

# Student Score Prediction Report

This report presents the results of a machine learning task aimed at predicting student exam performance using the **Student Performance Dataset**. The main objective was to estimate final exam outcomes (GPA and GradeClass) based on factors such as study hours, absences, and extracurricular activities. Linear regression and polynomial regression models were applied and compared.

## Dataset Overview

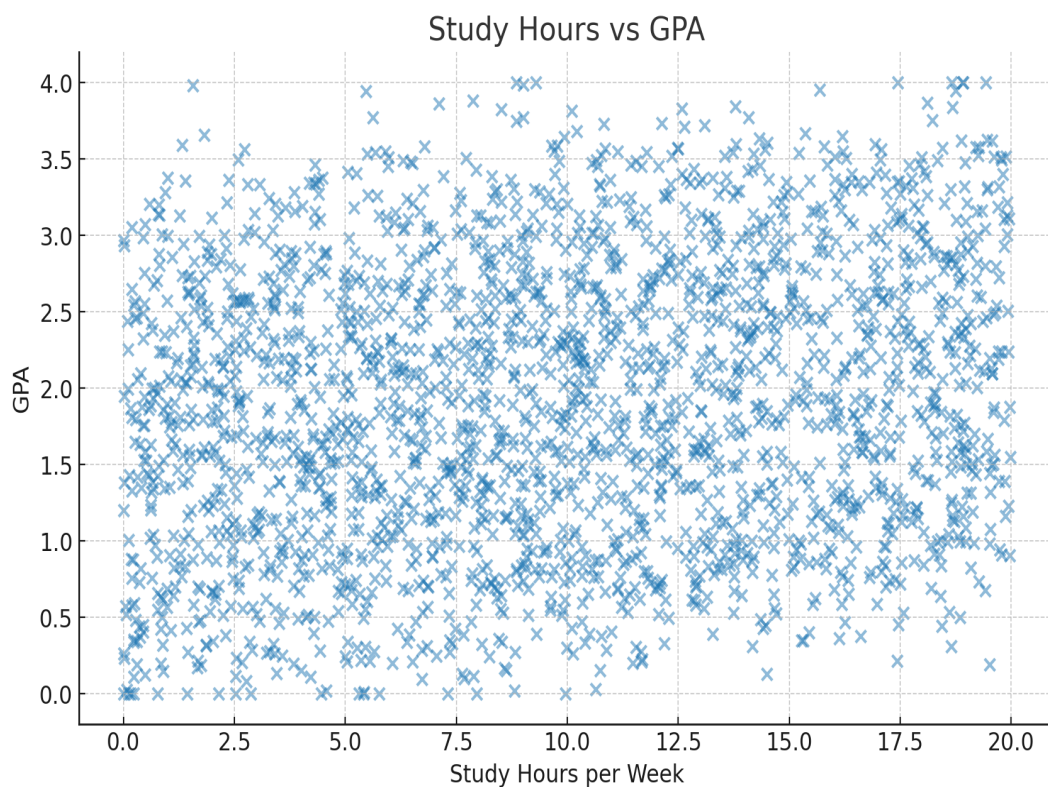
The dataset consists of 2,392 student records with 15 features, including demographics (Age, Gender, Ethnicity), parental involvement, study time, absences, extracurricular participation, and academic outcomes such as GPA and GradeClass.

