

Predicting Student Performance using ML

NAME: SUJAL SANTOSH SALUNKE

ID: cds05_036

Email: sujalsalunke9190@gmail.com

1. Project Objective

The primary objective of this project is to develop a machine learning model that accurately predicts the "Performance Index" of students based on factors such as study hours, prior academic scores, extracurricular participation, sleep hours, and practice with sample question papers. This predictive model aims to identify areas where students may need additional support, helping educators make data-driven decisions to enhance academic outcomes.

2. Methodology

- **Data Preprocessing:**

- Cleaned and prepared the dataset by handling any missing values, ensuring consistent data types, and normalizing features where necessary.
- Divided the dataset into training and testing subsets to assess model performance objectively.

- **Model Development:**

- Tested various regression algorithms to predict the "Performance Index," including:
 - **Linear Regression**
 - **Decision Tree**
 - **Random Forest**
 - **Gradient Boosting**
- Used standard metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and R-squared to evaluate and compare model accuracy.

- **Visualization:**

- Created visual comparisons (scatter plots) to observe how well the predicted values aligned with actual values for each feature, allowing for easy identification of trends and outliers.

3. Key Findings

- **Model Accuracy:**
 - **Linear Regression** outperformed other models with an R-squared of 0.989, making it the most effective predictor of student performance among the tested models.
 - **Gradient Boosting** and **Random Forest** also showed strong performance, but had slightly higher MSE values compared to Linear Regression.
- **Significant Features:**
 - Study hours and practice with sample question papers showed the highest correlation with student performance, indicating that focused study efforts and exam preparation significantly impact academic outcomes.

4. Model Comparisons

Model	Mean Absolute Error (MAE)	Mean Squared Error (MSE)	R-squared
Linear Regression	1.61	4.08	0.989
Decision Tree	2.34	8.81	0.976
Random Forest	1.81	5.16	0.986
Gradient Boosting	1.66	4.36	0.988

5. Actionable Insights & Recommendations

- **Targeted Study Plans:** Encourage students to focus on practicing sample question papers and increase their study hours in manageable increments, as these factors have the highest positive influence on performance.
- **Balanced Lifestyle:** Adequate sleep and extracurricular activities also play a role in sustaining academic success, suggesting that students should balance study with rest and hobbies.
- **Data-Driven Support Programs:** Schools can use similar models to proactively identify students needing additional academic support and design personalized interventions to improve overall student outcomes.

Project Link:

https://drive.google.com/drive/folders/1-KiWdFrCJXcnmHs1W1sPi27TnCZoE7N?usp=drive_link