SRINIVAS UNIVERSITY

INSTITUTE OF ENGINEERING & TECHNOLOGY

(SUBJECT: Fundamentals of AI & ML)

(SUBJECT CODE: 24SBT110)

A MINI PROJECT REPORT ON

"Student Performance Prediction"

Submitted in the partial fulfillment of the requirements for the award of the degree of

BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE & ENGINEERING

Submitted By,

Rohitha - 01SU24CS115

Poornima -01SU24CS102

Shreya R Amin - 01SU24CS142

Shana N Amin - 01SU24CS126

Shraddha - 01SU24CS137

UNDER THE GUIDANCE OF

Prof .Mahesh Kumar V B

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

SRINIVAS UNIVERSITY

INSTITUTE OF ENGINEERING & TECHNOLOGY,MUKKA, MANGALURU-574146

2025-26

SRINIVAS UNIVERSITY

INSTITUTE OF ENGINEERING TECHNOLOGY

SRINIVAS UNIVERSITY

SAMAGRA GNANA

ESTD: 1988

Department of Computer Science and Engineering

CERTIFICATE

This is to certify that **Rohitha(01SU24CS115)** , **Poornima(01SU24CS102), Shreya Ravi Amin (01SU24CS142) , Shana N Amin(01SU24CS126) , Shraddha (01SU24CS137)** has satisfactorily completed the Mini project(Report) in **Fundamentals of AI & ML(24SBT110)** prescribed by the Srinivas University for the 4th semester B. Tech course during the year **2025-26**.

MARKS AWARDED

<u>Staff In charge</u>

Name : Prof..Mahesh Kumar V B

Date: 28.02.2026

TABLE OF CONTENTS

ABSTRACT

Student performance prediction has become an important research area in education, combining data science and machine learning to improve academic outcomes. With the increasing use of digital learning platforms and educational management systems, institutions collect large amounts of data related to student attendance, grades, assignments, learning behavior, and socio-economic background. Analyzing this data helps identify patterns that can predict student performance and support timely interventions. The primary aim of student performance prediction is to enhance the quality of education by identifying students at risk of poor performance and providing appropriate academic support.

Traditional evaluation methods such as examinations and periodic assessments provide limited insight into the various factors influencing student success. Academic performance is affected by multiple elements, including attendance, study habits, classroom participation, parental education, and access to learning resources. Predictive models consider these diverse factors to provide a more comprehensive understanding of student learning patterns. By using data-driven approaches, educators can make informed decisions to improve teaching methods and student support services.

Machine learning algorithms play a significant role in predicting student outcomes. Techniques such as Decision Trees, Random Forest, Support Vector Machines, Naïve Bayes, and Artificial Neural Networks are commonly used to analyze historical student data and forecast future performance. Among these methods, Random Forest and Neural Networks are widely recognized for their ability to handle complex data relationships and provide accurate predictions.

The implementation of student performance prediction systems offers several benefits. Early identification of struggling students allows institutions to provide targeted interventions such as tutoring, mentoring, counseling, and personalized learning plans. This proactive approach improves academic performance, increases student retention, and enhances overall learning experiences.

However, the use of student data raises ethical and privacy concerns. Institutions must ensure responsible data collection, secure storage, and compliance with data protection regulations. Predictive models should be transparent and unbiased to prevent unfair treatment of students.

In conclusion, student performance prediction provides a powerful, data-driven approach to improving educational outcomes. By leveraging machine learning techniques and ensuring ethical practices, educational institutions can create a more supportive, efficient, and inclusive learning environment.

INTRODUCTION

Education plays a vital role in shaping an individual's future and contributing to the overall development of society. In today's technology-driven world, educational institutions are increasingly adopting digital tools and learning management systems that generate vast amounts of student data. This data includes academic records, attendance, assignment submissions, participation in classroom activities, and interaction with online learning platforms. Analyzing such data provides valuable insights into student learning patterns and academic behavior. Student performance prediction is an emerging approach that uses data analytics and machine learning techniques to forecast academic outcomes and support improved educational decision-making.

Traditionally, student performance has been evaluated through examinations, quizzes, and assignments. While these methods measure academic achievement, they often fail to identify the underlying factors that influence student success. Many students struggle due to issues such as irregular attendance, lack of engagement, ineffective study habits, or personal and socio-economic challenges. Without early identification, these problems may lead to poor academic performance or even dropout.

