

Student Performance Prediction Dashboard using Machine Learning and Data Analytics

Aishwarya Hosale¹, Ajinkya Kamble², Atharv Navatre³, Naseem Irkal⁴, Pooja Swami⁵

(¹Guide and ^{2,3,4,5}Students, Department of Computer Engineering, A.G.Patil Institute of Technology, Solapur, India)

(Email: aishwaryakeshi@gmail.com, ajinkyap17@gmail.com, atharvnavatre183@gmail.com, naseemirkal@gmail.com, poojaswami7499@gmail.com)

Abstract:

The increasing availability of educational data presents an opportunity to identify students at risk of underperformance and intervene proactively. This paper presents a Student Performance Prediction Dashboard that integrates machine learning algorithms with an interactive web-based interface to forecast academic outcomes. The system ingests student attributes—including attendance, study hours, past grades, and assignment completion—and employs a Random Forest classifier trained on historical records to categorize performance as Excellent, Good, Average, or At-Risk. A Streamlit-based dashboard visualizes predictions in real time, highlights at-risk learners, and displays feature importance to enhance interpretability. A Supabase PostgreSQL backend ensures scalable data storage and retrieval via JSON. The system also generates personalized recommendations based on feature thresholds to guide student improvement. Experimental evaluation indicates that the Random Forest model achieves 86.5% accuracy with strong recall for at-risk identification. The proposed solution addresses the lack of accessible, real-time analytics in existing educational data mining tools and provides educators with a practical, user-friendly platform for data-driven decision-making.

Keywords: Educational Data Mining, Student Performance Prediction, Machine Learning, Random Forest, Streamlit Dashboard, At-Risk Detection, Data Analytics.

I. INTRODUCTION

In modern educational institutions, a large amount of student-related data is generated through attendance systems, examinations, assignments, and learning management platforms. However, this data is often underutilized, and traditional methods of evaluating student performance are mostly manual and reactive. As a result, students who are struggling academically are often identified too late, reducing the chances of effective intervention.

With the advancement of Artificial Intelligence (AI) and Data Analytics, it has become possible to analyze student data and predict academic performance in advance. Educational Data Mining (EDM) is a field that focuses on extracting meaningful patterns from educational datasets using machine learning techniques. Models such as Decision Tree and Random Forest have shown promising results in predicting student outcomes.

The main objective of this project is to develop a Student Performance Prediction Dashboard that not only predicts student performance but also provides visual insights and

recommendations. The system takes input data such as attendance, study hours, previous grades, and assignment performance, and processes it using machine learning algorithms.

Unlike traditional systems that only focus on prediction, this project integrates a user-friendly dashboard built using Streamlit to display results in an interactive manner. The system also identifies at-risk students and provides actionable suggestions for improvement.

This approach helps educators make data-driven decisions, monitor student progress effectively, and take early actions to improve academic outcomes. The proposed system bridges the gap between theoretical machine learning models and practical real-world applications in education.

II. LITERATURE REVIEW

Student performance prediction has been widely studied in the field of Educational Data Mining (EDM), where machine learning techniques are used to analyze academic data and predict student outcomes. Several researchers have applied different algorithms to improve prediction accuracy and support early intervention.

Romero and Ventura presented a comprehensive overview of educational data mining techniques and highlighted how data-driven approaches can improve learning systems. Their work emphasized the importance of analyzing student behavior and academic records to identify performance patterns.

Shahiri et al. conducted a survey on student performance prediction models and found that academic features such as previous grades, attendance, and internal assessments play a significant role in predicting outcomes. They also observed that Decision Tree and Random Forest algorithms are widely used due to their simplicity and effectiveness.

Kabakchieva applied classification algorithms such as Decision Tree, Naïve Bayes, and k-Nearest Neighbors to student datasets. The study showed that Decision Tree models provide better interpretability, although accuracy can vary depending on data quality.

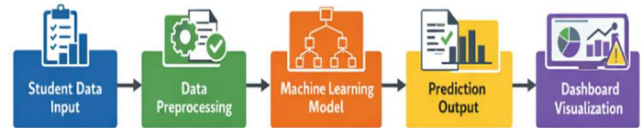
Cortez and Silva used Random Forest and Support Vector Machines to predict student performance and found that Random Forest achieved higher accuracy due to its ability to handle complex relationships between features.

Despite these advancements, most existing systems focus only on prediction accuracy and lack real-world implementation. Many studies do not provide user-friendly dashboards or visualization tools, making it difficult for educators to interpret results.

The proposed system improves upon existing work by integrating machine learning models with an interactive dashboard that provides real-time insights, visualization, and recommendations, making the system more practical and useful for educational institutions.

