Experimental Results and Evaluation

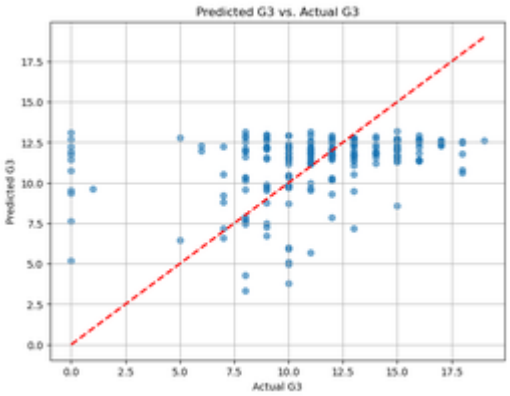
Results

Linear Regression

- **Cross-Validation Mean Squared Error (MSE):** 13.245372100163866
- **Cross-Validation R-squared (R^2):** 0.02662466381021269
- **Test Set Mean Squared Error (MSE):** 14.443667363778466
- **Test Set R-squared (R^2):** 0.06578893998139157

The performance metrics indicate that the linear regression model struggles to capture the variability in G3 using the selected features. The low R^2 values in both the cross-validation (2.66%) and the test set (6.58%) suggest that the model explains only a small fraction of the variance in student grades. This implies that the model is underfitting, meaning it lacks the complexity to accurately model the underlying relationships in the data.

Plot 1: Predicted G3 vs. Actual G3



The large spread suggests that the model is underfitting, particularly for higher actual scores where the model struggles to predict accurately. This aligns with the low R^2 value, which indicates that only a small portion of the variance is explained by the model.

Polynomial Regression

Degree 2 Polynomial Regression

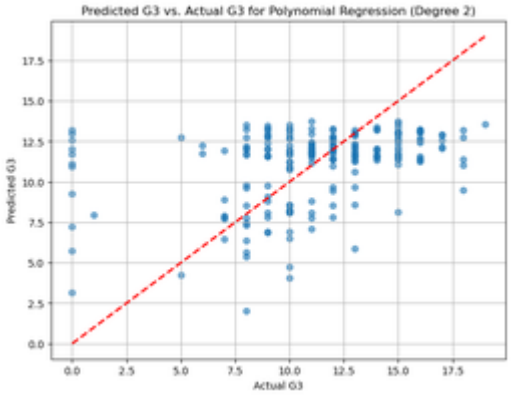
- **Mean Squared Error (MSE):** 14.158877893257824
- **R-squared (R^2):** 0.08420901754454779

Degree 3 Polynomial Regression

- **Mean Squared Error (MSE):** 21.1068426051264
- **R-squared (R^2):** -0.3651827688326632

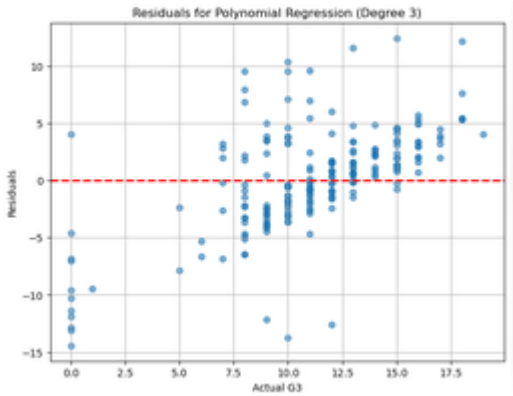
Polynomial regression improves model fit compared to regular linear regression when using degree 2, as seen by the slightly higher R^2 value. However, the improvement is not substantial, and increasing the polynomial degree to 3 leads to severe overfitting.

Plot 2: Predicted G3 vs. Actual G3 for Polynomial Regression (Degree 2)



The scatter plot shows the relationship between predicted and actual G3 values for degree 2 polynomial regression. Some clustering along the ideal fit line is observed, but the spread remains large. This indicates that the model still has difficulty accurately predicting G3 for many students. The low R^2 value (8.42%) suggests that the model does not capture the underlying complexity effectively.

Plot 3: Residuals for Polynomial Regression (Degree 3)



The residuals plot for degree 3 polynomial regression shows a significant spread, with no clear pattern around zero. Large positive and negative residuals indicate that the model is **overfitting** and fails to generalize well to the test data. Residuals are not randomly distributed, and the wide variance highlights the instability of the degree 3 model. The negative R^2 value supports that the model performs worse than a simple baseline prediction.