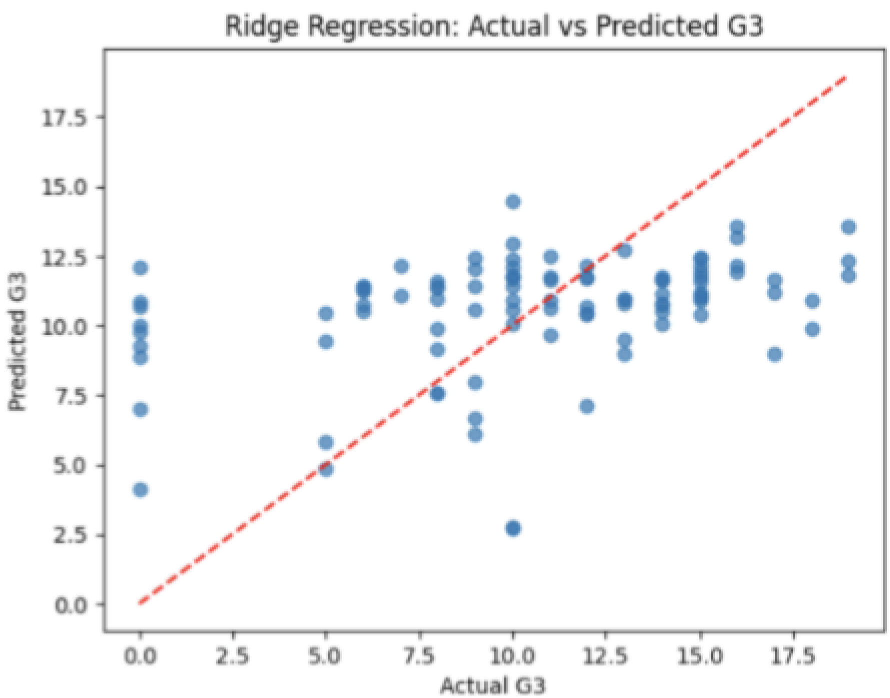
Ridge and Lasso Regression

In this study, Ridge and Lasso Regression were used to predict students' final grades based on a variety of features, including parental education, study habits, extracurricular involvement, and support from family and school. These methods were chosen because of their ability to handle multicollinearity and potentially select the most important predictors.

Ridge Regression

Ridge Regression is particularly effective when multicollinearity exists among features, as it adds an L2 penalty term to the loss function. This penalty helps shrink the coefficients, thereby reducing model complexity and variance. The approach is robust in maintaining model stability while retaining all features in the dataset, making it useful for problems where predictors are highly correlated.

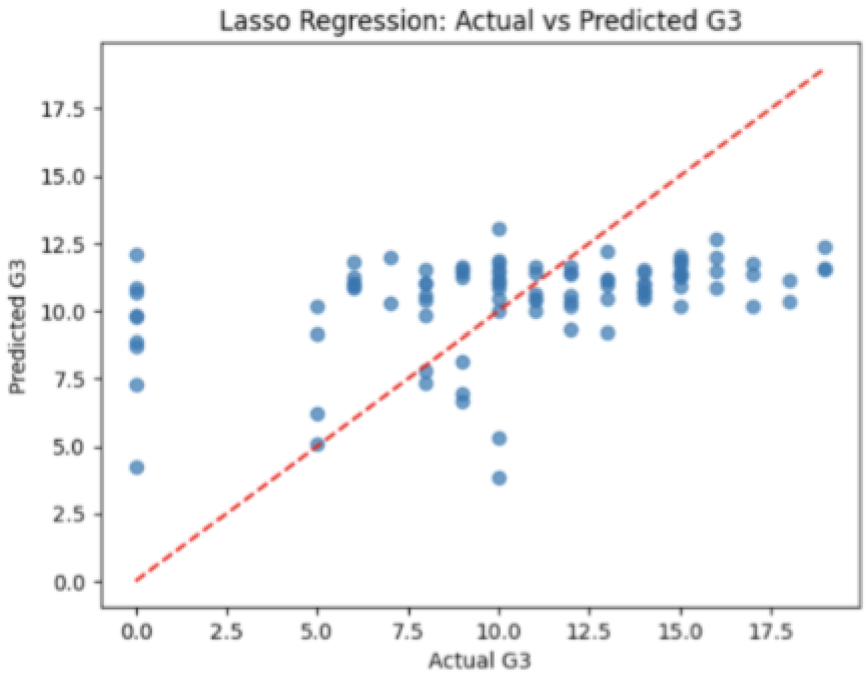
For this analysis, Ridge Regression was tested with various regularization strengths (alpha values) ranging from 0.01 to 100. The model demonstrated moderate predictive performance, with R-squared values ranging between **0.12 and 0.15**. This indicates that only a limited portion of the variance in students' final grades was explained by the model. While Ridge performed slightly better than baseline linear regression, the improvements were marginal. The results suggest that the model might not fully capture the non-linear relationships or other key factors influencing academic success.



Lasso Regression

Lasso Regression offers an alternative to Ridge by introducing an L1 penalty term in its loss function. This penalty has a unique property: it drives some feature coefficients to zero, effectively performing feature selection. This makes Lasso particularly valuable in datasets with a large number of predictors, where some features may be redundant or irrelevant.

In this study, Lasso Regression was applied with a range of alpha values, and the best performance was observed at an alpha value of **0.1**, yielding an R-squared value of **0.16**. Although this represents only a slight improvement over Ridge Regression, it suggests that some predictors in the dataset may not contribute significantly to the target variable. By shrinking the coefficients of less important features to zero, Lasso provided a slightly more interpretable model while maintaining similar predictive power.



Gradient Boosting Regression

Inspired by previous studies, Gradient Boosting Regression was applied to capture complex relationships in the data. We experimented with various learning rates (0.01, 0.05, 0.1, 0.2). The results are summarized below:

Learning Rate	Mean Squared Error	R^2
0.01	18.48	0.10
0.05	16.91	0.18
0.1	17.93	0.13
0.2	19.29	0.06

Conclusion

Gradient Boosting showed a marginal improvement in predictive performance compared to Ridge and Lasso Regression. However, the relatively low R-squared values across all models suggest that predicting student performance remains challenging. Further improvement requires richer datasets and additional features to better capture the complexities influencing student success.