

NAME: RAJA ZEESHAN AHMED

Project Report: Student Exam Score Prediction Task

1. The Task I Was Assigned

I was given the task of building a model to predict student exam scores using the provided StudentPerformanceFactors.csv dataset. The initial requirement was to see if a student's score could be predicted based on their study hours alone. The assignment also included bonus objectives to experiment with more advanced models to see if I could improve the prediction accuracy.

2. How I Approached the Task

To complete the assignment, I broke down the problem into a clear, multi-stage process. I decided to build three different models, starting with the basic requirement and then progressing to the bonus objectives to demonstrate a thorough analysis.

Model 1: Fulfilling the Core Requirement (Simple Linear Regression)

- **My Goal:** First, I needed to address the main task: predict exam scores using only the hours a student studied.
- **What I Did:** I built a simple linear regression model that used just one piece of information (Hours_Studied) to predict the Exam_Score.
- **The Result:** This initial model achieved an **R-squared (R^2) value of 0.23**. This metric tells us how much of the outcome the model can explain. A score of 0.23 meant that knowing the study hours could only explain about **23%** of why a student got their specific score.
- **My Conclusion:** I successfully completed the primary task and had a working model. However, the result showed that while studying helps, it wasn't the only important factor, and a more advanced model was needed to get a truly useful prediction.

Model 2: The First Bonus Task (Polynomial Regression)

- **My Goal:** The first bonus objective was to see if a more complex relationship, like a curve, could improve the prediction.
- **What I Did:** I experimented with a polynomial regression model, which can find curved patterns in data instead of just straight lines.
- **The Result:** The R-squared value was still **0.23**, showing no real improvement.

- **My Conclusion:** I completed the first bonus task and determined that a more complex curve didn't fit the data any better than a simple straight line. This was a valuable finding, as it proved that this approach was not the key to better accuracy.

Model 3: The Second Bonus Task (Multiple Linear Regression)

- **My Goal:** The second bonus objective was to improve the model by using a combination of different features.
- **What I Did:** I chose several new features that I believed would be influential: Previous_Scores, Attendance, Parental_Involvement, and Teacher_Quality. I prepared this new data by cleaning it and converting it into a format the model could use.
- **The Result:** This approach was a major success. The R-squared score jumped to **0.66**. My model's average error was also cut in half, making its predictions much more accurate.
- **My Conclusion:** I successfully completed the second bonus task and created a model that was far superior to the first two.

3. Final Result of My Work

After completing all assigned tasks, my final conclusion is that the **Multiple Linear Regression model was the most effective solution**.

By fulfilling the bonus objectives and incorporating more data, I was able to build a model that could explain **66%** of the variance in student scores, a nearly threefold improvement on the baseline model required by the initial task.

This project demonstrated that while the initial task was achievable with a simple model, a much more accurate and useful prediction can be made by thoughtfully selecting more features. The final visualization I created, which plots the model's predictions against the actual scores, clearly shows the high accuracy of this final model and serves as a successful completion of the assignment.