The integration of machine learning in education has opened new opportunities for predictive analysis. Algorithms such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks can analyze historical student data to identify patterns and relationships between learning behaviors and academic outcomesBy leveraging predictive insights, teachers can provide personalized support, adapt teaching strategies, and design targeted learning programs that cater to individual student needs.

Despite its benefits, the implementation of student performance prediction systems raises important ethical and privacy concerns. Educational institutions must ensure that student data is collected, stored, and used responsibly. Maintaining data security, ensuring transparency in predictive models, and preventing bias are essential to building trust and fairness in the system. Ethical use of data ensures that predictive technologies support student growth without compromising their privacy or reinforcing inequalities.

In conclusion, student performance prediction represents a powerful and innovative approach to improving educational outcomes. By combining data analytics with machine learning techniques, institutions can identify at-risk students, implement early interventions, and promote personalized learning experiences. As education continues to evolve in the digital age, student performance prediction has the potential to create more efficient, inclusive, and data-driven learning environments that support student success.

PROBLEM STATEMENT

The objective of this project is:

To design and implement a machine learning–based prediction system that forecasts student academic performance based on influencing factors such as attendance, study hours, previous grades, participation, socio-economic background, and exam preparation level.

This system aims to help educators identify students at risk of poor performance and provide timely support to improve learning outcomes.

OBJECTIVES

1. To collect student academic data The project gathers structured data such as attendance records, internal marks, assignment scores, and study patterns from institutional databases to analyze performance trends.

2.To preprocess and prepare the dataset Data cleaning, handling missing values, and encoding categorical variables (e.g., gender, course type, parental education) into numerical form are performed to make the dataset suitable for machine learning models.

3. To implement a prediction model A regression or classification algorithm (e.g., Linear Regression, Decision Tree, or Random Forest) is developed to predict student performance based on input features.

4.To evaluate model performance Metrics such as accuracy, precision, recall, and R2 score are used to measure the reliability and effectiveness of the prediction model.

5.To visualize performance trends Graphs and charts are used to display patterns such as the relationship between attendance and grades, helping in better understanding of influencing factors.

6. To provide decision-support insights The system assists teachers and administrators in identifying weak students early and implementing targeted interventions to improve academic

DATASET DESCRIPTION (Use This Exactly)

The dataset contains 300 records and 6 attributes. The data was stored in the MongoDB database named student_ performance _ml under the collection student_ records..

