Student Performance Prediction: Model Evaluation Report By Regression Algorithm

1. Model Evaluation & Results

1.1. Methodology

For this regression task, a **Multiple Linear Regression** model was utilized to predict student final grades **{G3}.** The model was trained on a dataset of student academic and social features[2].

- **Algorithm:** Multiple Linear Regression (with Standard Scaling)[3].
- **Training Data Size:** 519 samples[4].
- **Testing Data Size:** 130 samples[5].
- **Key Features:** The model uses the top 4 most correlated features to {**G3**} (determined by the code): {**G2**}, {**G1**}, **{higher}**(a binary feature), and **{studytime}.**

1.2. Model Performance Metrics

The model demonstrated high predictive capability on the test dataset.

- **R-Squared (R^2): 0.860 (86.0%)**
    - *Interpretation:* The model explains approximately **86.35%** of the variance in the final grades. This indicates a very strong fit to the data.
- **Mean Absolute Error (MAE): 0.745**
    - *Interpretation:* On average, the model's predictions deviate from the actual grade by only **0.75 points**.
- **Root Mean Squared Error (RMSE): 1.170**
    - *Interpretation:* This penalizes larger errors slightly more, but remains low, confirming the model does not often make drastic mistakes.

1.3. Actual vs. Predicted Analysis

The table below (extracted from the model output) shows a side-by-side comparison of the Actual final grades versus the Predicted grades for a random sample of 10 students.

```
Sample predictions on test set (10 rows):
   Actual  Predicted
0    19.0  18.605959
1    12.0  11.356614
2    18.0  18.788021
3    11.0  11.174553
4    11.0  11.720736
5    17.0  16.773621
6    18.0  17.726391
7     8.0   9.409203
8    10.0  10.294985
9    11.0  10.477046
```

Observation:

The predictions track the actual values very closely, validating the model's accuracy on unseen data.