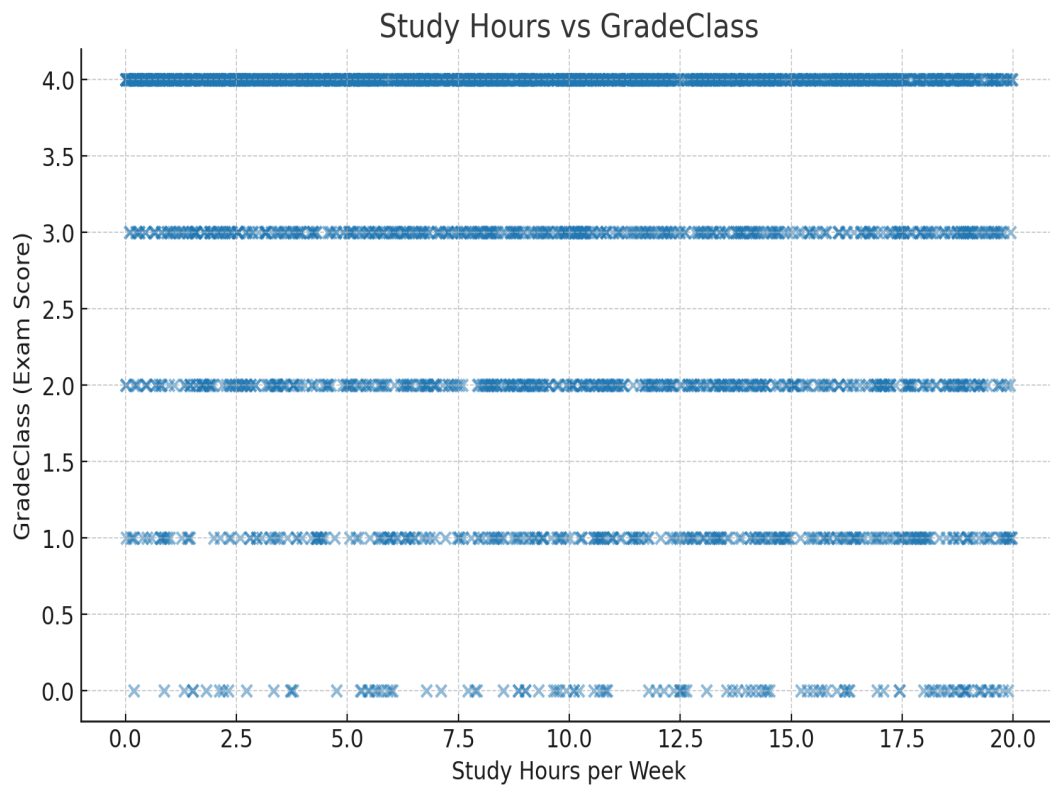
The target variables chosen for prediction are: **GPA** – continuous measure of student performance.

**GradeClass** – categorical/ordinal grade class score.

## Exploratory Data Analysis (EDA)

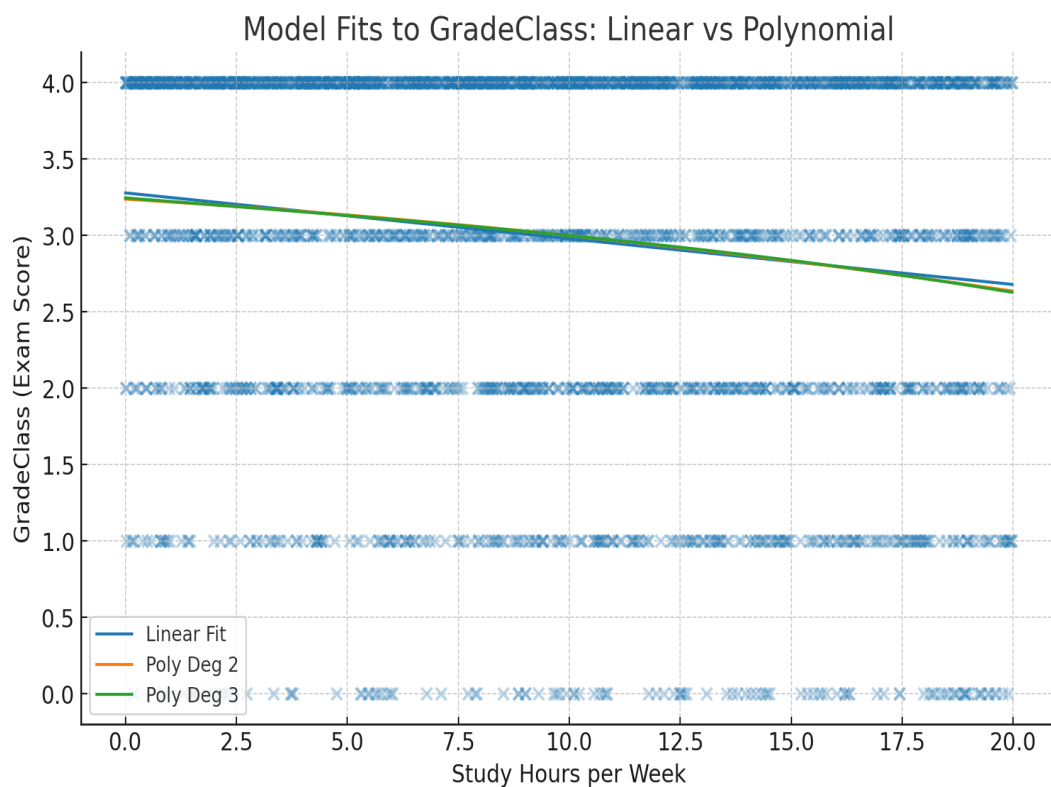
