PREDICTING HOUSE PRICE AND MACHINE LEARNING

TEAM MEMBER

810621104007: BHAVANI.S

Phase-3 submission document

Project Title: House Price Prediction

Phase 3: Development Part1

Topic: start building the house price prediction model by using 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 preprocessing 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 preprocessing.

➢ This introduction will guide you through the initial steps of the process. Well explore how to import essential libraries, load the housing data, and perform critical preprocessing steps. Data preprocessing is crucial as it help clean , format, and prepare the data for further analysis. This includes handling missing values, encoding categorical variables, and ensuring that the data is appropriately scaled.

GIVEN DATA SET:

Avg. Area Income	Avg. Area House Age	Avg. Area Number of	Avg. Area Number of Bedrooms	Area Population	Price	Address

		Rooms				
79545.46	5.682861	7.009188	4.09	23086.8	1059034	208 Michael Ferry Apt. 674 Laurabury, NE 37010-5101
79248.64	6.0029	6.730821	3.09	40173.07	1505891	188 Johnson Views Suite 079 Lake Kathleen, CA 48958
61287.07	5.86589	8.512727	5.13	36882.16	1058988	9127 Elizabeth Stravenue Danieltown, WI 06482-3489
63345.24	7.188236	5.586729	3.26	34310.24	1260617	USS Barnett FPO AP 44820
59982.2	5.040555	7.839388	4.23	26354.11	630943.5	USNS Raymond FPO AE 09386
80175.75	4.988408	6.104512	4.04	26748.43	1068138	06039 Jennifer Islands Apt. 443 Tracyport, KS 16077
64698.46	6.025336	8.14776	3.41	60828.25	1502056	4759 Daniel Shoals Suite 442 Nguyenburgh, CO 20247
78394.34	6.98978	6.620478	2.42	36516.36	1573937	972 Joyce Viaduct Lake William, TN 17778-6483
59927.66	5.362126	6.393121	2.3	29387.4	798869.5	USS Gilbert FPO AA 20957

NECESSARY STEO 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 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

Printf(df.isnull()sum())

#Explore statistics

Printf(df.describe())

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

Data Preprocessing:

- ➢ **Incomplete data** can occur for a number of reasons. Attributes of interest may not always be available, such as customer information for sales transaction data. Relevant data may not be recorded due to a misunderstanding, or because of equipment malfunctions.

- ➢ There are many possible reasons for **noisy data** (having incorrect attribute values). The data collection instruments used may be faulty. There may have been human or computer errors occurring at data entry. Errors in data transmission can also occur. Incorrect data may also result from inconsistencies in Data preprocessing is a predominant step in machine learning to yield highly accurate and insightful results. Greater the quality of data, greater is the reliance on the produced results. **Incomplete, noisy, and inconsistent data** are the properties of large real-world datasets. Data preprocessing helps in increasing the quality of data by filling in missing incomplete data, smoothing noise and resolving inconsistencies.

- ➢ naming conventions or data codes used, or inconsistent formats for input fields, such as date.
- ➢ There are a number of data preprocessing techniques available such as,

- ➢ **Data Cleaning**
- ➢ **Data Integration**
- ➢ **Data Transformation**
- ➢ **Data Reduction**

- ➢ **Data cleaning** can be applied to filling in missing values, remove noise, resolving inconsistencies, identifying and removing outliers in the data.

- ➤ **Data integration** merges data from multiple sources into a coherent data store, such as a data warehouse.
- ➤ **Data transformations**, such as normalization, may be applied. For example, normalization may improve the accuracy and efficiency of mining algorithms involving distance measurements.
- ➤ **Data reduction** can reduce the data size by eliminating redundant features, or clustering, for instance.