**Lab Task 1**  
**Name :** Ali Afzal   
**Roll no. :** 102

**Report: House Price Prediction using Random Forest Regressor**

## **1. Introduction**

This report presents a machine learning approach to predicting house prices using the Random Forest Regressor. The dataset is analyzed, preprocessed, and used to train a predictive model that estimates house prices based on selected features.

## **2. Dataset Description**

The dataset used consists of housing attributes and their respective sale prices. The training data is loaded from train.csv, while test data is read from test.csv. The target variable is SalePrice, representing the house price.

## **3. Methodology**

### **3.1 Data Preprocessing**

* The dataset is loaded using pandas.
* A subset of relevant features is selected for training.
* The dataset is inspected using head() and info() to understand the structure and identify missing values.

### **3.2 Feature Selection**

The following features were selected for training the model:

* LotArea - Lot size in square feet
* YearBuilt - Year the house was built
* 1stFlrSF - First floor square footage
* 2ndFlrSF - Second floor square footage
* GrLivArea - Above-ground living area
* FullBath - Number of full bathrooms
* HalfBath - Number of half bathrooms
* BedroomAbvGr - Number of bedrooms above ground
* KitchenAbvGr - Number of kitchens above ground
* Fireplaces - Number of fireplaces
* PoolArea - Pool area in square feet
* TotRmsAbvGrd - Total rooms above ground

### **3.3 Model Training**

A RandomForestRegressor from sklearn.ensemble was used to train the model. The steps involved:

* Initializing the model with a fixed random state for reproducibility.
* Training the model using the selected features (X) and target variable (y).

### **3.4 Predictions**

* The test dataset is loaded.
* The same features as the training data are extracted.
* The trained model is used to predict house prices on the test dataset.

## **4. Results and Discussion**

* The model successfully predicts house prices based on selected features.
* Feature importance analysis can be conducted to determine which features have the highest impact on price prediction.
* Further improvements can be made by hyperparameter tuning or using advanced techniques such as boosting algorithms.

## **5. Conclusion**

This study demonstrates the effectiveness of the Random Forest Regressor in predicting house prices based on various housing features. The model provides valuable insights and can be further refined to improve prediction accuracy.

## **6. Future Work**

* Handling missing values and outliers more effectively.
* Experimenting with different regression models for comparison.
* Using additional features to improve model performance.

**Screenshots**  