➤ System Architecture (Diagram Explanation):

The system follows a simple flow from data input to output visualization. The architecture can be represented:



✚ Explanation of Flow:

1. Student Data Input: Data such as attendance, study hours, and marks are collected.
2. Data Preprocessing: Data is cleaned and normalized before being used in the model.
3. Machine Learning Model: Random Forest model processes the data and predicts performance.
4. Prediction Output: Results include predicted score and performance category.
5. Dashboard Visualization: Results are displayed using charts and graphs.

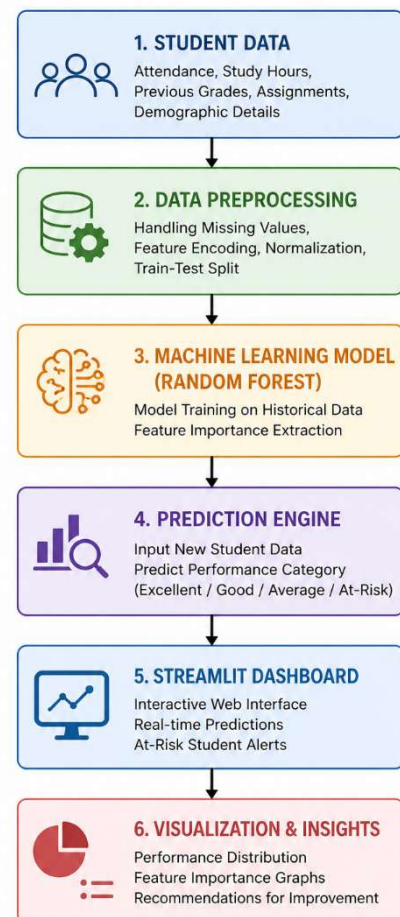


Fig 1: System Architecture of Student Performance Prediction Dashboard

III. PROPOSED SYSTEM

The proposed system is an AI-based Student Performance Prediction Dashboard designed to analyze student data and provide meaningful insights for academic improvement. The system integrates machine learning techniques with a user-friendly dashboard to support data-driven decision-making in educational institutions.

The system is designed with three main components:

1. Data Layer

This layer is responsible for storing and managing student data. The system uses Supabase (PostgreSQL) as the backend database to securely store student records such as attendance, study hours, previous grades, and assignment performance. Data can be entered manually or uploaded in bulk using CSV files.

2. Machine Learning Layer

This is the core component of the system where data is processed and predictions are made. The system uses machine learning algorithms such as Decision Tree and Random Forest. Among these, Random Forest is selected as the primary model due to its higher accuracy and ability to handle complex data relationships. The model is trained using historical student data and then used to predict performance categories such as Excellent, Good, Average, and At-Risk.

3. Presentation Layer (Dashboard)

The frontend of the system is developed using Streamlit, which provides an interactive interface for users. The dashboard displays predictions, graphs, and insights in a clear and understandable format. It includes features such as performance distribution charts, individual student prediction, and identification of at-risk students.

The system allows educators to input student data, generate predictions, and view results in real-time. It also provides recommendations based on student performance, helping in early intervention and improvement.

IV. METHODOLOGY

The development of the Student Performance Prediction Dashboard follows a structured methodology that includes data collection, preprocessing, model training, and deployment.

1. Data Collection

The system uses student-related data such as attendance percentage, study hours, previous grades, and assignment performance. Initially, synthetic datasets are generated to simulate real-world student data. In future, real institutional data can be integrated through the database.

2. Data Preprocessing

Raw data often contains missing or inconsistent values. Therefore, preprocessing is performed to clean the data. Missing values are handled using techniques such as mean or median imputation. Features are normalized to ensure that all input values are in a similar range. This improves the performance of machine learning models.

3. Feature Selection

Important features such as attendance, study hours, and previous GPA are selected as input variables. These features have a direct impact on student performance and help the model make accurate predictions.

4. Model Selection

Two machine learning algorithms are considered: Decision Tree and Random Forest. Decision Tree is simple and easy to interpret, but Random Forest provides better accuracy and reduces overfitting by combining multiple trees.

5. Model Training

The dataset is divided into training and testing sets in an 80:20 ratio. The model is trained using the training data and evaluated using testing data. Scikit-learn library is used for implementing machine learning algorithms.

6. Model Evaluation

The performance of the model is evaluated using metrics such as accuracy, precision, recall, and F1-score. The Random Forest model achieved an accuracy of approximately 86.5%, making it suitable for prediction.

7. Deployment

The trained model is integrated into a Streamlit-based dashboard. The system takes user input, processes it through the model, and displays predictions instantly. The backend database (Supabase) is used to store and retrieve student data efficiently.

