

# Predicting House Prices Using Machine Learning

## Introduction

Predicting house prices is a valuable task in real estate, finance, and investment. Machine learning can help us make accurate predictions based on historical data. In this document, we will guide you through the steps to build a house price prediction model.

### Step 1: Data Collection

Gather a dataset containing information on past house sales. This should include features like square footage, number of bedrooms, location, and sale prices.

### Step 2: Data Preprocessing

Handle missing data: Decide whether to impute or remove missing values.

Encode categorical variables: Convert categorical features (e.g., location) into

numerical form using techniques like one-hot encoding.

Split the data into training and testing sets: Typically, you'd use 70-80% of the data for training and the rest for testing.

### **Step 3: Feature Selection and Engineering**

Identify important features: Use techniques like correlation analysis or feature importance scores to select relevant features.

Create new features: Sometimes, creating derived features (e.g., price per square foot) can improve model performance.

### **Step 4: Model Selection**

Choose a regression algorithm: Popular choices include Linear Regression, Decision Trees, Random Forests, or Gradient Boosting.

Train multiple models: Experiment with different algorithms to find the best-performing one.

### **Step 5: Model Training**

Train the selected model using the training dataset .

Evaluate performance using appropriate metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE).

### **Step 6: Model Tuning**

Fine-tune hyperparameters: Adjust parameters like learning rate, tree depth, or regularization strength to optimize model performance.

### **Step 7: Model Evaluation**

Assess the model's performance on the testing dataset.

Visualize predictions vs. actual prices to understand how well the model is doing.

### **Step 8: Deployment**

Deploy the trained model into a production environment where it can make predictions on new data.

## **Step 9: Continuous Monitoring and Updating**

Regularly update the model with new data to keep it accurate over time.

Monitor for performance degradation and retrain as needed.

### **Conclusion**

Predicting house prices using machine learning is a valuable tool in the real estate industry. By following these steps and continually improving your model, you can make more accurate and informed predictions, aiding in property valuation and investment decisions

```
3 from sklearn.model_selection import
    train_test_split
4 from sklearn.linear_model import
    LinearRegression
5 from sklearn.metrics import
    mean_squared_error
6 data = pd.read_csv("house_data.csv")
7 X = data[["SquareFootage",
    "NumBedrooms", "Location"]]
8 y = data["Price"]
9 X_train, X_test, y_train, y_test =
    train_test_split(X, y, test_size=0
        .2, random_state=42)
10 model = LinearRegression()
11 model.fit(X_train, y_train)
12 y_pred = model.predict(X_test)
13 mse = mean_squared_error(y_test, y_pred
    )
14 print(f"Mean Squared Error: {mse}")
15 new_data = pd.DataFrame
    ({ "SquareFootage": [2000],
      "NumBedrooms": [3], "Location":
        ["Suburb"] })
16 predicted_price = model.predict
    (new_data)
17 print(f"Predicted Price: ")
```

Run