
Predicting Student Performance
Insights from Data-Driven Analysis and Algorithmic Optimization

Agenda

- Dataset overview
- Methodology
- Models Evaluated
- Hyperparameter Tuning
- Model Performance
- Conclusion
- References

Dataset Overview

Portuguese (649R, 33C)

Math (395R, 33C)

Methodology

- Data Preprocessing
- Model Selection
- Hyperparameter Tuning
- Evaluation

Data Preprocessing

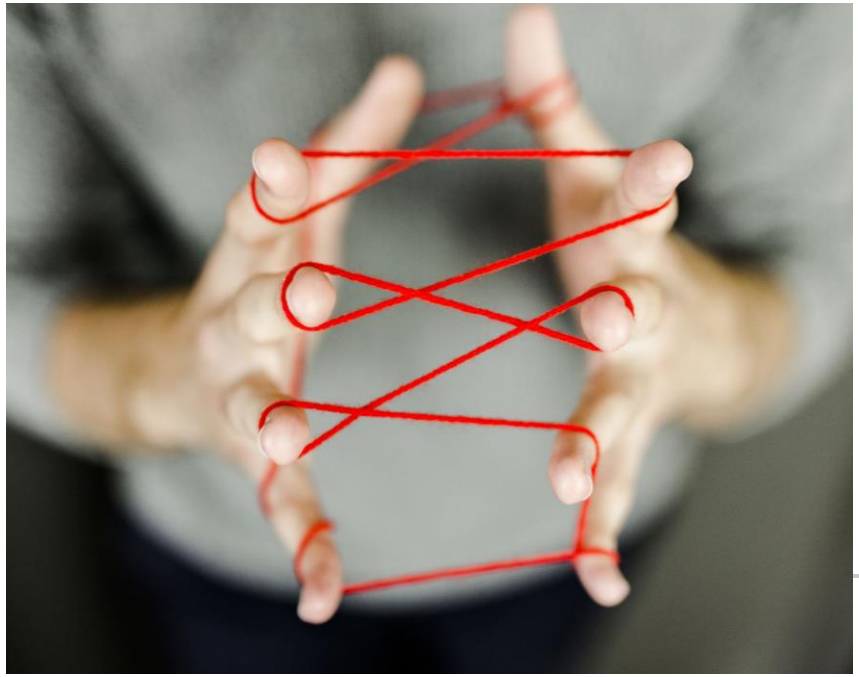
Data merging

Handling Columns (Numeric, Categorical, Binary)

One-Hot Encoding

Normalize

Features Correlation



+Pos Correlations

- G2: 0.93
- G1: 0.87
- Higher Education: 0.29
- Mother’s Education Level: 0.25
- Father’s Education Level: 0.20
- Study Time: 0.18

–Ne Correlations:

- Failures: -0.42
- Age: -0.16
- Travel Time: -0.17
- Go out: 0.-14
- Health: -0.13

