**📘 Task 1: Student Score Prediction — Final Report**

**1. Objective**

To build a machine learning model that predicts students’ **exam scores** based on their study hours and related factors such as sleep hours, attendance, motivation level, and previous scores.

**2. Dataset**

**Source:** *Student Performance Factors* (Kaggle)  
**Description:** The dataset includes various academic, social, and personal factors influencing student performance.  
**Rows & Columns:** 6608 × 20   
**Target Variable:** Exam\_Score

**3. Tools & Libraries**

* **Python**
* **Pandas** – Data handling and cleaning
* **Matplotlib / Seaborn** – Data visualization
* **Scikit-learn** – Machine learning modeling and evaluation

**4. Methodology**

| **Step** | **Description** |
| --- | --- |
| **Data Cleaning** | Verified no missing values; encoded categorical columns (e.g., Motivation\_Level) using pd.get\_dummies(). |
| **Exploratory Data Analysis (EDA)** | Visualized relationships — *Hours\_Studied vs Exam\_Score* showed a clear positive correlation. |
| **Feature Selection** | Started with Hours\_Studied only, later included Sleep\_Hours, Previous\_Scores, Attendance, and encoded Motivation\_Level. |
| **Train-Test Split** | 80 % training / 20 % testing via train\_test\_split(). |
| **Modeling** | Trained Linear Regression and Polynomial Regression (degree 2). |
| **Evaluation Metrics** | MAE, RMSE, and R² score used to measure accuracy. |

**5. Results**

| **Model** | **Features Used** | **MAE** | **RMSE** | **R²** |
| --- | --- | --- | --- | --- |
| **Linear Regression (1 feature)** | Hours\_Studied | 2.53 | 3.51 | 0.205 |
| **Linear Regression (5 features)** | + Sleep, Attendance, Motivation, Previous Scores | 1.40 | 2.53 | 0.587 |
| **Polynomial Regression (deg 2)** | Hours\_Studied only | 2.53 | 3.51 | 0.205 |

✅ *Adding multiple features boosted performance significantly (R² ≈ 0.59).*  
❌ *Polynomial regression didn’t improve accuracy → relationship is linear.*

**6. Feature Importance (Coefficients)**

| **Feature** | **Effect on Score** | **Interpretation** |
| --- | --- | --- |
| **Hours\_Studied** | Positive (≈ 0.0 – 0.2) | More study → higher score |
| **Attendance** | Positive (≈ 0.0 – 0.2) | Regular attendance helps |
| **Previous\_Scores** | Positive (≈ 0.0 – 0.2) | Past performance predicts future success |
| **Sleep\_Hours** | Slightly negative (< 0) | Too much/little sleep may hurt focus |
| **Motivation\_Level (Medium/Low)** | Negative (-0.6 to -1.0) | Lower motivation reduces performance |

**7. Visualizations**

* Scatter plot – Hours Studied vs Exam Score
* Actual vs Predicted exam scores
* Feature importance bar chart
* Correlation heatmap (optional)

*(All produced and verified in Jupyter Notebook.)*

**8. Conclusions**

* Study hours, prior performance, and attendance are **strong predictors** of exam success.
* Motivation levels show a clear psychological impact.
* Sleep requires balance — both deprivation and oversleeping correlate with lower results.
* The model explains roughly **59 %** of exam score variation, making it a strong baseline for academic performance prediction.