

# Prediction of House Prices

## 1. Introduction:

- This task's main goal was to develop and assess machine learning models that could forecast home values based on a variety of characteristics.
- This included feature importance analysis, model development, evaluation, preprocessing, and data exploration.
- The findings show how well machine learning works to solve this practical issue and offer insightful information about the major variables affecting home values.

## 2. Data Exploration And Preprocessing:

- Data Loading: Dataset was loaded into the notebook using Pandas library.
- Descriptive Statistics: Understanding the distribution, dispersion, and central tendency of numerical features through the use of functions.
- Data Visualization: Use of scatter plots, boxplots, and histograms to identify trends in data, correlations between variables, and possible outliers.
- Categorical Variable Analysis: Analyzing categorical variables involves using cross-tabulations and bar charts to look at their frequency distribution.

- Missing Value Handling: Functions were used to thoroughly check for missing values. Thankfully, there were no missing values in the dataset.
- Duplicate Detection: Functions were used to find and eliminate duplicate instances. There were no duplicate entries discovered.
- Outlier Treatment: Winsorization, a rigorous technique that clips extreme values at designated percentiles.
- Feature Scaling: To guarantee a uniform range and keep features with higher values from taking over the model, Min-Max scaling was used to normalize numerical features.
- Data Splitting: Training, testing and splitting functions from scikit-learn were used to

separate the dataset into training and testing sets.

### **3. Model Development:**

Three machine learning models were created from the ground up:

- **Linear Regression:** Least squares optimization was used to create a basic linear regression model.
- **Random Forest Regressor:** A simple random forest model was constructed with modifiable parameters.
- **XGBoost Regressor:** A simple XGBoost model was put into practice, featuring modifiable parameters.

#### 4. Model Evaluation:

Two metrics were used to assess the model's performance:

- Root Mean Squared Error: The average difference between expected and actual values is measured by the Root Mean Squared Error, or RMSE.
- R-Squared : Indicates the percentage of the target variable's variance that the model can account for.

#### 5. Feature Importance Analysis:

- To determine which features were most important in predicting home prices, feature importance analysis was done for each model.
- This was accomplished either by employing feature importance scores for Random Forest or by examining the model's coefficients for Linear Regression and XGBoost.

## 6. Conclusion:

- For this task, a thorough process of developing and accessing machine learning models for predicting home prices was required.
- The knowledge gathered from feature importance analysis, model comparison, and data exploration is crucial for comprehending the variables affecting home prices.