Model Selection

Ridge Regression

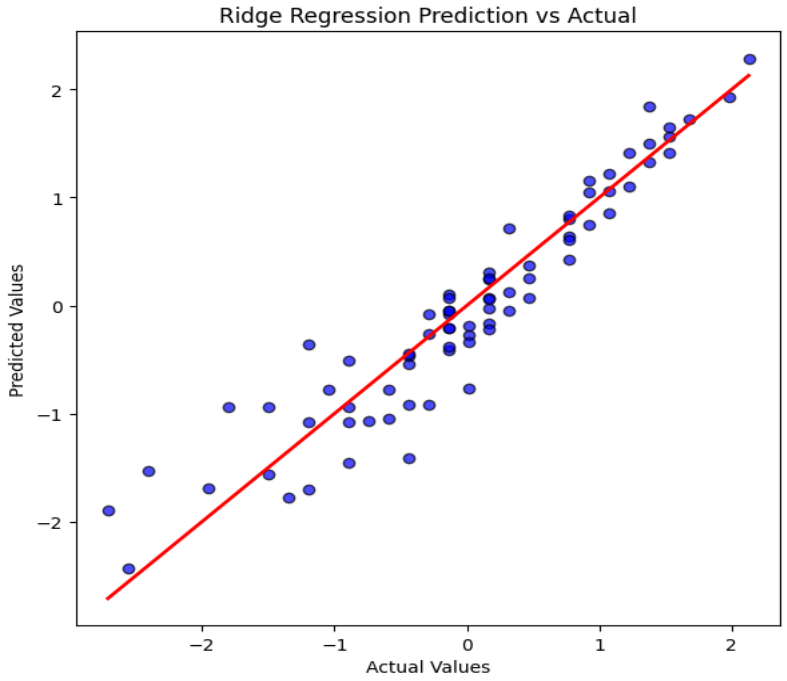
Decision Tree

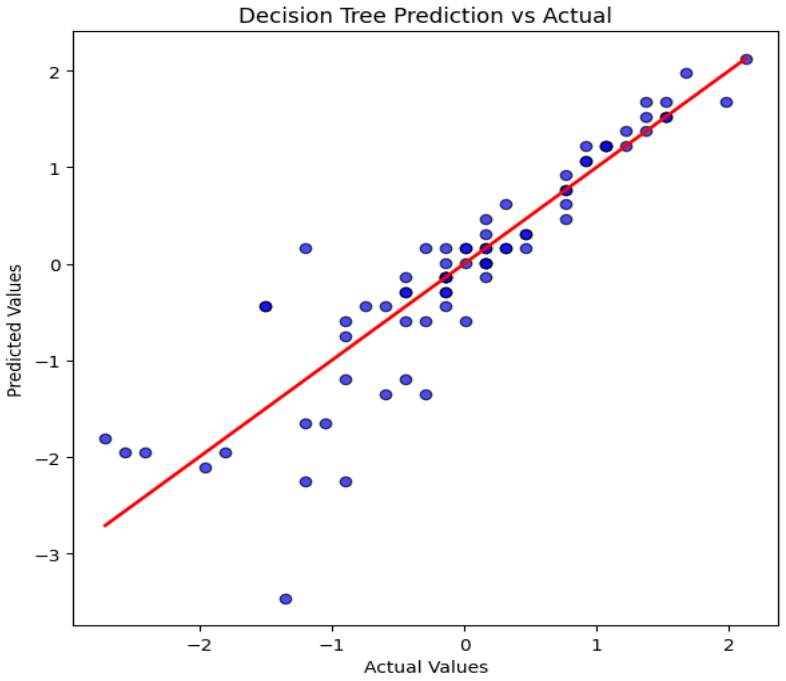
Random Forest

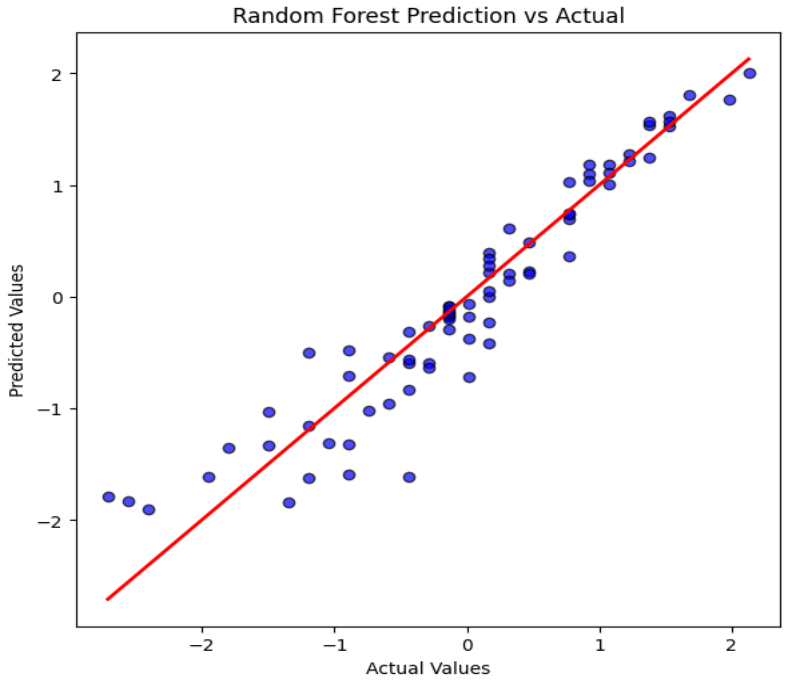
Gradient Boost

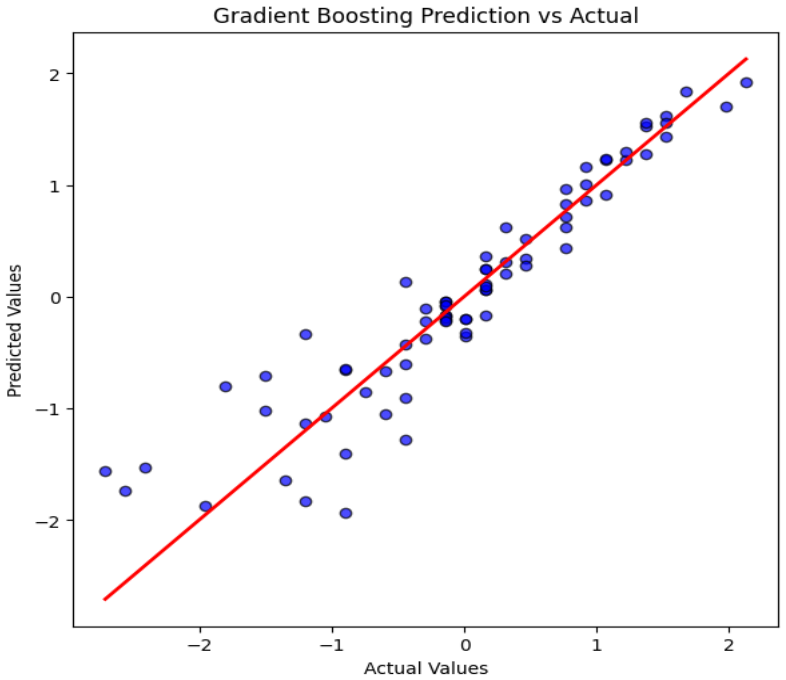
```
Ridge Regression Performance with Best CV (5 splits):
Test MSE: 0.1172
MAE: 0.2550
R2 Score: 0.8904
--------------------------------------------------
Decision Tree Performance with Best CV (4 splits):
Test MSE: 0.2598
MAE: 0.3262
R2 Score: 0.7569
--------------------------------------------------
Random Forest Performance with Best CV (4 splits):
Test MSE: 0.1086
MAE: 0.2357
R2 Score: 0.8984
--------------------------------------------------
Gradient Boosting Performance with Best CV (9 splits):
Test MSE: 0.1309
MAE: 0.2498
R2 Score: 0.8775
--------------------------------------------------
```

