STUDENT PERFORMANCE PREDICTION

Suresh Kampe [1], Chenagalla Tejasree[2] Dasari Bhavani[3] Dama Yashodha[4] Alge Lavanya[5]
[1]Assistant Professor Department *of Computer Science and Engineering, Joginpally B.R. Engineering College, Telangana, India.*

[2]*B.Tech Student, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Telangana, India.* tejasreechenagalla11@gmail.com

[3]*B.Tech Student, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Telangana, India.* dasaribhavani796@gmail.com
[4]*B.Tech Student, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Telangana, India.* yashodadama@gmail.com
[5]*B.Tech Student, Department of Computer Science and Engineering, Joginpally B.R. Engineering College, Telangana, India.* algelavanya@gmail.com

ABSTRACT: this project aims to predict student performance in advance using Machine Learning and provide personalized recommendations for improvement. Student data such as grades, attendance, study hours, internet usage, and stress levels is collected and pre-processed. Models like Random Forest and XGBoost are trained to predict outcomes such as grades or dropout risk. The system's unique feature is its recommendation engine, which suggests actions like increasing study time or improving attendance. A user-friendly dashboard displays predictions and advice, enabling teachers, parents, and students to take timely action, reduce failure rates, and enhance overall academic performance. Educational institutions generate large volumes of student-related data, but this data is often underutilized. By applying data mining techniques and machine learning algorithms such as Decision Trees, Random Forest, Naïve Bayes, Logistic Regression, or Support Vector Machines, meaningful patterns can be extracted to forecast student outcomes. These predictions help in identifying students who are at risk of poor performance at an early stage. The system collects, pre-processes, and analyzes historical student data to build predictive models. The model is trained using labeled datasets and evaluated using performance metrics such as accuracy, precision, recall, and F1-score. Based on the trained model, the system predicts whether a student is likely to pass, fail, or achieve a certain grade. This project aims to support teachers, administrators, and academic planners in making informed decisions, enabling timely interventions, personalized guidance, and improved learning strategies. Ultimately, Student Performance Prediction contributes to enhancing academic success rates and reducing dropout rates through data-driven decision-making.

Keywords Student Performance Prediction, Machine Learning, Data Mining, Random Forest, Education Analytics, Predictive Modeling

1. INTRODUCTION

Student Performance Prediction is an emerging application of data mining and machine learning in the field of education. It focuses on analyzing students' academic data to predict their future performance, such as grades, pass/fail outcomes, or overall achievement levels. With the rapid growth of educational institutions and digital learning platforms, a large amount of studentrelated data is generated. This data can be effectively utilized to identify patterns and trends that influence academic success.

The primary objective of student performance prediction is to detect students who may be at risk of poor academic performance at an early stage. By predicting outcomes in advance, educators and institutions can take preventive measures such as providing additional support, counseling, personalized learning plans, or remedial classes. This helps improve overall academic results and reduces dropout rates.

Various factors influence student performance, including attendance, previous academic records, socio-economic background, study habits, participation in activities, and internal assessment scores. Machine learning algorithms such as Decision Trees, Random Forest, Naïve Bayes, Support Vector Machines, and Neural Networks are commonly used to build predictive models. These models are trained using historical data and then tested to evaluate their accuracy and reliability. Student performance prediction systems not only benefit students but also assist teachers, administrators, and educational policymakers in making data-driven decisions. By identifying key performance indicators, institutions can enhance teaching strategies, optimize resource allocation, and improve curriculum planning.

2. Existing System:

The existing system for student performance evaluation relies on manual assessment methods and focuses mainly on after-exam analysis. This approach is time-consuming and reactive, as it evaluates student outcomes only after results are available. Additionally, it does not provide any early warning mechanisms to identify students who may be at risk of poor performance. The absence of analytical tools further limits its effectiveness, as it lacks data-driven insights to support informed decision-making and timely interventions.

3. Proposed System:

The proposed system introduces an automated prediction approach to evaluate student performance before exams take place. By leveraging predictive techniques, it enables early identification of at-risk students, allowing educators to intervene proactively. This system is designed to be faster and more accurate compared to traditional methods. Moreover, it incorporates data-driven insights and analytics, which help in understanding student behavior and performance trends, ultimately leading to improved academic outcomes.

