed Methods

1. Overview

This study proposes a data-driven Student Performance Prediction System that integrates educational data collection, preprocessing, feature engineering, machine learning modeling, and performance evaluation. The primary objective of the proposed method is to accurately predict student academic outcomes and identify at-risk students at an early stage to enable timely intervention.

2. Data Collection

The proposed system collects multidimensional data from the educational environment (traditional and e-learning platforms).

3. Data Preprocessing

Before model development, the collected data undergoes preprocessing to improve quality and consistency.

4. Feature Engineering and Selection

Feature engineering is performed to enhance model performance. Derived features include:

- Engagement score (based on login frequency and time spent)
- Assignment consistency index
- Participation ratio
- Performance improvement trend

5. Model Development

The proposed system employs a hybrid machine learning approach combining supervised learning techniques.

6. Model Evaluation

The performance of the proposed models is evaluated using standard metrics.

7. Prediction and Early Warning Mechanism

The final trained model is integrated into the educational system to provide real-time predictions.

8. System Workflow Summary

1. Collect multidimensional student data.
2. Clean and preprocess the dataset.
3. Extract and select relevant features.
4. Train multiple machine learning models.
5. Evaluate and select the optimal model.
6. Deploy the model for continuous prediction and intervention.

9. Advantages of the Proposed Method

- Early identification of at-risk students
- Data-driven academic decision-making
- Improved teaching strategies
- Reduced failure and dropout rates
- Scalable for both traditional and online education systems

IV. OBJECTIVES OF THE PROJECT

The primary objective of this project is to design and implement an intelligent Student Performance Prediction System capable of analyzing academic and behavioral data to forecast student outcomes accurately. The system aims to support educators and institutions in making data-driven decisions that enhance learning effectiveness and reduce academic failure rates.

The specific objectives of the project are as follows:

1. To Develop a Data-Driven Prediction Framework

To design a comprehensive framework that integrates academic records, interaction data, course information, and student usage patterns into a unified predictive model. The system should efficiently handle structured and semi-structured educational data collected from traditional and e-learning environments.

2. To Identify Key Factors Influencing Student Performance

To analyze and determine the most significant academic and behavioral features that influence student success or failure. This includes evaluating factors such as attendance, participation, assignment submission patterns, assessment scores, and engagement levels.

3. To Implement and Compare Machine Learning Models

To implement multiple machine learning and data mining algorithms (e.g., Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, Neural Networks) and evaluate their effectiveness in predicting student performance. The objective is to identify the most accurate and reliable model for the given dataset.

4. To Predict Academic Outcomes at an Early Stage

To develop a predictive mechanism capable of identifying at-risk students early in the academic term. Early prediction enables timely intervention strategies, mentoring, and personalized academic support.

5. To Design an Early Warning and Feedback System

To create a system that provides actionable insights and performance reports to both educators and students. Teachers can use this information to modify instructional strategies, while students can receive personalized recommendations to improve their performance.

6. To Improve Educational Quality and Reduce Failure Rates

To contribute to the enhancement of overall educational effectiveness by reducing student dropout rates, minimizing academic failures, and promoting student success through data-driven interventions.

7. To Ensure Scalability and Adaptability

To design the system in a scalable manner so that it can be applied across different institutions, courses, and learning environments, including traditional classrooms, blended learning systems, and fully online platforms.

8. To Enhance Model Accuracy and Generalization

To apply appropriate preprocessing, feature selection, and validation techniques to ensure that the developed model maintains high accuracy, robustness, and generalizability across different student populations.

CONCLUSIONS

This study presented a comprehensive Student Performance Prediction System designed to analyze academic and behavioral data for forecasting student outcomes. By integrating course information, academic records, interaction logs, and usage patterns within a unified framework, the proposed system demonstrates the effectiveness of data-driven approaches in educational environments.