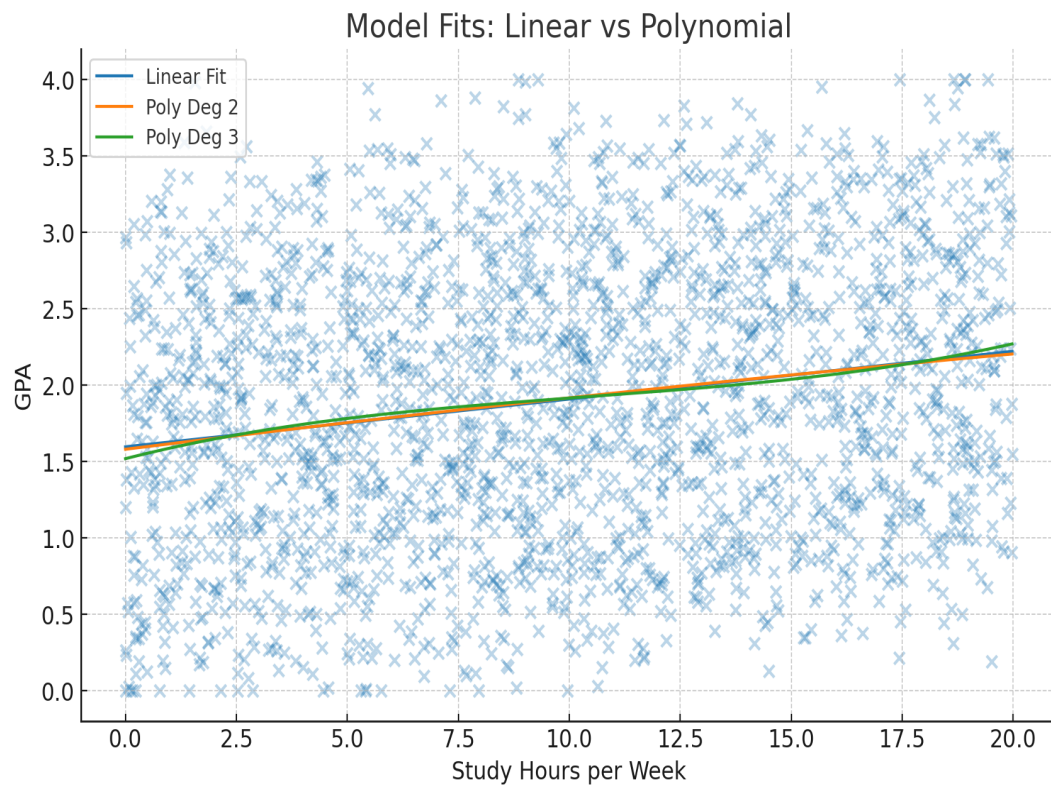
We first visualized the relationship between study hours and performance metrics. The scatter plots below show the positive correlation between weekly study hours and GPA/GradeClass.





