Student Score Prediction Based on Study Habits

Introduction

Student performance is influenced by multiple factors, including study habits and attendance. This project investigates whether it is possible to predict a student's final exam score using just two variables: hours studied and attendance percentage. By building a predictive model, educators can identify students at risk of underperformance and provide timely academic interventions.

Methodology

Dataset

The dataset contains three columns:

- Hours_Studied – number of hours a student studied before the exam
- Attendance – attendance percentage throughout the semester
- Final_Score – actual exam score obtained by the student

Modeling Approach

1. Data Preprocessing: Handled missing values and ensured valid ranges for all fields. Split dataset into training and testing sets (80/20).
2. Visualization: Scatter plot of study hours vs. final score (colored by attendance), correlation heatmap.
3. Regression Model: Linear Regression with Hours_Studied and Attendance as predictors, Final_Score as target.
4. Evaluation Metrics: R^2 Score and Mean Absolute Error (MAE) were used for performance evaluation.

Results

- Model Performance:
- R^2 Score: ~0.97 (very strong correlation)
- Mean Absolute Error: ~2–3 points

- Example Prediction:

Input $\rightarrow$ 4 hours of study and 80% attendance

Output $\rightarrow$ Predicted Final Score $\approx$ 72–74

This shows the model can predict student performance with high accuracy using just two features.

Conclusion & Improvements

This project demonstrates that study hours and attendance are strong predictors of exam performance. The linear regression model achieves high accuracy and can be useful in academic monitoring systems.

Future improvements:

- Expand dataset with more students for better generalization.

- Include additional features like assignment scores, class participation, or sleep patterns.
- Experiment with advanced models (Random Forest, Gradient Boosting) for comparison.
- Deploy as a simple web app to allow real-time score prediction.