4. Literature Review

Student performance prediction has gained significant attention in the field of Educational Data Mining (EDM) and Machine Learning due to its potential to improve academic outcomes through early intervention.

Educational Data Mining plays a crucial role in extracting meaningful insights from educational datasets. One of the earliest works by D. Kabakchieva [1] applied classification algorithms such as Decision Trees and Naïve Bayes to predict student outcomes. The study demonstrated that data mining techniques can effectively classify students based on academic performance, though it primarily relied on limited academic features.

Similarly, A. Altabrawee et al. [2] explored multiple machine learning techniques including Random Forest and Support Vector Machines. Their work highlighted that ensemble methods significantly improve prediction accuracy compared to traditional statistical approaches.

A comprehensive survey by Cristóbal Romero and Sebastián Ventura [3] provided an extensive overview of EDM techniques applied in education. The study emphasized the importance of data preprocessing, feature selection, and algorithm selection in improving prediction performance. It also identified gaps such as lack of real-time systems and limited personalization.

Further research by H. Huang and N. Fang [4] demonstrated the effectiveness of machine learning models in predicting student

performance using academic and behavioral data. Their study confirmed that combining multiple features leads to higher predictive accuracy.

Recent advancements focus on more sophisticated models. W. Zhou et al. [5] proposed a prediction model based on Interval Belief Rule Base, which improves decision-making under uncertainty. Similarly, M. Ahmed et al. [6] compared various machine learning algorithms and concluded that ensemble models such as Random Forest outperform single classifiers in most scenarios.

In addition, A. A. Elrahman et al. [7] introduced a predictive model using student interaction data from eTextbooks, highlighting the importance of behavioral data in improving prediction accuracy. This approach moves beyond traditional academic metrics and incorporates learning patterns.

Recent studies also explore advanced machine learning techniques. Research on Random Forest Regressor [8] and Gradient Boosting methods [9] shows improved performance in handling large and complex datasets. These models are capable of capturing nonlinear relationships between variables and provide higher accuracy compared to traditional approaches.

Despite these advancements, several challenges remain:

- Limited use of real-time data
- Lack of integrated recommendation systems
- Insufficient focus on personalized learning

The proposed system addresses these gaps by combining multiple data sources, applying ensemble learning techniques, and providing actionable recommendations through an interactive dashboard.

5. System architecture

The system architecture of a Student Performance Prediction system represents the structured framework that defines how data flows through different components to analyze and predict students' academic performance. The architecture is designed to ensure efficient data handling, accurate prediction, and meaningful output for decision-making. The architecture begins with data acquisition, where student-related data such as academic scores, attendance records, demographic details, and learning behavior are collected from institutional databases or learning platforms. This data serves as the primary input to the system. Once the data is collected, it moves to the preprocessing stage. In this stage, the raw data is cleaned and transformed to improve its quality. Preprocessing includes handling missing values, removing inconsistencies, and converting data into a suitable format. This step is essential to ensure reliable predictions.

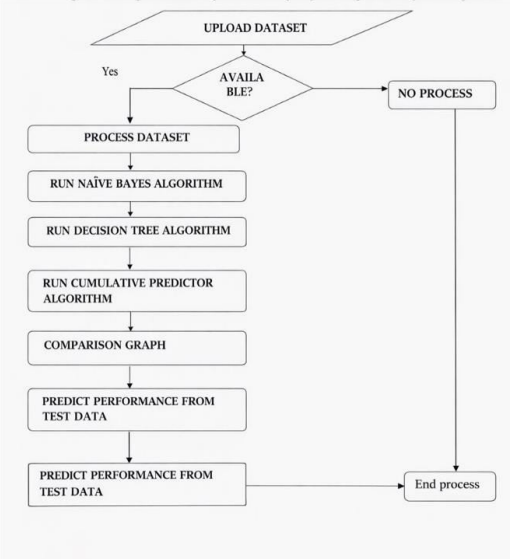


Fig.1: System architecture

After preprocessing, the system performs feature extraction and selection. Relevant attributes that significantly influence student performance are identified and selected. This helps reduce data complexity and improves the efficiency and accuracy of the prediction