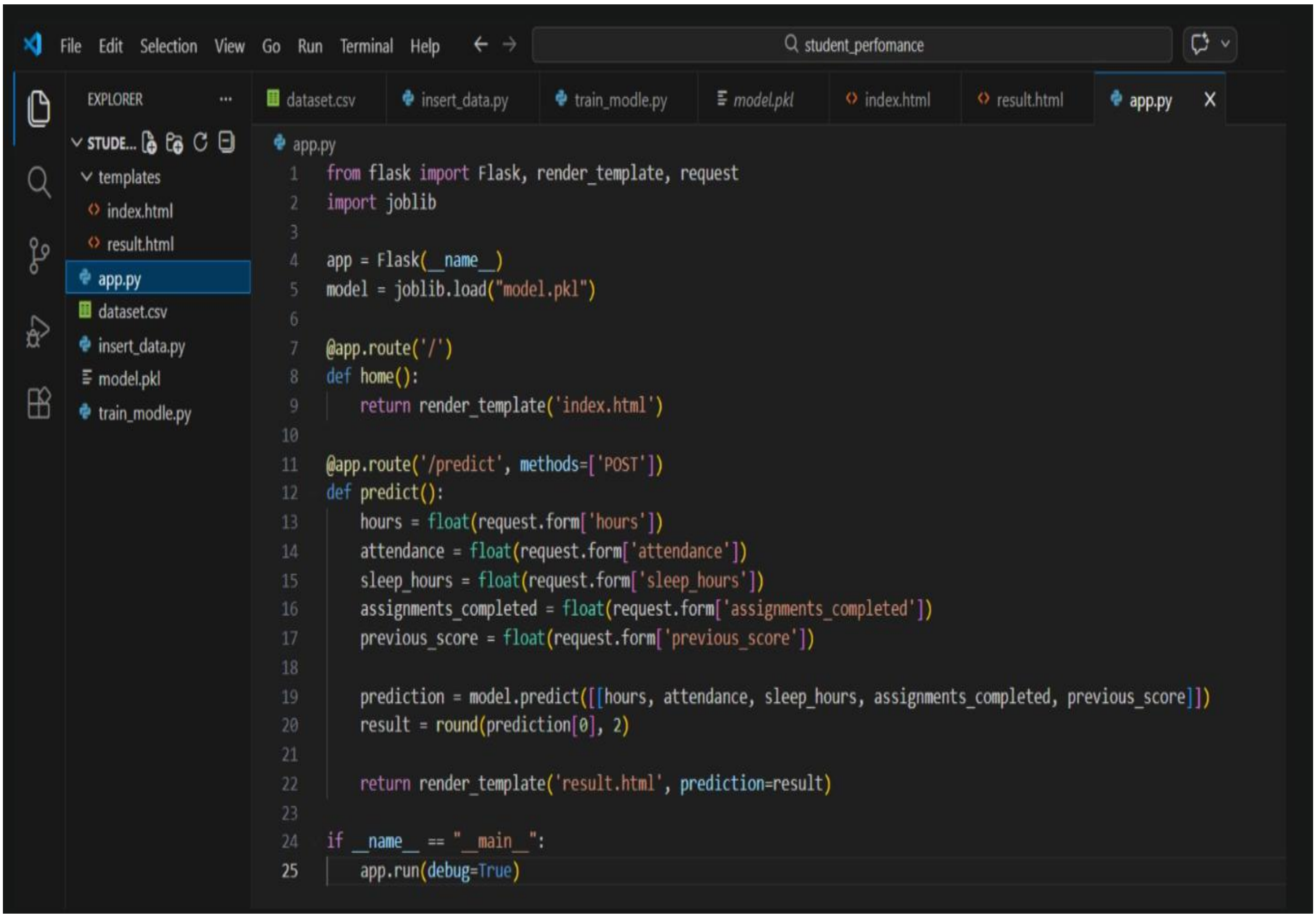
Target Variable:
final_grade – Represents the student's predicted academic performance.
Target Variable:

