PREDICTING HOUSE PRICE USING MACHINE LEARNING

TEAM MEMBER

510321104003 – DIVYA R

Phase-1 Project Submission

FIG : HOUSE PRICE PREDICTION

Abstract:

House price prediction is a complex task that is influenced by a variety of factors, including the property's location, size, condition, and amenities, as well as the overall market conditions. Machine learning algorithms can be used to develop predictive models that can estimate house prices with a high degree of accuracy.

Data Source

A data source for house price predicting to use machine learning should be Accurate, Complete the geographic area of interest, Accessible.

Dataset Link: **https://www.kaggle.com/datasets/vedavyasv/usa-housing**

Data collection and preparation:

This module involves collecting a dataset of historical sales data and preparing it for use in the machine learning model. This may involve cleaning the data, imputing missing values, and converting categorical variables to numerical variables.

Feature engineering:

This module involves creating new features from the existing data that are more informative for the machine learning model. For example, new features could be created to represent the property's proximity to schools, parks, and other amenities.

Model selection and training:

This module involves selecting a machine learning algorithm and training it on the prepared dataset. There are many different machine learning algorithms that can be used for house price prediction, such as linear regression, decision trees, random forests, and gradient boosting machines.

Model evaluation:

This module involves evaluating the performance of the trained model on a held-out dataset of unseen data. This helps to ensure that the model is able to generalize to new data.

Model deployment:

This module involves deploying the trained model to production so that it can be used to predict the prices of new properties. This may involve deploying the model to a web service or integrating it into a software application.

Modules for house price prediction using machine learning

The following are some of the key modules that can be used for house price prediction using machine learning:

- **Data collection and preparation:**
  - **Libraries:** NumPy, Pandas, SciPy
- **Feature engineering:**
  - **Libraries:** FeatureHasher, SelectKBest, OneHotEncoder
- **Model selection and training:**

- **Libraries:** scikit-learn (linear regression, decision trees, random forests, gradient boosting machines, etc.), TensorFlow, PyTorch

- **Model evaluation:**

- **Libraries:** scikit-learn (cross-validation, metrics, etc.)

- **Model deployment:**

- **Libraries:** Flask, Django, AWS SageMaker

Benefits and limitations of using machine learning for house price prediction

Benefits:

- Machine learning models can be very accurate in predicting house prices.

- Machine learning models can be used to predict the prices of new properties, even if there is no historical sales data for those properties.

- Machine learning models can be used to identify factors that influence house prices, which can be helpful for buyers and sellers.

Limitations:

- Machine learning models can be complex and difficult to interpret.

- Machine learning models are trained on historical data, so they may not be accurate in predicting house prices in a changing market.

- Machine learning models can be biased, depending on the data they are trained on.

Overall, machine learning can be a valuable tool for house price prediction. However, it is important to use machine learning models with caution and to be aware of their limitations.

PROGRAM:

```python
import numpy as np
import pandas as pd
from sklearn.linear_model import LinearRegression

# Load the house price dataset
df = pd.read_csv('house_price_dataset.csv')

# Prepare the data
X = df[['square_feet', 'bedrooms', 'bathrooms']]
y = df['price']
```

```
# Split the data into training and test sets
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.25)

# Train the linear regression model
model = LinearRegression()
model.fit(X_train, y_train)

# Evaluate the model on the test set
y_pred = model.predict(X_test)
rmse = np.sqrt(mean_squared_error(y_test, y_pred))
print('RMSE:', rmse)

# Deploy the model
# This could involve saving the model to a file or deploying it to
a web service
```

OUTPUT:

Dataset Preview:

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms \
0	59545.458574	5.65381	7.789188
1	69248.642545	6.066900	6.880821
2	61387.06359	5.855890	8.342727
3	63445.245546	7.788236	5.666729
4	55682.196626	5.047555	7.789388

	Avg. Area Number of Bedrooms	Area Population	Price \
0	4.09	23086.800503	1.059034e+06
1	3.09	40173.072174	1.505891e+06
2	5.13	36882.159400	1.058988e+06
3	3.26	34310.242831	1.260617e+06
4	4.23	26354.109472	6.309435e+05

                        Address
0 208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1 188 Johnson Views Suite 079\nLake Kathleen, CA...
2 9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3 USS Barnett\nFPO AP 44820 4 USNS Raymond\nFPO AE 09386

Preprocessed Data:
[[-0.19105816 -0.13226994 -0.13969293 0.12047677 -0.83757985 -1.00562872]
[-1.39450169 0.42786736 0.79541275 -0.55212509 1.15729018 1.61946754]
[-0.35137865 0.46394489 1.70199509 0.03133676 -0.32671213 1.63886651]

[-0.13944143 0.1104872   0.22289331 -0.75471601 -0.90401197 -1.54810704]
[0.62516685 2.20969666 0.42984356 -0.45488144 0.12566216 0.98830821]]
4227   1.034480e+06
4676   1.650389e+06
800   1.323172e+06
3671   1.077428e+06
4193   1.532887e+06
Name: Price, dtype: float64

CONCLUSION:

Machine learning can be a powerful tool for house price prediction. However, it is important to use machine learning models with caution and to be aware of their limitations. For example, machine learning models are trained on historical data, so they may not be accurate in predicting house prices in a changing market. Additionally, machine learning models can be biased, depending on the data they are trained on.

Overall, machine learning can be a valuable tool for house price prediction, but it should be used in conjunction with other factors, such as expert judgment and market research.