# Random Forest Regressor

Generated documentation for the project found in the notebook.

## 1. Overview

This project performs house price prediction using multiple machine learning techniques. The notebook walks through data loading, exploration, preprocessing, model training, evaluation, and recommendation.

## 2. Dataset Details

Dataset: The notebook should indicate dataset source/file (e.g., train.csv, housing.csv). Look for file read commands such as pd.read\_csv(...).

Typical features include: numerical features (lot area, year built), categorical features (neighborhood), and the target variable 'SalePrice' or 'price'.

## 3. Exploratory Data Analysis (EDA)

EDA steps to include:

- Summary statistics (mean, median, std).

- Distribution plots for target variable and key features.

- Correlation matrix and pairwise relationships.

- Handling outliers and skewness.

## 4. Preprocessing

Common preprocessing steps detected in the notebook:

- Scaling

## 5. Models and Techniques

Models detected or commonly used in such notebooks:

- DecisionTree

- KNN

- Lasso

- LinearRegression

- RandomForest

- Ridge

## 6. Evaluation Metrics

Common metrics for regression:

- Mean Absolute Error (MAE)

- Root Mean Squared Error (RMSE)

- R-squared (R²)

## 7. Results Summary

Refer to the notebook for numeric results and comparison tables showing model performance. Generally, ensemble methods like RandomForest and XGBoost perform well for tabular regression tasks.

## 8. How to Run

1. Ensure you have Python 3.8+ installed.

2. Create and activate a virtual environment.

3. Install requirements: pip install -r requirements.txt

4. Launch the notebook and run cells sequentially.

## 9. Requirements

Typical packages include: pandas, numpy, scikit-learn, matplotlib, seaborn, xgboost, lightgbm (optional), python-docx (for this document).

## 10. Future Work

- Hyperparameter tuning with GridSearchCV or RandomizedSearchCV.

- Cross-validation and model ensembling.

- Feature engineering (interactions, polynomial features).

- Deploy model as an API or web app.

## 11. Notebook Notes & Appendix

Snippet of detected libraries and models from the notebook:

Libraries: matplotlib, numpy, pandas, seaborn, sklearn

Models: DecisionTree, KNN, Lasso, LinearRegression, RandomForest, Ridge