PREDICTING HOUSE PRICE USING MACHINE LEARNING

BATCH MEMBER

DILLIBABU V (511321106009)
DILLIBABU V (511321106008)
DINESH K (511321106011)
SANTHOSH KUMAR K (511321106021)

PROJECT TITLE: HOUSE PRICE PREDICTION

PHASE 3

TOPIC: Start building the house price prediction model by loading and pre-processing the dataset.

House Price Prediction

Introduction:

Whether you're a homeowner looking to estimate the value of your property, a real estate investor seeking profitable opportunities, or a data scientist aiming to build a predictive model, the foundation of this endeavor lies in loading and prepossessing the dataset.

Building a house price prediction model is a data-driven process that involves harnessing the power of machine learning to analyze historical housing data and make informed price predictions. This journey begins with the fundamental steps of data loading and prepossessing.

Necessary step to follow:

1. Import Libraries:

Start by importing the necessary libraries:

Program:

```
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import StandardScaler
```

2.Load the Dataset:

Load your dataset into a Pandas DataFrame. You can typically find house price datasets in CSV format, but you can adapt this code to other formats as needed.

Program:

```
df = pd.read_csv(' E:\USA_Housing.csv ')
Pd.read()
```

3. Exploratory Data Analysis (EDA):

Perform EDA to understand your data better. This includes checking for missing values, exploring the data's statistics, and visualizing it to identify patterns.

Program:

```
# Check for missing values

print(df.isnull().sum())

# Explore statistics

print(df.describe())

# Visualize the data (e.g., histograms, scatter plots, etc.)
```

4.Feature Engineering:

Depending on your dataset, you may need to create new features or transform existing ones. This can involve one-hot encoding categorical variables, handling date/time data, or scaling numerical features.

Program:

```
# Example: One-hot encoding for categorical variables
df = pd.get_dummies(df, columns=[' Avg. Area Income ', ' Avg. Area House Age ])
```

5.Split the Data:

Split your dataset into training and testing sets. This helps you evaluateyour model's performance later.

```
X = df.drop('price', axis=1) # Features

y = df['price']  # Target variable

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

6.Feature Scaling:

Apply feature scaling to normalize your data, ensuring that all features have similar scales. Standardization (scaling to mean=0 and std=1) is a common choice.

Program:

```
scaler = StandardScaler()
X_train = scaler.fit_transform(X_train)
X_test = scaler.transform(X_test)
```

Program:

```
import pandas as pd

import numpy as np

import seaborn as sns

import matplotlib.pyplot as plt

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import StandardScaler
```

```
from sklearn.metrics import
r2_score,
mean_absolute_error,mean_squar
ed_error

from sklearn.linear_model import

LinearRegressionfrom

sklearn.linear_model import Lasso

from sklearn.ensemble import

RandomForestRegressorfrom sklearn.svm

import SVR

import xgboost as xg

%matplotl

ib inline

import

warnings

warnings.filterwarnings("ignore")

/opt/conda/lib/python3.10/site-packages/scipy/__init__.py:146:
UserWarning: A NumPy version >=1.16.5 and <1.23.0 is required for this
version of SciPy (detected version 1.23.5
warnings.warn(f"A NumPy version >={np_minversion} and
<{np_maxversion}"
```

Loading Dataset:

```
dataset = pd.read_csv('E:/USA_Housing.csv')
```

Conclusion:

- In the quest to build a house price prediction model, we have embarked on a critical journey that begins with loading and preprocessing the dataset.We have traversed through essential steps, starting with importing the necessary libraries to facilitatedata manipulation and analysis.
- Understanding the data's structure, characteristics, and any potential issues through exploratory data analysis (EDA) is essential for informed decision-making.
- Data preprocessing emerged as a pivotal aspect of this process. It involves cleaning, transforming, and refining the dataset to ensurethat it aligns with the requirements of machine learning algorithms.