5134-UNIVERSITY COLLEGE OF ENGINEERING KANCHEEPURAM

Kanchipuram -631 552

PREDICTION HOUSE PRICES USING MACHINE LEARNING

NAAN MUDHALVAN PROJECT

PREPARED BY :

J.DILLIRANI

REG NO :513421106009

ECE DEPARTMENT

3RD YEAR 5TH SEM

HOUSE PRICE PREDICTION USING MACHINE LEARNING

PROBLEM DEFINITION :

OBJECTIVE:

- People looking to buy a new home tend to be more conservative with their budgets and market strategies.

- This project aims to analyze various parameters like average income, average area etc. and predict the house price accordingly.

- This application will help customers to invest in an estate without approaching an agent

- To provide a better and fast way of performing operations. • To provide proper house price to the customers.

- To eliminate need of real estate agent to gain information regarding house prices.

- To provide best price to user without getting cheated.

- To enable user to search home as per the budget.

- The aim is to predict the efficient house pricing for real estate customers with respect to their budgets and priorities. By analyzing previous market trends and price ranges, and also upcoming developments future prices will be predicted.

- House prices increase every year, so there is a need for a system to predict house prices in the future.

- House price prediction can help the developer determine the selling price of a house and can help the customer to arrange the right time to purchase a house.

- We use Random forest regression algorithm in machine learning for predicting the house price trends

DATA SOURCE : Collection of data processing techniques and processes are numerous. We collected data for USA real estate properties from various real estate websites. The data would be having attributes such as Location, carpet area, built-up area, age of the property, zip code, price, no of bed rooms etc. We must collect the quantitative data which is structured and categorized. Data collection is needed before any kind of machine learning research is carried out. Dataset validity is a must otherwise there is no point in analyzing the data.

Data preprocessing :

Data preprocessing is the process of cleaning our data set. There might be missing values or outliers in the dataset. These can be handled by data cleaning. If there are many missing values in a variable we will drop those values or substitute it with the average value.

```
dataset = pd.read_csv('/kaggle/input/usa-housing/USA_Housing.csv')

dataset.info()

dataset.describe()
```

```
dataset.columns
```

VISUALISATION AND PRE PROCESSING DATA

```
sns.histplot(dataset, x='Price', bins=50, color='y')

sns.boxplot(dataset, x='Price',  palette='Blues')

sns.jointplot(dataset, x='Avg. Area House Age', y='Price', kind='hex')

sns.jointplot(dataset, x='Avg. Area Income', y='Price')

plt.figure(figsize=(12,8))
sns.pairplot(dataset)


dataset.hist(figsize=(10,8))

dataset.corr(numeric_only=True

plt.figure(figsize=(10,5))

sns.heatmap(dataset.corr(numeric_only = True), annot=True)
```

DIVIDE THE DATA SET

```
X = dataset[['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area
Number of Rooms',
'Avg. Area Number of Bedrooms', 'Area Population']]
Y = dataset['Price']
```

USING TRAIN TEST SPLIT

```
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=0.2,
random_state=101)
Y_train.head()

Y_train.shape

Y_test.head()

Y_test.shape

sc = StandardScaler()
```

```
X_train_scal = sc.fit_transform(X_train)
X_test_scal = sc.fit_transform(X_test)
```

MODEL BULIDING AND EVALUTION OF PREDICATED DATA

```
 model_lr=LinearRegression()

odel_lr.fit(X_train_scal, Y_train)

Prediction1 = model_lr.predict(X_test_scal)

plt.figure(figsize=(12,6))

 plt.plot(np.arange(len(Y_test)), Y_test, label='Actual Trend')
 plt.plot(np.arange(len(Y_test)), Prediction1, label='Predicted Trend')
 plt.xlabel('Data')
 plt.ylabel('Trend')
 plt.legend()
 plt.title('Actual vs Predicted')
 ns.histplot((Y_test-Prediction1), bins=50)
```

```
print(r2_score(Y_test, Prediction2))

print(mean_absolute_error(Y_test, Prediction2))
print(mean_squared_error(Y_test, Prediction2))
```

```
print(r2_score(Y_test, Prediction1))
print(mean_absolute_error(Y_test, Prediction1))
print(mean_squared_error(Y_test, Prediction1))
```

```
Model_rf = RandomForestRegressor(n_estimators=50)
```

```
 model_rf.fit(X_train_scal, Y_train)
```

CONCLUSION:

Thus the machine learning model to predict the house price based on given dataset is executed successfully using random forester (a upgraded/slighted boosted form of regular linear regression, this gives lesser error). This model further helps people understand whether this place is more suited for them based on heatmap correlation. It also helps people looking to sell a house at best time for greater profit. Any house price in any location can be predicted with minimum error by giving appropriate dataset.