

Machine Learning Project: House Price Prediction

Project Objective

The goal of this project is to predict house prices based on features such as area, number of bedrooms, bathrooms, and age of the house. This is a regression machine learning problem because we are predicting numerical values.

Step 1: Install Required Libraries

```
pip install numpy pandas matplotlib seaborn scikit-learn
```

Step 2: Import Libraries

```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
```

Step 3: Create Dataset

```
data = {
    'Area': [1000, 1500, 1800, 2400, 3000],
    'Bedrooms': [2, 3, 3, 4, 4],
    'Bathrooms': [1, 2, 2, 3, 3],
    'Age': [10, 5, 8, 2, 1],
    'Price': [200000, 300000, 350000, 500000, 600000]
}

df = pd.DataFrame(data)
print(df.head())
```

Step 4: Explore Data

```
print(df.info())
print(df.describe())
```

Step 5: Define Features and Target

```
X = df[['Area', 'Bedrooms', 'Bathrooms', 'Age']]
y = df['Price']
```

Step 6: Split Dataset

```
from sklearn.model_selection import train_test_split

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, random_state=42)
```

Step 7: Train Model

```
from sklearn.linear_model import LinearRegression

model = LinearRegression()
model.fit(X_train, y_train)
```

Step 8: Make Predictions

```
y_pred = model.predict(X_test)
print("Predicted Prices:", y_pred)
```

Step 9: Evaluate Model

```
from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score

print("MAE:", mean_absolute_error(y_test, y_pred))
print("MSE:", mean_squared_error(y_test, y_pred))
print("R2 Score:", r2_score(y_test, y_pred))
```

Step 10: Predict Price for New House

```
new_house = [[2000, 3, 2, 5]]
predicted_price = model.predict(new_house)
print("Predicted Price for New House:", predicted_price[0])
```

What You Learned

- 1 How to handle data using Pandas
- 2 How to split data into training and testing sets
- 3 How to train a Linear Regression model
- 4 How to make predictions using the model
- 5 How to evaluate model performance