# Conclusion

In this project, we explored various machine learning approaches to develop a model for predicting grades using the dataset from the Irvine repository. Our goal was to identify the most suitable model for this task by balancing underfitting and overfitting while maximizing predictive accuracy.

We began with linear regression and enhanced it with techniques such as gradient boosting and L1/L2 regularization. However, L1 and L2 regularization were found to over-penalize the model, limiting its effectiveness, especially since the regular linear regression was not prone to overfitting in the first place.

Polynomial regression with a degree of 2 (Poly2) emerged as the best-performing model. While it slightly underfits the data, it provided a reasonable balance between bias and variance. In contrast, polynomial regression with a degree of 3 (Poly3) tended to overfit, capturing noise rather than meaningful patterns.

Finally, we experimented with decision trees, but their performance fell short compared to the Poly2 model, likely due to their tendency to overfit small datasets or underperform with limited complexity.

In conclusion, the Poly2 model strikes the most appropriate balance for this dataset, demonstrating strong predictive capabilities while avoiding the pitfalls of overfitting or underfitting. Future work could explore ensemble methods or hybrid approaches to further refine the predictions.