model. The core component of the architecture is the prediction model. Machine learning or statistical techniques are applied to the selected features to learn patterns from historical data. The trained model is then used to predict future student performance, such as grades, pass/fail status, or risk level. 28 The predicted results are evaluated to measure the effectiveness of the system. Evaluation ensures that the model produces accurate and consistent predictions and supports further improvement of the system. Finally, the output is presented to users through the application layer. The system provides predicted performance reports that help educators and administrators identify students who need academic support. All data and results are securely stored in the database for future reference. Overall, the system architecture integrates data collection, processing, prediction, and presentation in a structured manner to support early intervention and improve student academic outcomes.

6. SYSTEM IMPLEMENTATION

6.1 Modules

The system is divided into different functional modules to ensure proper organization and smooth execution.

1. User Authentication Module

Purpose:

Manages login and access control.

Functions:

- User registration (Admin/Faculty)
- Login & Logout
- Role-based access control
- Password management

2. Data Collection Module

Purpose:

Collects student-related academic and personal data.

Inputs:

- Attendance
- Internal marks
- Assignment scores
- Previous semester results
- Study hours
- Demographic details (if required)

Functions:

- Manual data entry
- Bulk upload (CSV/Excel)
- Data validation

3. Data Preprocessing Module

Purpose:

Prepares raw data for analysis.

Functions:

- Handling missing values
- Removing duplicates
- Encoding categorical data
- Data normalization/scaling
- Feature selection

4. Database Management Module

Purpose:

Stores and manages student records.

Functions:

- Store student profiles
- Store academic records
- Store prediction results
- Data retrieval and updates

5. Prediction Module (Core Module)

Purpose:

Predicts student performance using machine learning algorithms.

Functions:

- Model training
- Model testing
- Running prediction on new student data
- Generating output (Pass/Fail, Grade, Score)

Algorithms Used:

- Logistic Regression
- Decision Tree
- Random Forest
- SVM
- KNN

6. Performance Evaluation Module

Purpose:

Evaluates accuracy of the prediction model.

Functions:

- Accuracy calculation
- Confusion matrix
- Precision, Recall, F1-score
- Error metrics (MAE, MSE for regression)

7. Report Generation Module

Purpose:

Generates reports for decision-making.

Outputs:

- Individual student performance report
- At-risk student list
- Class performance summary
- Graphical analysis (Bar charts, Pie charts)

8. Visualization Module

Purpose:

Displays data in graphical format.

Visuals:

- Attendance vs Marks graph
- Performance trend chart
- Subject-wise performance analysis

9. Recommendation Module (Optional Advanced Module)

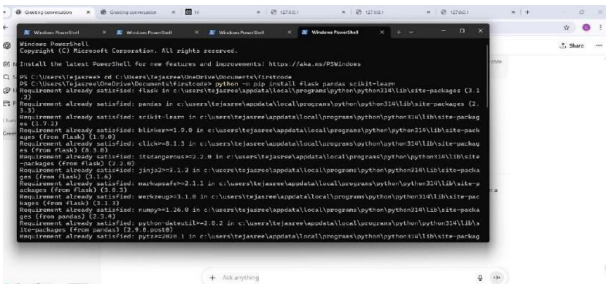
Purpose:

Provides suggestions for improvement.

Examples:

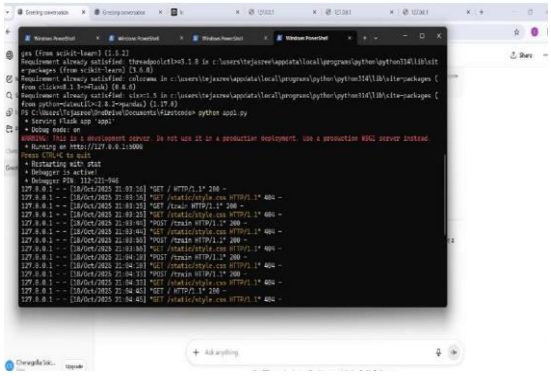
- Suggest extra classes
- Recommend study hours
- Identify weak subjects
- Alert faculty for intervention

