# House Price Prediction - Code Explanation

This document explains the steps involved in preprocessing the dataset, training the machine learning model, and making predictions for house price prediction using the Kaggle dataset.

## 1. Data Loading and Preprocessing

The dataset is loaded using pandas. The 'Id' column is removed as it is not useful for predictions. The target variable 'SalePrice' is separated from the features.

Missing values are handled separately for numerical and categorical columns. Numerical features are imputed using the median, while categorical features are filled with the most frequent value.

Categorical variables are then encoded using Label Encoding to convert text data into numerical form, which is necessary for machine learning algorithms.

## 2. Model Training

The dataset is split into training and validation sets. A Random Forest Regressor is used for training, which is an ensemble learning method based on decision trees.

The model is trained using 100 estimators, and its performance is evaluated using Mean Absolute Error (MAE), which measures the average error in predicted house prices.

## 3. Model Prediction

Once trained, the model is used to predict house prices on the test dataset. The predictions are saved in a CSV file ('submission.csv') for submission.

## 4. Conclusion

This project demonstrates a basic machine learning pipeline for house price prediction. Future improvements can be made by using advanced feature engineering, hyperparameter tuning, and other regression models.