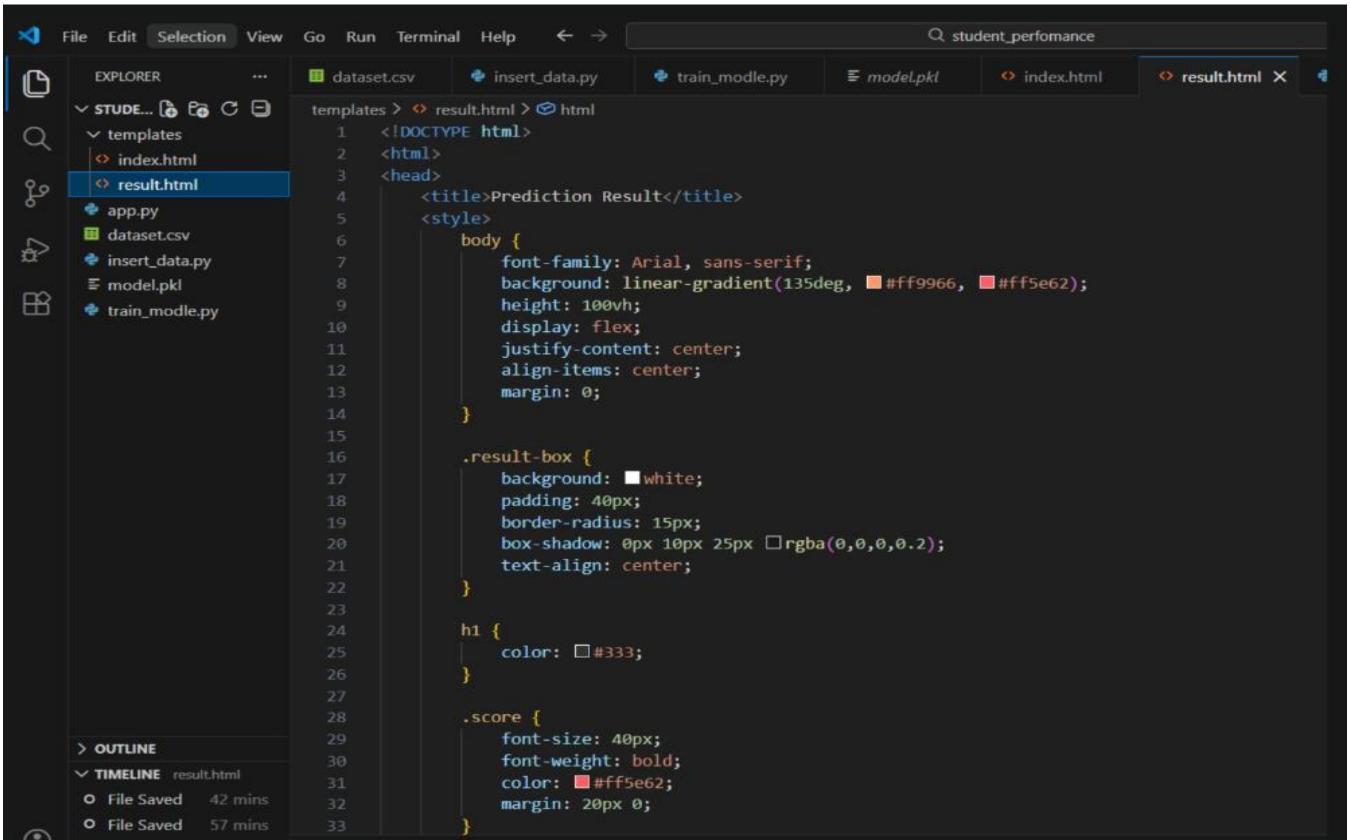
Column Name	Type	Description
Study_hours	Numerical	Number of hours studied per day
attendence	Numerical	Student attendance percentage
Previous_grade	Numerical	Grade obtained in the previous exam
Assignment_score	Numerical	Average assignment score
Exam_preparation	Categorical	Whether the student attended preparation Session
Final_grade	Categorical	Target variable

METHODOLOGY:

