Smart Study Score Predictor - Learn the Power of Regression

Name: Zafar Hussain

Course: Machine Learning Internship

Institute: [Your Institute Name]

Project: Task 01 - Regression Model

Date: November 2025

1. Objective

The objective of this project is to analyze how study habits, preparation, and lifestyle factors affect exam performance, and to predict student scores using regression algorithms.

2. Tools & Libraries Used

Python, Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn

3. Dataset Description

The dataset contains 10,000 records with the following columns:

- Hours Studied

- Previous Scores

- Extracurricular Activities

- Sleep Hours

- Sample Question Papers Practiced

- Performance Index (Target Variable)

4. Data Preprocessing

Missing values were checked and handled. The categorical column 'Extracurricular Activities' was encoded using One-Hot Encoding. The dataset was then split into training and testing sets (80/20 ratio).

5. Model Building

A Linear Regression model was trained to predict the 'Performance Index' based on input features. Model performance was evaluated using Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R^2 Score.

6. Model Evaluation

Mean Absolute Error (MAE): 1.61

Root Mean Squared Error (RMSE): 2.02

R2 Score: 0.989

The model achieved high accuracy, explaining 98.9% of performance variation.

7. Feature Importance

Feature Coefficients:

- Hours Studied: +2.85

- Previous Scores: +1.02

- Extracurricular Activities (Yes): +0.61

- Sleep Hours: +0.48

- Sample Question Papers Practiced: +0.19

All features positively influence student performance.

8. Prediction Example

Input:

Hours Studied: 6

Previous Scores: 80

Sleep Hours: 7

Sample Papers: 3

Extracurricular: Yes

Predicted Performance Index: 72.35

9. Conclusion

Regression analysis helps predict academic performance effectively based on behavioral and lifestyle factors. The model achieved a high R2 of 0.989, demonstrating strong accuracy and reliability for predictive analysis.