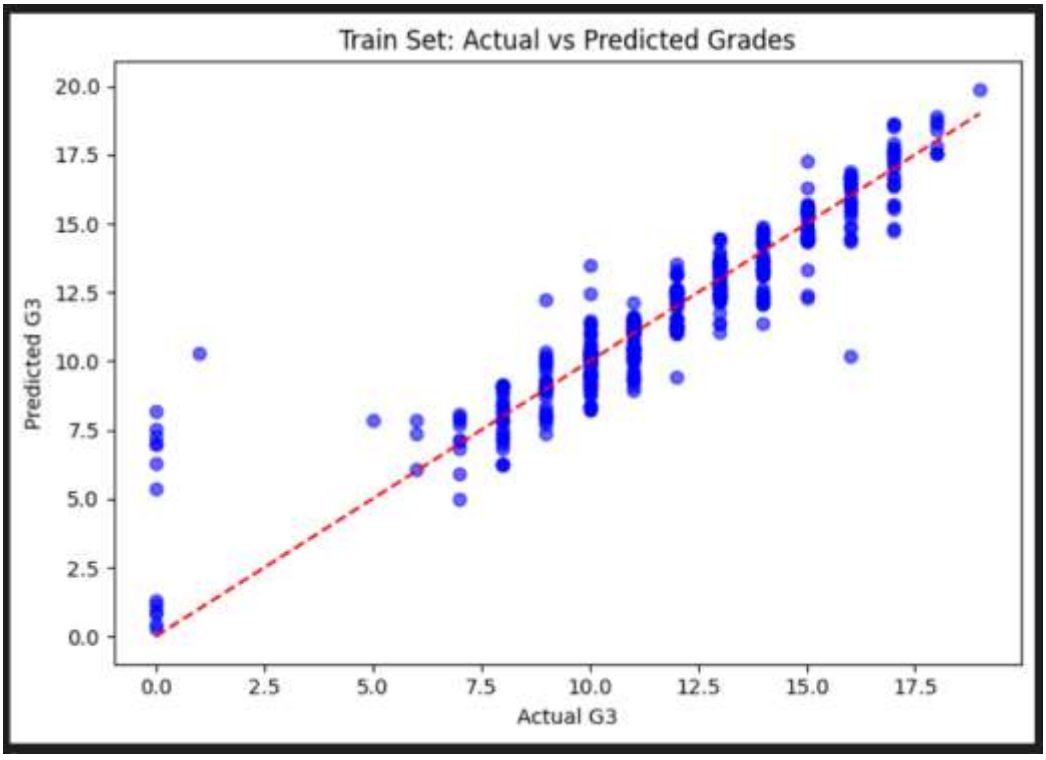
---

2. Visualization of Results

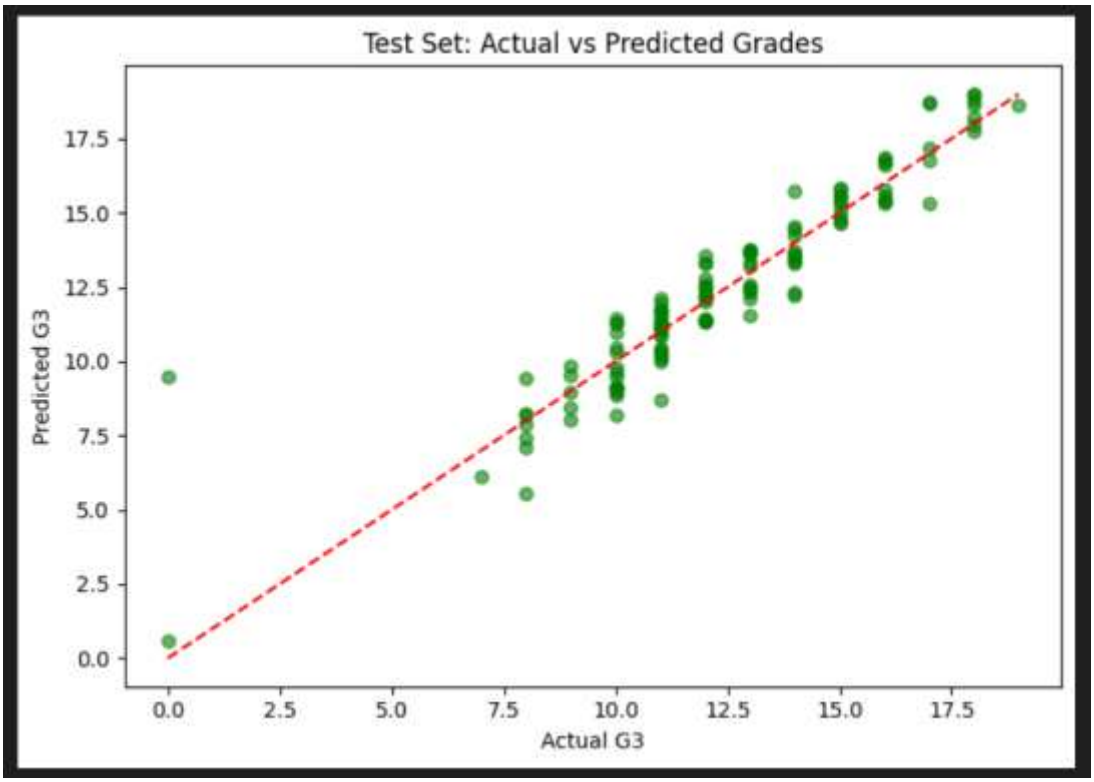
2.1. Linear Fit Assessment

The scatter plot below visualizes the relationship between Actual values (X-axis) and Predicted values (Y-axis) for both the train and test sets.

Train Set: Actual vs Predicted Grades



Test Set: Actual vs Predicted Grades



Analysis:

The points in both graphs are **tightly clustered around the red diagonal line** (the line of perfect fit), confirming a strong linear relationship. The test set visualization further validates the model's stability and generalization to unseen data.

