**TITLE :- STUDENT PREDICTION REPORT**

NAME :- ALOK KHARWAR
BRACH :- CSE(AI)
SEC :- A
ID :- 202401100300028

# Student Performance Prediction Report

## Introduction

This project aims to predict students' final exam scores based on their study hours and previous scores. The Linear Regression model was employed to analyze the data and predict outcomes effectively.

## Dataset Description

The dataset contains the following key attributes:

- **StudyHours**: The number of hours a student studied.
- **PreviousScores**: Scores obtained in earlier assessments.
- **FinalExamScore**: The target variable representing the final exam result.

## Steps Performed

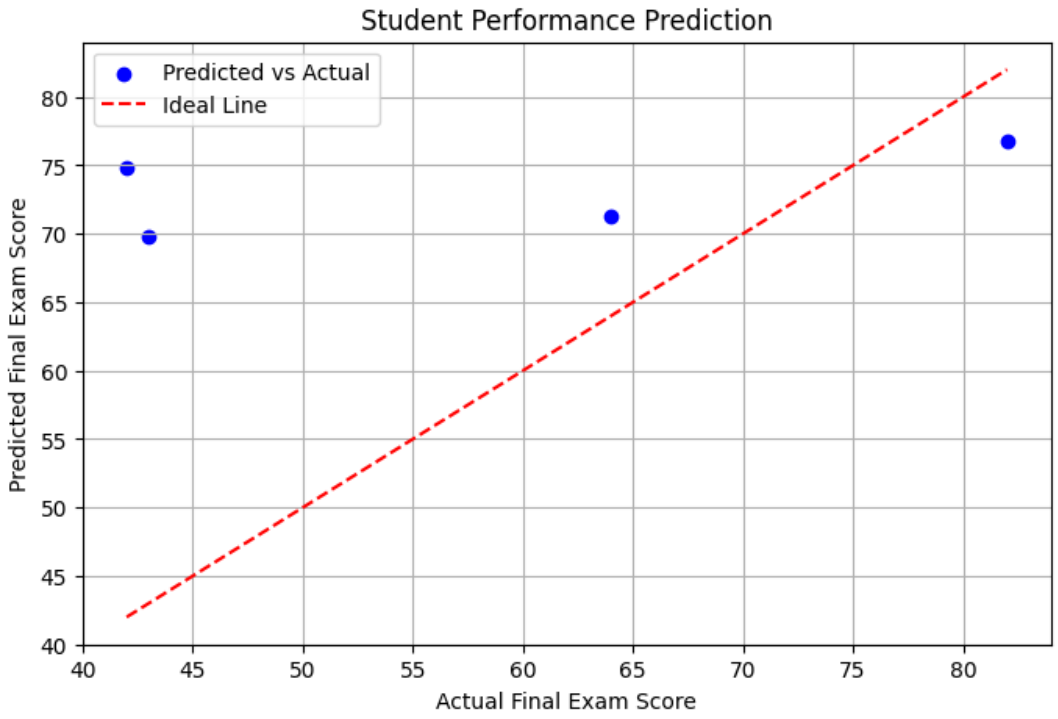
1. **Data Loading**: The dataset was imported using Pandas from the file `/content/student_data.csv`.
2. **Data Exploration**: Displayed the first few rows to understand the data structure and ensure correctness.
3. **Feature Selection**: Chose `StudyHours` and `PreviousScores` as input features (`X`) and `FinalExamScore` as the target variable (`y`).
4. **Data Splitting**: Divided the dataset into 80% training data and 20% testing data using `train_test_split()`.
5. **Model Training**: Applied the `LinearRegression()` model from scikit-learn.
6. **Prediction**: Used the trained model to predict test set results.
7. **Evaluation**: Assessed model performance using:
    - **Mean Squared Error (MSE)**: Measures average squared error between actual and predicted values.
    - **R-squared (R2) Score**: Represents the model's accuracy by indicating the proportion of variance explained by the model.

## Results

- **Mean Squared Error (MSE)**: The lower the value, the better the model's accuracy.
- **R-squared (R2) Score**: Closer to 1 indicates a strong correlation between predicted and actual values.

| | StudentID | StudyHours | PreviousScores | FinalExamScore |
|---|---|---|---|---|
| **0** | 1 | 8.777482 | 75 | 64 |
| **1** | 2 | 9.161915 | 55 | 82 |
| **2** | 3 | 3.278010 | 77 | 70 |
| **3** | 4 | 4.500247 | 60 | 60 |
| **4** | 5 | 2.264931 | 72 | 60 |

```
Mean Squared Error (MSE): 468.88511292976506
R-squared (R2) score: -0.716349075011723
```




## Visualization

A scatter plot was generated to visualize the relationship between actual and predicted exam scores. An ideal prediction line (in red) was added for comparison, demonstrating the model's accuracy.

## Conclusion

This Linear Regression model effectively predicts students' final exam scores based on their study hours and previous scores. Further improvements can be achieved by incorporating additional features or experimenting with different machine learning models.

## Recommendations

- Collect more data for better model generalization.
- Try advanced regression techniques like Ridge or Lasso for improved accuracy.
- Explore feature engineering for enhanced predictive performance.