7. RESULTS

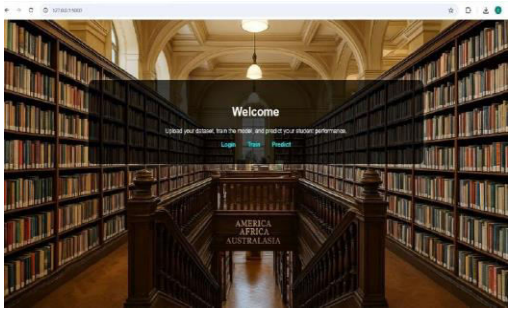


Copy the project path and navigate to it:cmd type cd

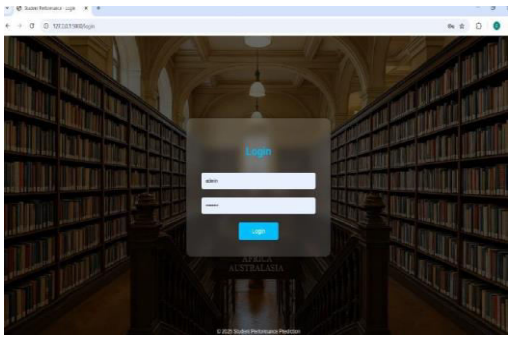
- "C:\Users\Tejasree\OneDrive\Documents\firstcode
- Python -m pip install flask pandas scikit-learn && python app.py
- Open http://127.0.0.1:5000 in your browser



- The image shows a Windows PowerShell window where you ran python -m pip install flask pandas scikit-learn.
- All the listed packages (flask, pandas, scikit-learn) and their dependencies are already installed, so pip reports "Requirement already satisfied" for each of them.

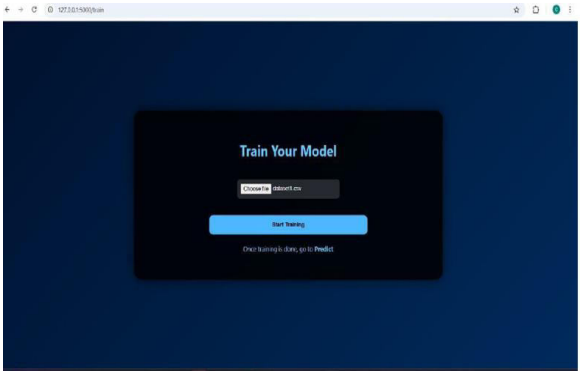


- The Flask app is running successfully.
- The home page at http://127.0.0.1:5000 displays a Welcome screen with options to Login, Train, and Predict for student performance analysis.



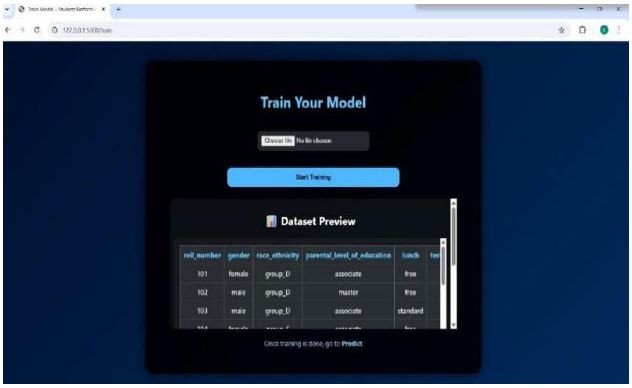
The image shows a web interface for training a machine learning model with the following elements:

- Functionality: The page allows users to "Train Your Model" by choosing a file (currently showing dataset1.csv) and clicking the "Start Training" button.
- Workflow: After training is complete, the user is instructed to go to the "Predict" section to use the trained model.
- Environment: The application is running locally on 127.0.0.1:5000/train, indicating a development server.



