Student Performance Analysis Project

Luv Kumar

B.Sc CSDA

Dataset

`student_performance_india_with_csda.csv` — 250 students (including B.Sc CSDA).

Problem Statement

Analyze and predict student performance based on study hours, attendance, and previous performance. Build regression models, visualize relationships, and evaluate model accuracy using R^2 and Mean Absolute Error (MAE).

Data Summary

- Total students: 250 (includes B.Sc CSDA students)
- Features: `Hours_Studied_per_week`, `Attendance_pct`, `Previous_Score`, Maths, CS, English
- Target: `Final_Score` (numeric), `Result` (Pass/Fail)

Models Trained

Linear Regression

- R^2: 0.8650
- MAE: 2.357

Random Forest Regressor

- R^2: 0.7943
- MAE: 2.998

Key Insights

- Study hours, attendance, and previous performance are all positively correlated with final scores.
- Linear Regression achieved strong accuracy ($R^2 \approx 0.86$).
- Random Forest gave a flexible but slightly lower accuracy baseline.
- Pass rate overall: 70.4%.

Recommendation

- Students should balance study hours and attendance for improved outcomes.
- Departments can use this model to identify at-risk students early.

Prepared by Luv Kumar for academic submission (B.Sc CSDA).