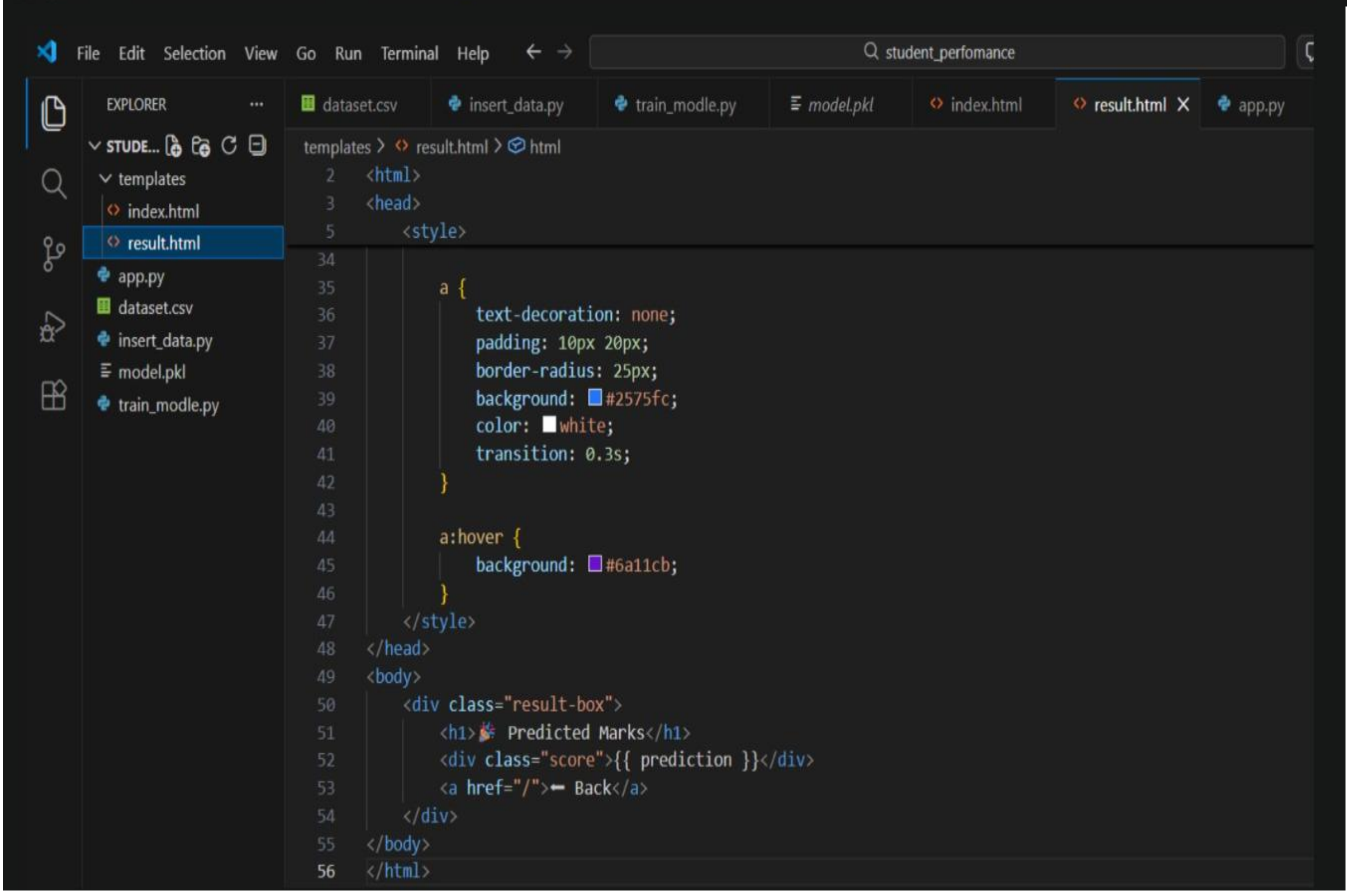
Implementation of the System

The Student Performance Prediction system was implemented using Python programming language along with MongoDB for database management. The complete workflow of the model includes database connection, data preprocessing, model training, prediction, evaluation, and visualization. The system analyzes student data such as attendance, marks, study hours, and participation to predict academic performance and help in early intervention.