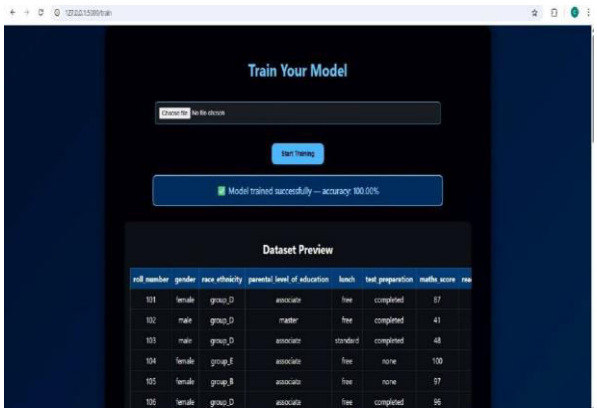
The image shows a web interface for training a machine learning model with the following elements

- File selection: A "Choose file" button is used to upload a dataset (here dataset1.csv is selected).
- Action: The blue "Start Training" button initiates the model training process on the uploaded file.
- Next step: After training completes, the user is instructed to go to Predict to use the model for predictions.

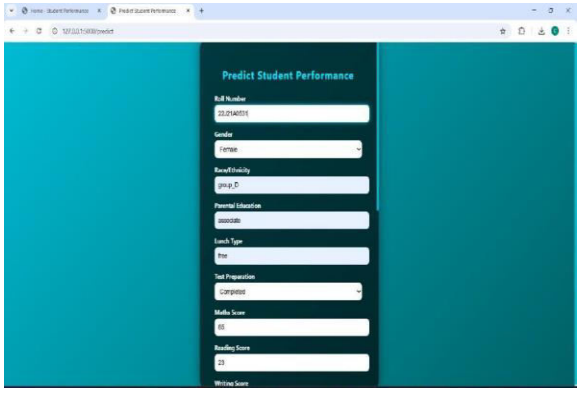


• The model has been trained successfully with an accuracy of 100.00%.

• The dataset preview shows columns: roll_number, gender, race_ethnicity,level_of_education,lunch,test_preparation,maths_score, and reading_score.



• The image shows a Dataset Preview table with columns: roll_number, gender, ethnicity, parental education, lunch type, test preparation status, and corresponding maths scores.

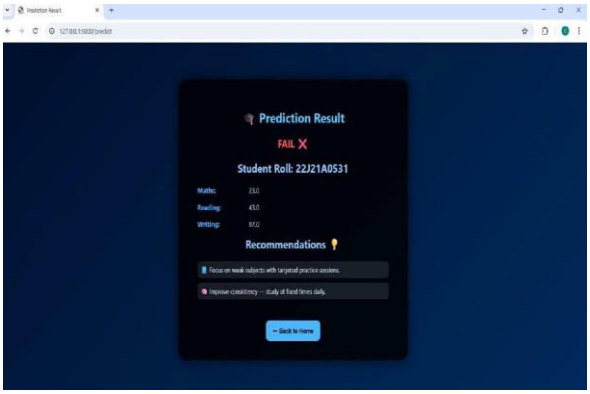


After training is complete, the user is instructed to go to the "Predict" section to use the trained model.

• Predicted Student Performance: Pass.

• Overall Grade:

• The system uses the performance database to forecast that the student will pass with a B grade, suggesting the earlier weak areas have been improved through targeted practice and consistent study.



- Prediction Result: FAIL for Student Roll 22J21A0531.
- Scores: Maths = 23.0, Reading = 43.0, Writing = 87.0.
- Finally it predicts the future of students using the database of performance by analyzing subject scores and providing recommendations like focusing on weak subjects and improving daily study consistency to turn the result into a pass.

