

# Project Report

---

## Student Score Prediction Based on Study Habits

### 1. Introduction

Student academic performance is influenced by multiple factors such as study hours, class attendance, participation, and motivation. Among these, study hours and attendance are two of the most measurable and impactful parameters. The aim of this project is to develop a regression-based predictive model that can estimate a student's final exam score using these two factors.

By using machine learning, we can create a data-driven approach that helps students and educators understand the relationship between study habits and performance, and provide guidance for improvement.

### 2. Problem Statement

**Question:** Can we predict a student's final exam score using study hours and attendance data?

The challenge is to build a machine learning model that takes input values of:

- Hours Studied
- Attendance (%)

and predicts:

- Final Exam Score

### 3. Objectives

- Collect and preprocess student data.
- Visualize relationships between study hours, attendance, and scores.
- Train and test a Linear Regression model.
- Evaluate the model using performance metrics such as  $R^2$  Score and Mean Absolute Error (MAE).
- Demonstrate predictions for new student inputs.

## 4. Tools & Technologies

- Python
- Pandas
- Matplotlib & Seaborn
- Scikit-learn

## 5. Methodology

### Step 1: Data Collection

- A sample dataset with three attributes: Hours\_Studied, Attendance (%), and Final\_Score.

Example:

```
Hours_Studied | Attendance | Final_Score
5 | 90 | 85
3 | 60 | 55
6 | 95 | 90
```

### Step 2: Data Preprocessing

- Handle missing values if any. Normalize/scale values if required. Split dataset into training (80%) and testing (20%) sets.

### Step 3: Data Visualization

- Scatter plot of Hours Studied vs Final Score.
- Scatter plot of Attendance vs Final Score.
- Correlation heatmap to check relationships.

### Step 4: Model Building

- Use Linear Regression from Scikit-learn. Train the model using training data. Fit line represents relationship between inputs and output.

### Step 5: Prediction & Evaluation

- Predict scores for test dataset.  
Example prediction:
  - Input: Hours Studied = 7, Attendance = 80%
  - Output: Predicted Score  $\approx$  85.5 (depends on dataset).
- Evaluation metrics:
  - $R^2$  Score – measures accuracy of model fit.
  - Mean Absolute Error (MAE) – average prediction error.

## 6. Results & Discussion

- The regression model successfully captured the positive correlation between study hours, attendance, and exam scores. Higher study hours and consistent attendance generally lead to better performance. Model evaluation showed a good fit with low error, demonstrating reliability in prediction.

## 7. Conclusion

This project demonstrates how simple student data like study hours and attendance can be used to predict academic performance using machine learning.

1. It can help students set study goals.
2. It can help teachers identify students at risk.

### **Future Work:**

Include more features like assignment scores, participation, or sleep patterns.

Apply more advanced models such as Random Forest Regression or Neural Networks for higher accuracy.

## 8. References

1. Scikit-learn Documentation: <https://scikit-learn.org>
2. Pandas Documentation: <https://pandas.pydata.org>
3. Matplotlib Documentation: <https://matplotlib.org>