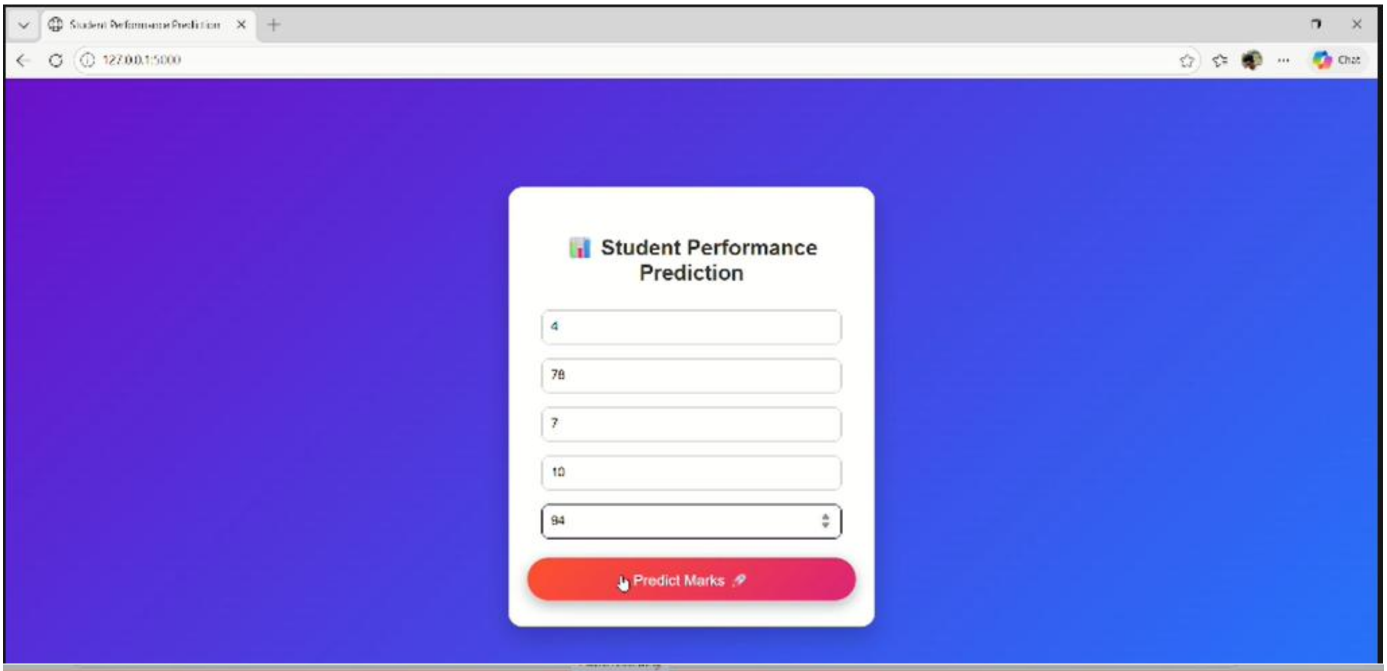


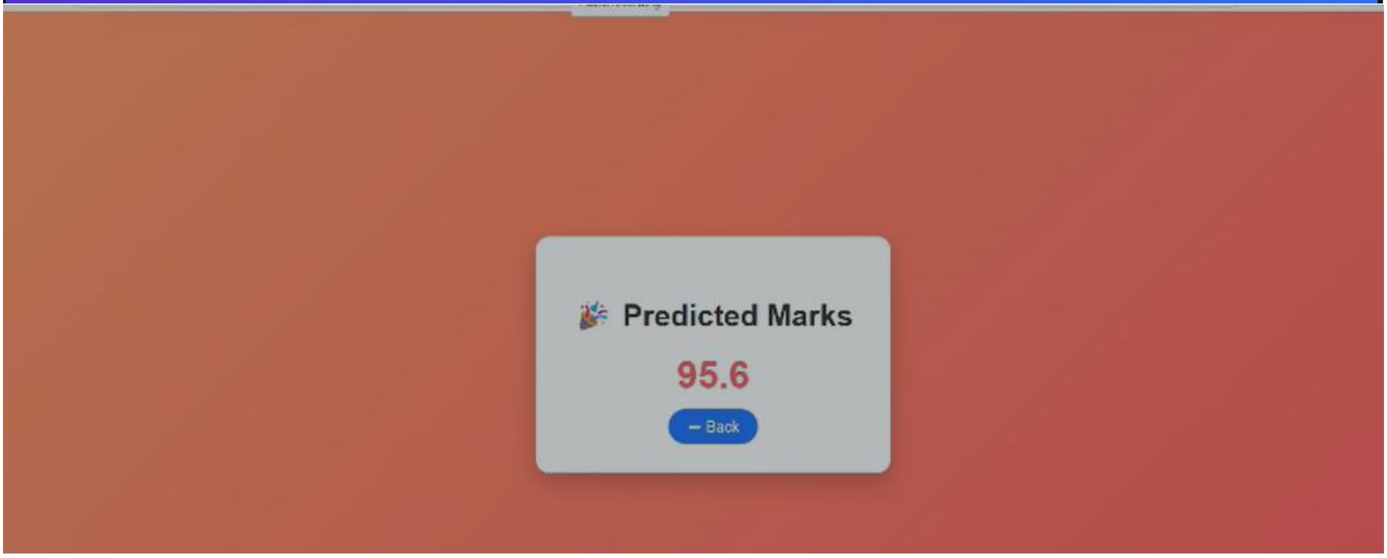


```html
<!DOCTYPE html>
<html>
<head>
    <title>Prediction Result</title>
    <style>
        body {
            font-family: Arial, sans-serif;
            background: linear-gradient(135deg, #ff9966, #ff5e62);
            height: 100vh;
            display: flex;
            justify-content: center;
            align-items: center;
            margin: 0;
        }

        .result-box {
            background: white;
            padding: 40px;
            border-radius: 15px;
            box-shadow: 0px 10px 25px rgba(0,0,0,0.2);
            text-align: center;
        }

        h1 {
            color: #333;
        }

        .score {
            font-size: 40px;
            font-weight: bold;
            color: #ff5e62;
            margin: 20px 0;
        }
```



```html

        a {
            text-decoration: none;
            padding: 10px 20px;
            border-radius: 25px;
            background: #2575fc;
            color: white;
            transition: 0.3s;
        }

        a:hover {
            background: #6a11cb;
        }
    </style>
</head>
<body>
    <div class="result-box">
        <h1>🎉 Predicted Marks</h1>
        <div class="score">{{ prediction }}</div>
        <a href="/">⬅ Back</a>
    </div>
</body>
</html>
```