The implementation of machine learning and data mining techniques enables accurate identification of patterns and relationships that influence student success. Through preprocessing, feature engineering, and model evaluation, the system effectively predicts academic performance and classifies students based on risk levels. Comparative analysis of multiple algorithms ensures the selection of an optimal predictive model with high accuracy and reliability.

One of the key contributions of this work is the development of an early warning mechanism that allows institutions and educators to detect at-risk students at an early stage. This proactive approach supports timely interventions, personalized guidance, and improved teaching strategies, ultimately reducing failure and dropout rates.

Furthermore, the system establishes a continuous feedback loop between students, teachers, and the educational platform. The insights generated from predictive analytics assist educators in refining course design and instructional methods, while students benefit from actionable recommendations that enhance their learning outcomes.

Although the proposed model demonstrates promising results, challenges such as data heterogeneity, privacy concerns, and model generalization across institutions

remain areas for further research. Future work may focus on incorporating advanced deep learning techniques, real-time analytics, explainable AI methods for model interpretabilityand cross-institutional validation to enhance robustness and scalability.

In conclusion, the Student Performance Prediction System represents a significant step toward intelligent, data-driven education. By leveraging predictive analytics, the system contributes to improved academic performance, informed decision-making, and the overall enhancement of educational quality.

REFERENCES

[1] 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, 2010.

[2] C. Romero and S. Ventura, "Data Mining in Education," Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12–27, 2013.

[3] S. B. Kotsiantis, C. J. Pierrakeas, and P. E. Pintelas, "Predicting Students' Performance in Distance Learning Using Machine Learning Techniques," Applied Artificial Intelligence, vol. 18, no. 5, pp. 411–426, 2004.

[4] A. L. Asif, A. Merceron, S. A. Ali, and N. G. Haider, "Analyzing Undergraduate Students' Performance Using Educational Data Mining," Computers & Education, vol. 113,

pp. 177–194, 2017.

[5] M. Dekker, M. Pechenizkiy, and J. M. Vleeshouwers, "Predicting Students Drop Out: A Case Study," in Proceedings of the International Working Group on Educational Data Mining, 2009.

[6] E. Fernandes, M. Holanda, M. Victorino, V. Borges, R. Carvalho, and G. Van Erven, "Educational Data Mining: Predictive Analysis of Academic Performance of Public School Students in the Capital of Brazil," Journal of Business Research, vol. 94, pp. 335–343, 2019.

[7] A. M. Shahiri, W. Husain, and N. A. Rashid, "A Review on Predicting Student's Performance Using Data Mining Techniques," Procedia Computer Science, vol. 72, pp. 414–422, 2015.

[8] J. Xu and K. H. Moon, "A Machine Learning Approach for Tracking and Predicting Student Performance in Degree Programs," IEEE Journal of Selected Topics in Signal Processing, vol. 11, no. 5, pp. 742–753, 2017.

[9] S. Helal, J. Li, L. Liu, E. Ebrahimie, S. Dawson, and D. J. Murray, "Predicting Academic Performance by Considering Student Heterogeneity," Knowledge-Based Systems, vol. 161,

pp. 134–146, 2018.

[10]H. Al-Balawi, M. Al-Ajlan, and A. Al-Ajlan, "Predicting Students' Performance in MOOCs Using Deep Learning Approaches," International Journal of Advanced Computer Science and Applications, vol. 10, no. 6, pp. 1–7, 2019.

BIOGRAPHIES

1) **PROF. S. CHECHARE**
Pursuing (CSE) ICEEM INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)

2) **Dr.S.N.KAKARWAL**
Pursuing (CSE) ICEEM INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)

3) **MS. POOJA.KAWDE**
Pursuing (CSE) ICEEM INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)

4) **MS. ANJALI SHENDGE**
Pursuing (CSE) ICEEM INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)

5) **MR.ANIKET PAYAGAVHAN**
Pursuing (CSE) ICEEM INTERNATIONAL CENTER OF EXCELLENCE IN ENGGINERING & MANAGEMENT (ENGINEERING)Z