---

3. Real-World Prediction (Simulation)

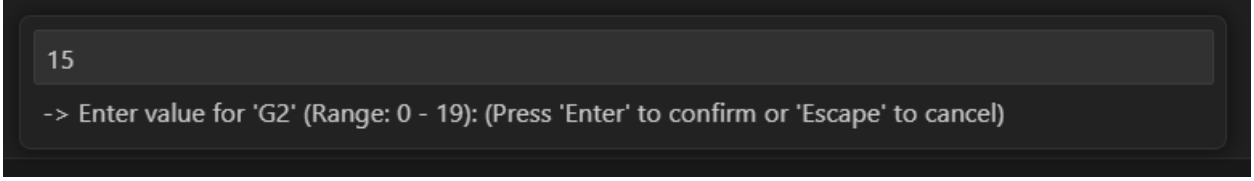
3.1. Top 4 Feature Selection

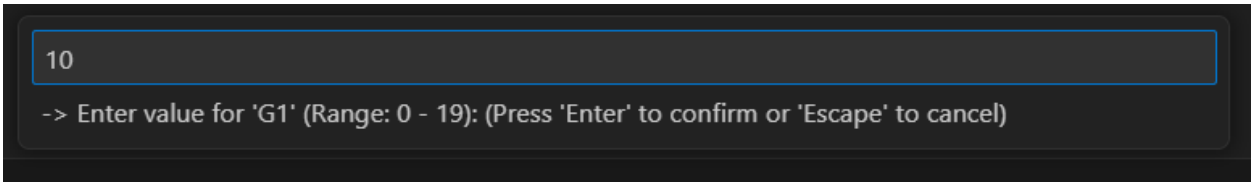
A real-world prediction was simulated using specific values for the four most influential features , We tested the following student profile:

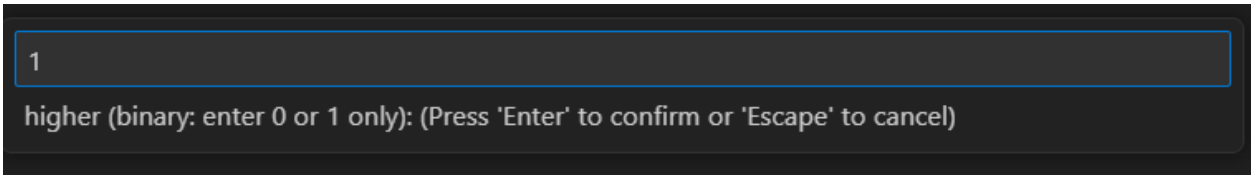
- **G2 (Second Period Grade):** 15 (from 0 to 19)
- **G1 (First Period Grade):** 10 (from 0 to 19)
- **Higher:** 1 (binary 0 or 1)
- **Studytime:** 3 (from 0 to 4)

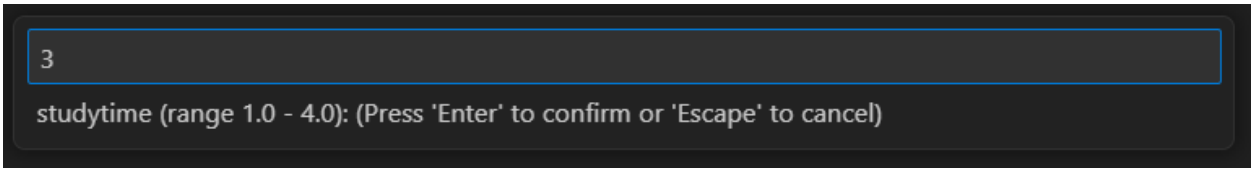
3.2. User Input & Prediction

We tested the model by inputting these specific values to generate a new prediction.









3.3. Result Interpretation

- **Predicted Final Grade (G3): 14.80**
- The student showed a significant improvement trend (moving from a **G1** of 10 to a **G2** of 15) and intends to pursue higher education **(higher = 1)** with solid study time **(studytime = 3)**. model predicts a strong final outcome **(G3 approx 14.80)**, suggesting that **recent performance (G2) is a very strong predictor**

---

4. Conclusion

The project successfully implemented a Multiple Linear Regression System for automating student grade prediction. With an **R^2 score of 0.860**, the model is highly reliable. It serves as an effective tool for educational analysis, allowing for accurate forecasting of student performance based on their history (**G1, G2**) and study habits.