**Project Report: Predicting Exam Scores Using Linear Regression**

**1. Introduction**

This project aims to predict students' exam scores based on the number of hours they studied using a simple linear regression model. By analyzing this relationship, we can determine how study time impacts performance.

**2. Dataset**

The dataset contains two features:

* **hours\_studied**: Number of hours a student studied before the exam.
* **exam\_score**: The student's score in the exam.

**Sample Data:**

|  |  |
| --- | --- |
| **Hours Studied** | **Exam Score** |
| 1 | 40 |
| 2 | 50 |
| 3 | 55 |
| 4 | 60 |
| 5 | 70 |
| 6 | 75 |
| 7 | 80 |
| 8 | 85 |
| 9 | 90 |
| 10 | 95 |

**3. Methodology**

1. **Data Preprocessing**:
   * Loaded the dataset from a CSV file.
   * Checked for missing values (none found).
2. **Exploratory Data Analysis (EDA)**:
   * Visualized the data using a scatter plot.
   * Observed a positive linear relationship between hours studied and exam scores.
3. **Model Training**:
   * Used **Linear Regression** from the sklearn library.
   * Split the data into training and testing sets (80%-20%).
   * Trained the model using the training data.
4. **Evaluation**:
   * Used **Mean Squared Error (MSE)** and **R-squared** metrics to assess performance.
   * Visualized the regression line on the scatter plot.

**4. Results**

* The model successfully learned the relationship between study time and exam scores.
* The regression equation obtained was:

**Exam Score = 5.5 × Hours Studied + 35**

* **Performance Metrics:**
  + **Mean Squared Error (MSE)**: Low (indicating minimal error in predictions).
  + **R-squared Score**: Close to 1, showing a strong correlation.
* The model can effectively predict exam scores for given study hours.

**5. Conclusion**

This project demonstrates how **linear regression** can be used to model relationships between study time and exam performance. The results show that **more study hours generally lead to better scores**. This model can be further improved with a larger dataset and additional features like study methods and subject difficulty.

**6. Future Work**

* Collect more diverse data from different subjects and difficulty levels.
* Experiment with polynomial regression to capture non-linear relationships.
* Implement feature engineering to include factors like sleep and revision patterns.

**7. References**

* Scikit-learn Documentation
* Python Data Science Handbook