

House Price Prediction – Model Optimization Report

1. Introduction

- This project enhances a basic regression model by applying:
 - Feature Scaling
 - Multiple Model Training
 - Structured Performance Comparison
- Dataset used: California Housing Dataset.

2. Data Preprocessing

- Loaded dataset using scikit-learn
- Checked missing values (none found)
- Separated features and target
- Applied Standard Scaler to normalize data

3. Models Trained

Model	Purpose
Linear Regression	Baseline model
Ridge Regression	Reduces overfitting
Decision Tree	Captures non-linear relationships

4. Evaluation Metrics

- We used:
 - RMSE (Root Mean Squared Error)
 - R² Score
- Interpretation:
 - Lower RMSE → Better accuracy
 - Higher R² → Better explanatory power

5. Results Summary

Model	RMSE	R ²
Linear Regression	0.73	0.60
Ridge Regression	0.73	0.61
Decision Tree	0.70	0.64

- Decision Tree performed best.

6. Conclusion

- Feature scaling improved model stability.
- Ridge slightly improved Linear Regression.
- Decision Tree captured complex patterns better.
- Final Selected Model: Decision Tree Regressor

7. Future Improvements

- Hyperparameter tuning using GridSearchCV
- Try Random Forest / XGBoost
- Cross-validation
- Feature engineering (income per room, etc.)
- Remove outliers