8. CONCLUSION AND FUTURE ENHANCEMENT

Student Performance Prediction is an effective approach that uses data analysis and machine learning techniques to forecast students' academic outcomes based on various influencing factors such as attendance, internal marks, study habits, socio-economic background, and participation in activities. The project demonstrates how educational data can be transformed into meaningful insights that help institutions identify at-risk students at an early stage. By applying suitable algorithms like Decision Trees, Random Forest, Logistic Regression, or Neural Networks, accurate predictions can be generated to support timely intervention strategies. This system improves academic planning, enhances student support services, and assists teachers in making data-driven decisions. It also promotes personalized learning by understanding individual student strengths and weaknesses. Although challenges such as data quality, privacy concerns, and model accuracy exist, proper data preprocessing, validation techniques, and

ethical data handling can minimize these issues. In conclusion, Student Performance Prediction is a valuable tool in modern education systems. It not only enhances institutional effectiveness but also contributes to improving student success rates, reducing dropout rates, and creating a smarter, data-driven academic environment.

Future Enhancement

To improve the efficiency and impact of the Student Performance Prediction system, the following future enhancements can be implemented:

1. Integration of Advanced Machine Learning Models
   - Implement Deep Learning models such as Artificial Neural Networks (ANN) and LSTM for higher prediction accuracy.
   - Use ensemble techniques (Random Forest, Gradient Boosting) to improve model performance.

2. Real-Time Data Processing
   - Integrate the system with Learning Management Systems (LMS) to collect real-time attendance, assignment, and exam data.
   - Enable continuous performance monitoring instead of semester-based prediction.

3. Personalized Recommendation System
   - Provide customized study plans and learning resources based on individual student weaknesses.
   - Suggest remedial classes or mentoring support automatically.

4. Mobile Application Development
   - Develop a mobile app for students, teachers, and parents to track academic progress easily.
   - Send alerts and notifications for low performance or attendance shortage.

5. Data Visualization Dashboard
   - Create interactive dashboards for administrators and faculty to analyze trends and performance reports.
   - Include graphical reports such as bar charts, line graphs, and heat maps.

6. Inclusion of Behavioural and Psychological Factors
   - Incorporate stress levels, participation, motivation, and extracurricular involvement to improve prediction accuracy.

7. Early Dropout Detection System
   - Enhance the model to predict dropout risks at an early stage.
   - Provide intervention strategies for high-risk students.

8. Cloud-Based Deployment

Deploy the system on cloud platforms for scalability, security, and easy access across institutions.

9. Improved Data Security and Privacy
   - Implement encryption and role-based access control.
   - Ensure compliance with educational data protection policies.

10. Multi-Institutional Dataset Integration
   - Train models using data from multiple colleges/schools to improve generalization and robustness.

9. REFERENCES

[1] D. Kabakchieva, "Student Performance Prediction by Using Data Mining Classification

Algorithms," International Journal of Computer Science and Management Research, vol. 1, no. 4, pp. 686–690, 2012. https://ieeexplore.ieee.org/document/6974687

[2] A. Altabrawee et al., "Student Performance Prediction Using Machine Learning Techniques," IEEE, 2019. https://ieeexplore.ieee.org/document/8756583

[3] C. Romero and S. Ventura, "Educational Data Mining: A Survey and a Data Mining-Based Analysis of Recent Works," Computers & Education, vol. 68, pp. 1–17, 2013. https://www.sciencedirect.com/science/article/pii/S0360131513000298

[4] H. Huang and N. Fang, "Student Performance Prediction Using Machine Learning," IEEE, 2018. https://ieeexplore.ieee.org/document/8460031

[5] W. Zhou et al., "A Student Academic Performance Prediction Model Based on the Interval Belief Rule Base," Scientific Reports, vol. 15, 2025. https://www.nature.com/articles/s41598-025-92324-x 0

[6] M. Ahmed et al., "Student Performance Prediction Using Machine Learning Algorithms," Education and Information Technologies, 2024. https://link.springer.com/article/10.1007/s10639-024-12619-w

[7] A. A. Elrahman et al., "A Predictive Model for Student Performance in Classrooms Using Student Interactions With an eTextbook," arXiv preprint, 2022. https://arxiv.org/abs/2203.03713 1

[8] "Student Performance Prediction Using Random Forest Regressor," arXiv preprint, 2026. https://arxiv.org/abs/2603.00608

[9] "Student Performance Prediction Using Gradient Boosting Techniques," ScienceDirect, 2026. https://www.sciencedirect.com/science/article/pii/S155018762600003X