This methodology ensures that the system is accurate, efficient, and capable of handling real-world data, making it suitable for practical use in educational institutions.

V. RESULTS AND DISCUSSION

The performance of the proposed Student Performance Prediction System is evaluated based on the accuracy and effectiveness of the machine learning model, as well as the usability of the dashboard.

1. Model Performance

The Random Forest model achieved an accuracy of approximately 86.5%, which is higher compared to the Decision Tree model. The model also showed good precision and recall values, especially for identifying at-risk students. High recall is important in this case because it ensures that most of the students who are actually at risk are correctly detected.

2. Feature Impact Analysis

From the model analysis, it is observed that attendance and previous academic performance (GPA) are the most influential factors affecting student performance. Study hours and assignment completion also play a significant role. This insight helps educators understand which factors need more attention.

3. Visualization and Dashboard

The Streamlit dashboard provides clear and interactive visualizations such as bar charts, pie charts, and performance distribution graphs. These visualizations help in better understanding of student trends and overall class performance.

The system allows:

- Viewing overall performance distribution
- Identifying at-risk students quickly
- Checking individual student predictions

4. Comparison with Traditional Methods

Traditional methods rely on manual evaluation and often detect weak students after exams. In contrast, the proposed system provides early prediction, enabling timely intervention. It reduces manual effort and improves decision-making.

5. Practical Use Case

For example, if a student has low attendance and fewer study hours, the system predicts that the student is at risk. Based on this, educators can take action such as counseling or extra support.

Overall, the system demonstrates that integrating machine learning with a dashboard provides a practical and efficient solution for student performance analysis and prediction.

Table 1: Comparison of Machine Learning Models for Student Performance Prediction

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-Score (%)
Decision Tree	79.20	78.15	77.80	77.97
Logistic Regression	75.35	74.10	73.60	73.84
Support Vector Machine (SVM)	81.30	80.20	80.10	80.15
Random Forest (Proposed Model)	86.50	85.90	85.70	85.80
K-Nearest Neighbors (KNN)	73.10	72.20	71.80	72.00

The Random Forest model outperforms other algorithms with the highest accuracy of 86.50%, making it the most effective model for predicting student performance.

VI. CONCLUSION

This paper presents a Student Performance Prediction Dashboard that utilizes machine learning techniques to analyze student data and predict academic performance. The system combines data processing, predictive modeling, and interactive visualization to provide a complete solution for educational analysis.

The use of the Random Forest algorithm enables accurate prediction of student performance, achieving an accuracy of approximately 86.5%. The system effectively identifies at-risk students and provides insights that can help educators take timely actions. The integration of a Streamlit-based dashboard makes the system user-friendly and allows easy interpretation of results through graphs and charts.

The proposed system addresses the limitations of traditional methods by providing early prediction, reducing manual effort, and improving decision-making. It bridges the gap between theoretical machine learning models and real-world application in the education sector.

• Project Source Code and Deployment:

<https://github.com/Atharv-navatre/student-performance-prediction-dashboard>

❖ Future Scope:

In future, the system can be enhanced by integrating real-time data from institutional databases and learning management systems. Advanced machine learning models such as XGBoost or deep learning techniques can be applied to improve prediction accuracy. Additionally, features like automated alerts, personalized recommendations, and mobile application support can be added to make the system more effective and widely usable.

Overall, the proposed system provides a scalable and practical solution for improving student performance analysis using modern AI technologies.

VII. ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Megha Deshmukh for her valuable guidance, continuous support, and encouragement throughout the development of this project. Her insights and suggestions have greatly contributed to the successful completion of this research work.

We also thank the Department of Computer Engineering, AGPIT, Solapur, for providing the necessary resources and environment to carry out this project. We are grateful to our faculty members and peers for their support and motivation during the project development.

VIII. 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*, vol. 40, no. 6, pp. 601–618, 2010.
- [2] 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.

[3] D. Kabakchieva, "Predicting Student Performance by Using Data Mining Methods for Classification," *Cybernetics and Information Technologies*, vol. 13, no. 1, pp. 61–72, 2013.

[4] P. Cortez and A. M. G. Silva, "Using Data Mining to Predict Secondary School Student Performance," in *Proceedings of Future Business Technology Conference*, 2008.

[5] R. S. J. d. Baker and K. Yacef, "The State of Educational Data Mining in 2009: A Review and Future Visions," *Journal of Educational Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.

[6] F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[7] Streamlit Inc., "Streamlit: The Fastest Way to Build Data Apps," 2019. [Online]. Available: <https://streamlit.io/>

[8] Supabase Inc., "Supabase: Open Source Firebase Alternative," 2023. [Online]. Available: <https://supabase.com/>