Models Performance

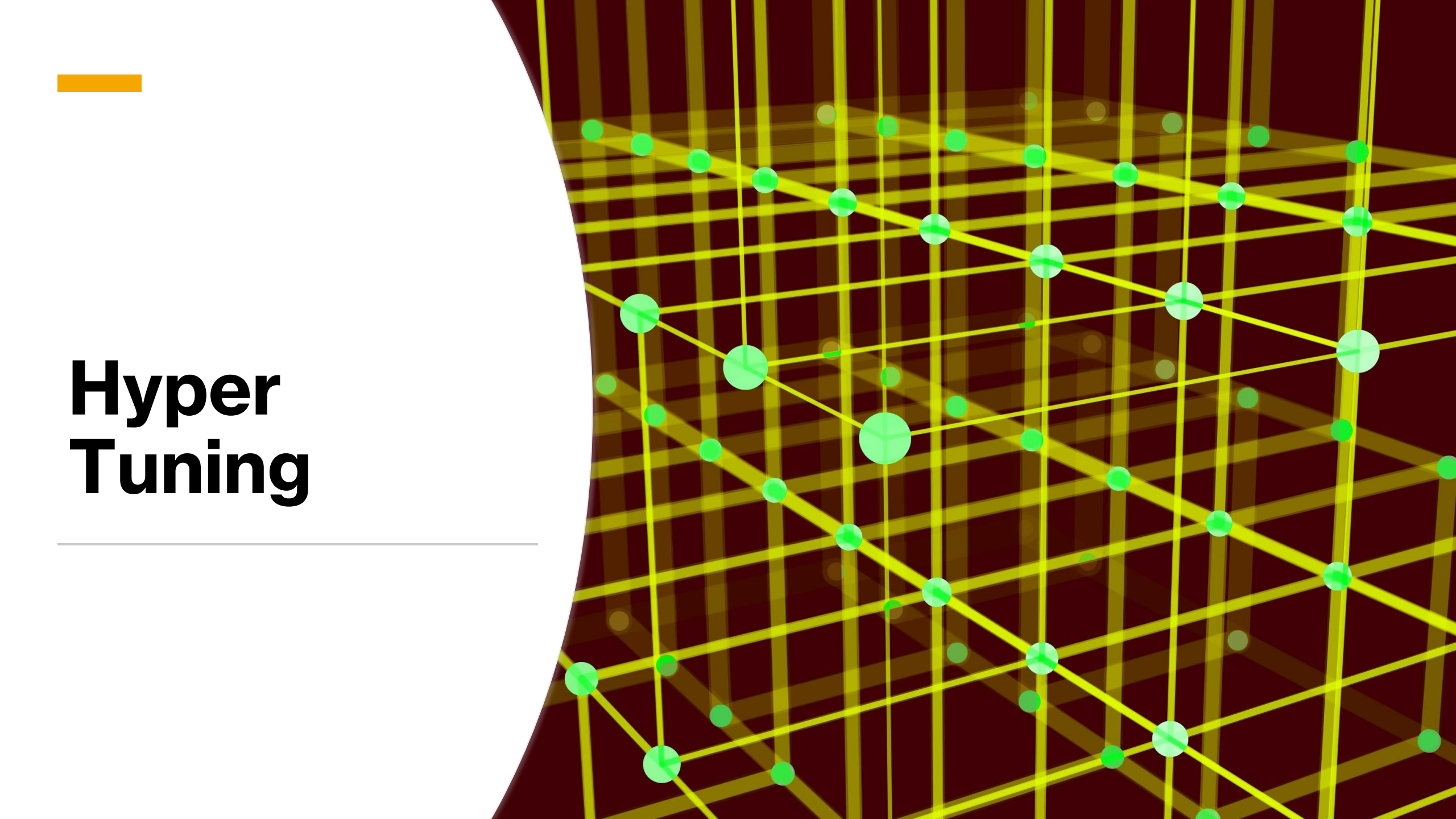

Ridge Regression Prediction vs Actual
Predicted Values
Actual Values


Decision Tree Prediction vs Actual
Predicted Values
Actual Values


Random Forest Prediction vs Actual
Predicted Values
Actual Values

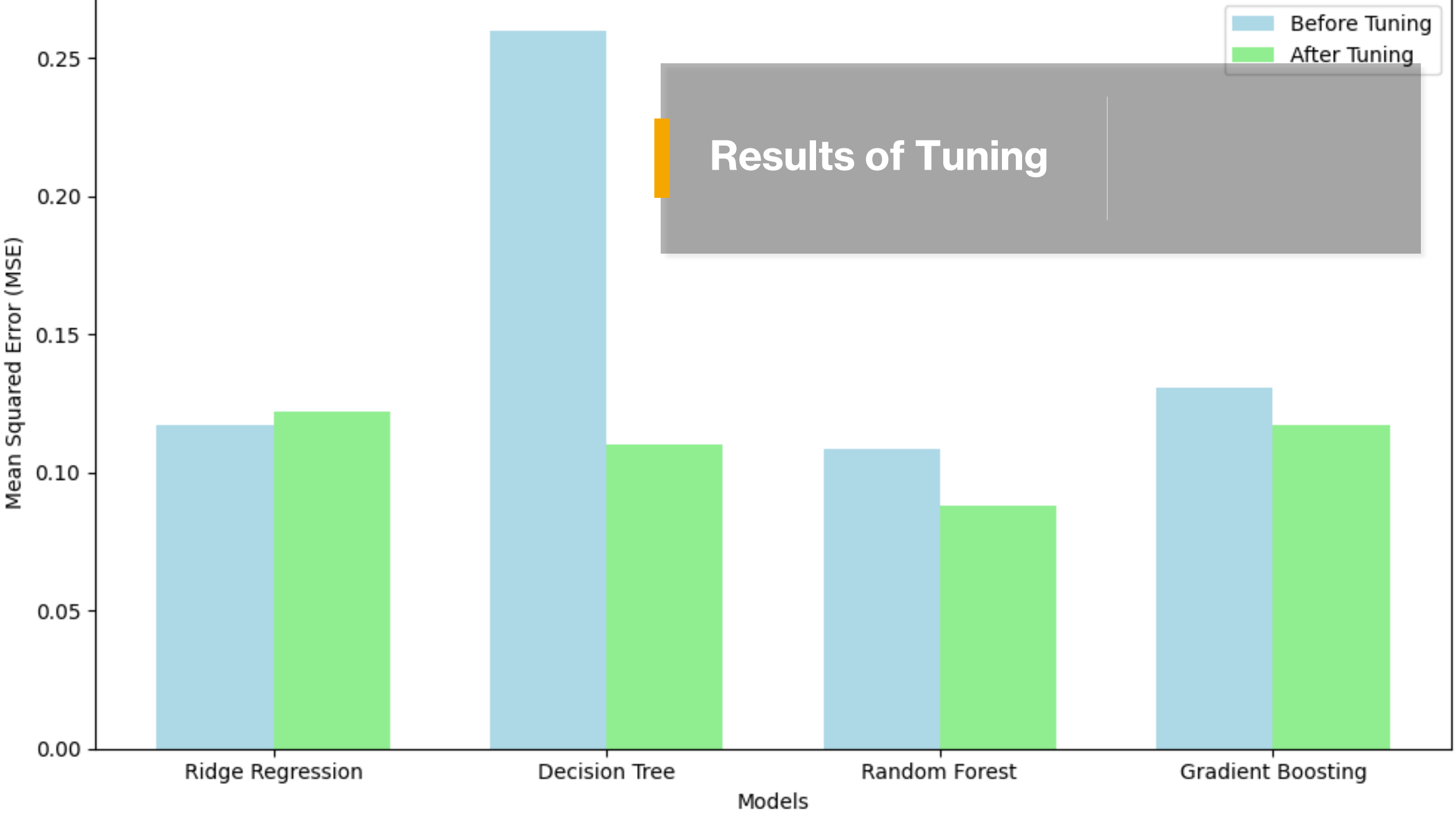

Gradient Boosting Prediction vs Actual
Predicted Values
Actual Values

Hyper Tuning



```
Tuning Ridge Regression...
Best Parameters for Ridge Regression: {'alpha': 10}
Best Cross-Validated MSE: 0.1494
Test MSE: 0.1223
MAE: 0.2571
R2 Score: 0.8856
--------------------------------------------------
Tuning Decision Tree...
Best Parameters for Decision Tree: {'max_depth': 3, 'min_samples_leaf': 1, 'min_samples_split': 2}
Best Cross-Validated MSE: 0.1552
Test MSE: 0.1102
MAE: 0.2436
R2 Score: 0.8969
--------------------------------------------------
Tuning Random Forest...
Best Parameters for Random Forest: {'max_depth': 10, 'min_samples_leaf': 4, 'min_samples_split': 2, 'n_estimators': 50}
Best Cross-Validated MSE: 0.1307
Test MSE: 0.0882
MAE: 0.2134
R2 Score: 0.9175
--------------------------------------------------
Tuning Gradient Boosting...
Best Parameters for Gradient Boosting: {'learning_rate': 0.1, 'max_depth': 3, 'min_samples_leaf': 4, 'min_samples_split': 2
_estimators': 50}
Best Cross-Validated MSE: 0.1337
Test MSE: 0.1170
MAE: 0.2368
R2 Score: 0.8905
--------------------------------------------------
```

Results of Tuning



Conclusion

Random Forest Regressor is the top performing model.

Hyper Tuning Significantly improved Decision Tree

Reference

Cortez, P. (2008). Student Performance [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5TG7T.