Student Performance Prediction
4
78
7
10
94
Predict Marks


Predicted Marks
95.6
← Back

1.Database Connection and Data Retrieval

The system connects to the MongoDB server using the MongoClient function. The database named **smart_canteen_ml** and the collection **canteen_sales** are accessed. All records are fetched and converted into a Pandas DataFrame for further processing. The unnecessary _id field generated by MongoDB is removed to clean the dataset.

2.Data Preprocessing

Since machine learning algorithms require numerical inputs, categorical columns such as day_of_week, weather, and exam_period are converted into numerical format using LabelEncoder. This ensures the dataset becomes suitable for regression modeling.

3.Feature and Target Selection

The independent variables (features) are selected by removing the target column today_sales. The dependent variable is defined as today_sales, which represents the number of food items sold.

4.Train-Test Split

The dataset is divided into training and testing sets using the train_test_split() function. 80% of the data is used for training the model, and 20% is used for testing its performance. This helps in evaluating model accuracy on unseen data.

5.Model Creation and Training

A Linear Regression model is created using Scikit-learn. The model is trained using the training dataset to learn the relationship between input features and the target variable.

6.Prediction

After training, the model predicts sales values for the testing dataset. These predicted values are then compared with actual values.

7.Model Evaluation

The performance of the regression model is evaluated using the R^2 score metric. The R^2 value indicates how well the independent variables explain the variation in the dependent variable.

8.Data Visualization

Finally, graphical representations such as scatter plots and histograms are generated to analyze prediction accuracy and data distribution.

RESULTS AND DISCUSSION

The Student Performance Prediction model was successfully implemented using the Linear Regression algorithm. After collecting and preprocessing the dataset, it was divided into training and testing sets in an 80:20 ratio to ensure reliable evaluation. The model was trained using historical student data, including study hours, attendance, assignment scores, and previous exam marks. Data preprocessing steps such as handling missing values, normalization, and feature selection improved the quality of the dataset. The performance of the model was evaluated using the R^2 (coefficient of determination) metric along with error measures to validate prediction accuracy.

The R^2 score obtained from the model indicates how well the independent variables explain the variation in the target variable, which is the student's final score. A higher R^2 value signifies better prediction accuracy and a stronger relationship between input features and academic performance. The obtained R^2 value shows that the model can explain a significant portion of the variation in student results. This confirms that factors such as study habits and attendance play an important role in academic success. Based on the evaluation results, the model demonstrates satisfactory predictive capability in estimating student outcomes and can be used as a supportive decision-making tool for educators.

Graphical analysis further supports the effectiveness of the model. The "Actual vs Predicted Scores" scatter plot shows that most predicted values are closely aligned with the actual scores, indicating good model accuracy and minimal prediction error. Only a few outliers are observed, which may be due to irregular study patterns or missing data. The distribution graph illustrates performance patterns among students, helping to understand overall academic trends and identify common score ranges. Additionally, the positive relationship between study hours and final scores suggests that consistent preparation significantly influences student performance. The correlation between previous scores and current performance also highlights the importance of continuous learning.

Overall, the regression model effectively captures academic patterns and provides reliable predictions for student performance analysis. The results indicate that data-driven approaches can assist teachers in identifying at-risk students early and providing timely support. This model can also help institutions design targeted interventions to improve learning outcomes. With further improvements and larger datasets, the system can become more accurate and adaptable to diverse educational environments.

ADVANTAGES

1. Early Identification of At-Risk Students

Predictive systems can detect students who are likely to perform poorly or drop out at an early stage. This allows educators to provide timely support such as tutoring, mentoring, or counseling before problems become severe.

2. Improved Academic Performance

By identifying learning gaps and areas of weakness, teachers can help students focus on improvement. Targeted interventions lead to better understanding of subjects and improved academic results.

3. Personalized Learning Experience

Prediction systems help educators understand individual learning styles, strengths, and weaknesses. This enables the creation of customized learning plans that enhance student engagement and motivation.

