# Student Score Prediction using Linear Regression

## 1. Objective of the Task

The objective of this project is to develop a Linear Regression model to predict students' exam scores based on their study hours. The aim is to analyze the relationship between study time and academic performance and evaluate how effectively study hours explain variations in exam scores.

## 2. Dataset Description

The dataset used is 'Student Performance Factors' from Kaggle, containing 6,607 student records with academic and lifestyle-related features. For this task, the target variable was Exam_Score, and the primary feature used was Hours_Studied.

## 3. Methodology

- Data Loading: The dataset was loaded using Pandas in Google Colab.
- Data Cleaning: Checked for missing values and ensured correct data types.
- Feature Selection: Selected Hours_Studied as the independent variable and Exam_Score as the target variable.
- Exploratory Data Analysis: Created scatter plots to visualize the relationship between study hours and exam score.
- Train-Test Split: Split the dataset into 80% training data and 20% testing data.
- Model Training: Applied Linear Regression using Scikit-learn.
- Model Evaluation: Evaluated performance using MAE, RMSE, and $R^2$ score.
- Bonus Experiment 1: Applied Polynomial Regression (degree 2) to test for non-linear patterns.
- Bonus Experiment 2: Added Previous_Scores feature and compared model performance.

## 4. Results

| Model | $R^2$ Score |
|---|---|
| Linear Regression (Hours Only) | 0.23 |
| Polynomial Regression | 0.23 |
| Multiple Linear Regression | 0.25 |

## 5. Conclusion

The analysis shows that study hours have a positive relationship with exam scores. However, the $R^2$ score of 0.23 indicates that study hours alone explain only 23% of the variation in student performance. Polynomial regression did not significantly improve performance, suggesting the relationship is mostly linear. Adding an additional feature (Previous_Scores) slightly improved model accuracy, highlighting the importance of multiple contributing factors in academic performance prediction.