International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 13 Issue 3 ǁ March 2025 ǁ PP. 70-72

Prediction of Student Performance Using Machine Learning

T D Divya Dharshini, R Vijaya Varshini, Dr.P.Meenakshi Sundari, Msc., M.Phil., SET,Ph.D.
Department of Computer Science, Kamaraj University, Fatima College, Madurai.

Date of Submission: 09-03-2025 Date of acceptance: 23-03-2025

I. INTRODUCTION

Predicting student performance is an essential aspect of the education system, allowing educators to identify areas where students may need additional support. Factors such as study habits, assignment completion, class participation, tutoring, and attendance play a significant role in academic success. By analyzing these factors, we can develop predictive models that help students optimize their study strategies and improve their performance.

Machine learning techniques provide a powerful way to analyze student performance data. By applying a **Linear Regression** model, we aim to predict student performance scores and provide insights into academic trends. This project highlights the importance of data-driven decision-making in education and demonstrates how predictive models can enhance learning outcomes.

MACHINE LEARNING IN STUDENT PERFORMANCE PREDICTION

Machine learning plays a crucial role in predicting student performance. By analyzing large datasets, machine learning models identify patterns and predict academic success. The application of supervised learning techniques, such as **Linear Regression**, enables accurate estimation of student scores based on key performance indicators.

Supervised learning involves training a model using labeled data, allowing it to learn relationships between input features (study habits, assignments, attendance) and the target variable (performance score). This approach enhances decision-making in education by providing actionable insights to students and educators.

LINEAR REGRESSION

Linear Regression is a widely used technique for predicting student performance. It analyzes the relationship between independent variables (study habits, class participation, etc.) and the dependent variable (student performance score). The regression equation is given by:
where:

- is the predicted student performance score,
- is the intercept,
- are the coefficients,
- are the input features.

Linear Regression is easy to implement, interpretable, and effective for numerical predictions, making it an ideal choice for student performance analysis.

MACHINE LEARNING STEPS

To build an efficient machine learning model for student performance prediction, the following key steps were followed:

1. DATA COLLECTION

The dataset was collected through a Google Forms survey, where students provided responses related to:

- **Study Hours** – Daily study duration.
- **Assignment Completion** – Frequency of submitting assignments.
- **Class Participation** – Engagement in classroom discussions.
- **Tutoring** – Attendance in tutoring sessions.
- **Days Before Exam Start** – Preparation time before exams.
- **Class Attendance** – Attendance percentage categories.
- **Performance Score** – The dependent variable representing academic success.

2. DATA PRE-PROCESSING

The collected data underwent preprocessing to ensure consistency and accuracy.

- **Encoding Categorical Variables**: Categorical responses were converted into numerical values.
- **Handling Missing Values**: Missing or inconsistent values were handled using data imputation.
- **Feature Selection**: The most relevant academic factors were selected to train the model.

3. MODEL TRAINING

The pre-processed dataset was split into training (80%) and testing (20%) sets using the **train_test_split** function. The Linear Regression model was then trained using the **fit()** function, learning the relationship between input features and student performance scores.

4. MODEL EVALUATION

The trained model was evaluated using the following metrics:

- **Mean Absolute Error (MAE)** – Measures the average absolute error between predicted and actual scores.
- **Mean Squared Error (MSE)** – Evaluates the average squared differences between predicted and actual scores.
- **R-squared Score (R^2)** – Represents the proportion of variance explained by the model.

5. USER INPUT PREDICTION

A function was developed to allow users to input their study habits and receive predicted performance scores, providing students with an estimation of their academic standing based on their behaviors.

HOW TO ANALYZE

Linear Regression helps analyze student performance by considering various academic factors. The total score is calculated based on input responses, and the model predicts student performance accordingly.

Table 1 : Example of Scoring System for Student Performance Prediction

QUESTIONS	OPTIONS	ASSIGNED SCORE
Study Hours	<2 hrs,2-4 hrs,4-6 hrs,>6 hrs	0,1,3,5
Assignment Completion	Never, Rarely, Always	0,3,5
Class Participation	Never, Rarely, Always	0,3,5
Tutoring	No, Yes	0,5
Days Before Exam Start	1-2 days, 3-5 days, >5 days	0,3,5
Class Attendance	0-25%, 26-50%, 51-75%, 76-100%	0,1,3,5

The model was applied to classify student performance based on this scoring system. The predictions were compared with actual performance scores to evaluate accuracy.

Table 2 : Predicted vs Actual Student Performance Scores

S.NO	TOTAL SCORE	ACTUAL PERFORMANCE	PREDICTED PERFORMANCE
1	5	Low Performance	Low Performance
2	25	High Performance	High Performance
3	15	Moderate Performance	Moderate Performance
4	2	Low Performance	Low Performance
5	11	Moderate Performance	Moderate Performance

OUTPUT

```
= RESTART: D:\Python\ddd.py
How many hours do you study per week? 2
How often do you complete assignments? (Always/Rarely/Never): rarely
How often do you participate in class discussions? (Frequently/Rarely/Never): rarely
Do you attend tutoring sessions? (Yes/No): no
How many days before the exam do you start preparing? 1
What is your class attendance? (0-25%, 26-50%, 51-75%, 76-100%): 76-100

Predicted Student Percentage: 69.25%
```

Figure 1 : Prediction of Student Percentage

II. CONCLUSION

This study highlights the significance of predicting student performance using machine learning techniques. By analyzing key academic factors, the model successfully estimates performance scores, enabling educators and students to make informed decisions.

The **Linear Regression** model demonstrates efficiency in predicting student performance, achieving an **R^2 score of 0.88**. Future improvements could involve integrating additional features such as **motivation levels, parental support, and learning styles** to enhance prediction accuracy.

The insights from this project emphasize the growing importance of data-driven education and the need for proactive learning strategies to support student success.