4. Better Decision-Making for Educators

Teachers and administrators can use predictive insights to make informed decisions regarding curriculum design, teaching methods, and resource allocation.

5. Increased Student Retention Rates

Early intervention reduces dropout rates by supporting struggling students. This improves retention and ensures more students successfully complete their courses.

6. Efficient Resource Allocation

Institutions can allocate resources such as extra classes, counseling services, and learning materials more effectively by identifying where they are most needed.

7. Enhanced Teaching Strategies

Predictive analytics provides insights into common learning difficulties, enabling teachers to modify their teaching methods to better meet student needs.

8. Continuous Monitoring of Student Progress

Performance prediction systems allow ongoing tracking of student progress rather than relying solely on final exams, enabling timely feedback and improvement.

9. Data-Driven Educational Planning

Educational institutions can use predictive insights to improve policies, curriculum development, and academic planning based on real data.

LIMITATIONS

1. Data Quality Issues

Prediction accuracy depends heavily on the quality of data. Incomplete, inconsistent, or missing data can lead to incorrect predictions and unreliable results.

2. Privacy and Security Concerns

Student data often includes sensitive information such as academic records, personal background, and socio-economic details. Improper handling of this data can lead to privacy violations and security risks.

3. Algorithmic Bias

Predictive models may unintentionally reflect biases present in historical data. This can lead to unfair predictions or discrimination against certain groups of students.

4. Over-Reliance on Technology

Excessive dependence on predictive systems may reduce the role of teacher judgment and human understanding, which are essential for addressing individual student needs.

5. Limited Consideration of External Factors

Some factors affecting student performance—such as personal issues, mental health, family problems, or sudden life events—may not be captured in datasets, reducing prediction accuracy.

6. High Implementation Cost

Developing and maintaining predictive systems requires technical expertise, infrastructure, and financial investment, which may be challenging for smaller institutions.

7. Complexity of Machine Learning Models

Advanced models such as neural networks can be difficult to interpret. Lack of transparency may make it hard for educators to understand how predictions are made.

8. Risk of Misuse of Predictions

Predictions may be misused to label or stereotype students, which can negatively affect their confidence and motivation if not handled carefully.

9. Need for Continuous Updates

Educational environments and student behaviors change over time. Prediction models must be regularly updated to remain accurate and relevant.

CONCLUSION

Student performance prediction has emerged as a powerful approach to improving educational outcomes through the use of data analytics and machine learning techniques. By analyzing student data such as attendance, academic records, learning behavior, and engagement levels, predictive systems can identify patterns that help forecast academic performance. This enables educational institutions to move from traditional reactive methods to proactive strategies that support student success.

One of the key benefits of student performance prediction is the early identification of at-risk students. Detecting potential academic difficulties at an early stage allows educators to implement timely interventions such as tutoring, mentoring, counseling, and personalized learning plans. These measures not only improve individual student performance but also contribute to higher retention rates and reduced dropout levels. Additionally, predictive insights help teachers adapt their teaching methods to address diverse learning needs, promoting a more inclusive and effective learning environment.

The integration of machine learning models such as Decision Trees, Random Forest, Support Vector Machines, and Neural Networks has enhanced the accuracy and efficiency of performance prediction. These models analyze complex relationships among various factors influencing student success, enabling data-driven decision-making. Furthermore, predictive systems support personalized learning by identifying individual strengths and weaknesses, allowing educators to tailor instructional strategies accordingly.

Despite its advantages, the implementation of student performance prediction must address important challenges, including data privacy, ethical concerns, algorithmic bias, and the need for high-quality data. Ensuring transparency, fairness, and responsible use of student information is essential to building trust among students, educators, and institutions. Predictive systems should be used as supportive tools rather than replacements for teacher judgment and human understanding.

In conclusion, student performance prediction offers significant potential to enhance the quality of education by enabling early intervention, personalized learning, and informed decision-making. With careful implementation and ethical considerations, it can contribute to creating a more efficient, inclusive, and student-centered educational system. As educational institutions continue to adopt data-driven approaches, student performance prediction will play a crucial role in shaping the future of learning and academic success.

REFERENCES

Python Official Documentation
MongoDB Official Documentation
Scikit-learn Machine Learning Library Documentation
Educational Data Mining Research Papers