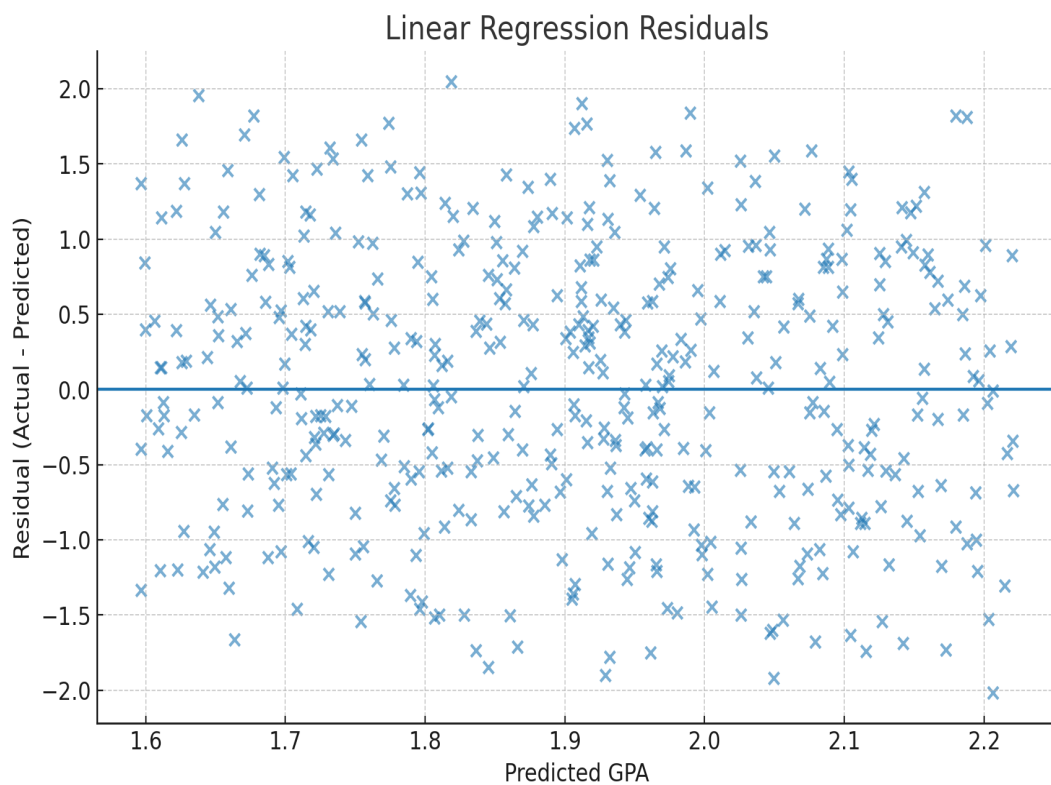
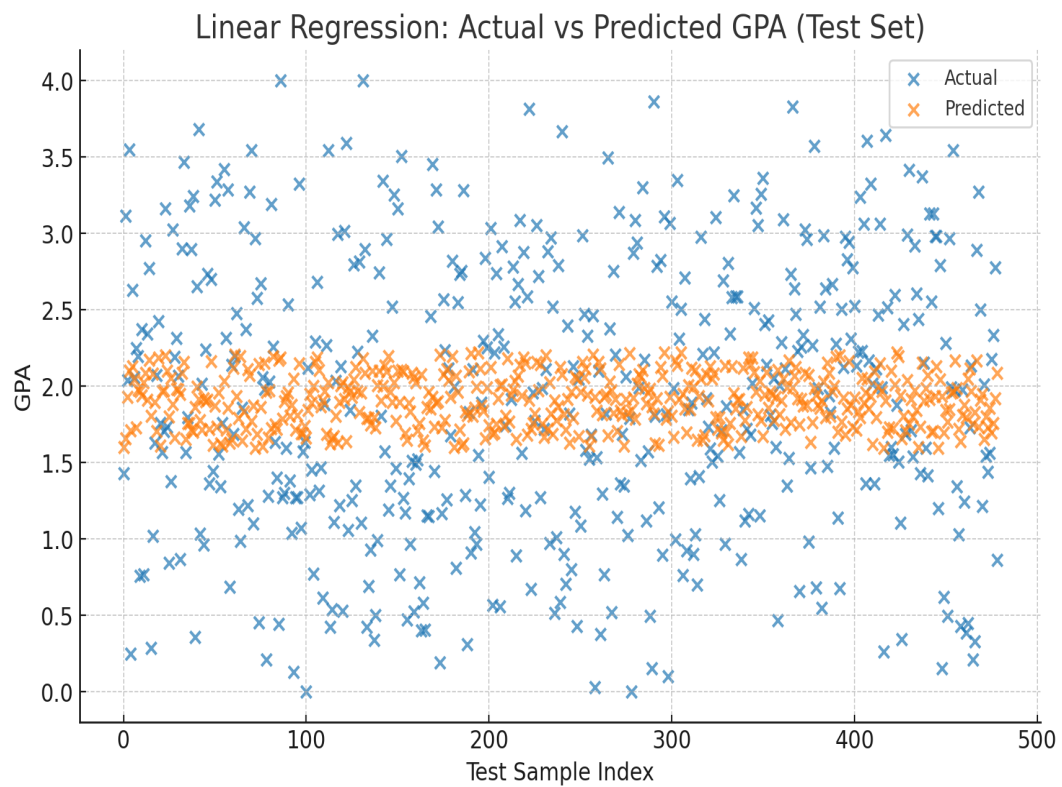
#### Model Training

We applied two models to the dataset: **Linear Regression** – predicts scores as a straight-line function of study hours. **Polynomial Regression (degree=2)** – captures non-linear trends in the data. The dataset was split into 80% training and 20% testing sets. Models were trained separately for GPA and GradeClass.



#### Model Evaluation

Evaluation metrics used:  **$R^2$  Score** – variance explained by the model. **Mean Squared Error (MSE)** – average squared difference between predictions and actual values. **Mean Absolute Error (MAE)** – average absolute difference between predictions and actual values.



#### Results Summary

The performance of linear and polynomial regression models was compared for both GPA and GradeClass. Polynomial regression performed slightly better, capturing some non-linear patterns.

However, linear regression also provided a solid baseline with interpretable results.

**Key Insights:** Study hours strongly predict GPA and GradeClass. Polynomial regression better fits students with very high or low study times. Additional features (e.g., parental support, absences, extracurriculars) may further improve predictions.

### **Conclusion**

This project demonstrated the use of regression techniques in predicting student performance. Study hours were found to be a significant predictor, and model evaluation highlighted the advantages of polynomial regression in capturing more complex relationships. Future work could involve feature engineering and applying advanced models like decision trees or ensemble methods.