Project Title: Predicting House Rent Using Python

Short Description: Develop a model to estimate house rental prices based on features like size, location, and amenities, using Python and machine learning techniques.

Difficulty Level: Intermediate

AI-ML Concepts Used: Linear Regression, Feature Engineering

Learning Outcome: Predict house rents using data

Programming Skills Mastered: Data Preprocessing, Model Evaluation

---

Introduction

House rent prices are influenced by various factors such as location, size, number of rooms, and amenities. Accurately predicting rental prices can assist both tenants and landlords in making informed decisions. In this project, we will develop a machine learning model using Python to predict house rents based on specific features. This project is tailored for Class 9 and 10 CBSE/ICSE students, introducing data analysis and machine learning concepts.

Implementation Steps

1. **Import Necessary Libraries**

   Begin by importing essential Python libraries for data manipulation, visualization, and modeling.

   ```python
   CopyEdit
   import pandas as pd
   import numpy as np
   import matplotlib.pyplot as plt
   import seaborn as sns
   from sklearn.model_selection import train_test_split
   from sklearn.linear_model import LinearRegression
   from sklearn.metrics import mean_absolute_error
   ```

2. **Load the Dataset**

   Load the dataset containing information about house rentals. For this project, we will use a sample dataset.

   ```python
   CopyEdit
   # Sample data creation
   data = pd.DataFrame({
       'Size': [850, 900, 1200, 1500, 1750],
       'Location': ['CityA', 'CityB', 'CityA', 'CityC', 'CityB'],
       'BHK': [2, 2, 3, 3, 4],
       'Bathroom': [1, 2, 2, 3, 3],
       'Rent': [15000, 18000, 22000, 26000, 30000]
   })
   ```

3. **Data Preprocessing**
   - **Handle Categorical Variables:** Convert categorical variables like 'Location' into numerical values using one-hot encoding.

     ```python
     CopyEdit
     data = pd.get_dummies(data, columns=['Location'],
     drop_first=True)
     ```

   - **Feature Selection:** Define the independent variables (features) and the dependent variable (target).

     ```python
     CopyEdit
     X = data.drop('Rent', axis=1)
     y = data['Rent']
     ```

4. **Split the Dataset**

   Divide the dataset into training and testing sets to evaluate the model's performance.

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

5. **Train the Model**

   Utilize Linear Regression to train the model on the training data.

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

6. **Make Predictions**

   Predict rental prices on the testing set.

   ```python
   CopyEdit
   y_pred = model.predict(X_test)
   ```

7. **Evaluate the Model**

   Assess the model's performance using Mean Absolute Error (MAE).

   ```python
   CopyEdit
   mae = mean_absolute_error(y_test, y_pred)
   print(f"Mean Absolute Error: {mae}")
   ```

8. **Visualize the Results**

   Plot the actual vs. predicted rental prices to visualize the model's accuracy.

   ```python
   CopyEdit
   plt.scatter(y_test, y_pred)
   plt.xlabel('Actual Rent')
   plt.ylabel('Predicted Rent')
   plt.title('Actual vs. Predicted Rent')
   plt.show()
   ```

Conclusion

By completing this project, students will learn how to preprocess data, handle categorical variables, and apply linear regression to predict house rental prices. This hands-on experience will provide a foundational understanding of machine learning workflows and data-driven decision-making.