COLLEGE CODE: 9133

COURSE:Artificial intelligence

PHASE 4:Development Part 2

PROJECT TITLE: House price predictor

Team Members:

maalinivcse2021@gmail.com- Maalini.V

pavithrancse2021@gmail.com- Pavithra.N

divyadharshiniscse2021@gmail.com- Divyadharshini.S

manoranjithampcse2021@gmail.com- Manoranjitham.P

priyadharshiniscse2021@gmail.com- Priyadharshini.S

DATASET: https://www.kaggle.com/datasets/vedavyasv/usa-housin

TRAINING AND TESTING THE MODEL:

Model Training:

Train the selected model on the training data. The model learns patterns and relationships within the data during this process. The training process typically involves adjusting model parameters to minimize a loss function.

Model Evaluation:

After training, you need to evaluate the model's performance on the testing data. Common evaluation metrics include accuracy, precision, recall, F1-score for classification, and mean squared error, R-squared for regression.

"Splitting the dataset into training and testing set"

```
[19]: X_train,X_test,y_train,y_test=train_test_split(X,y,test_size=0.2) #20% of observations are for test and 80%
```

```
[20]: X_train
```

	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population
1411	71328.704441	6.621666	6.932756	4.36	28053.016397
908	80000.764172	5.612802	5.910013	3.17	42245.446401
4131	76901.904610	6.050385	7.565421	5.46	50668.141364
3675	69462.148920	3.986452	8.293966	5.16	46158.752246
2707	90493.598097	4.606563	6.532432	2.48	45211.008064
...	...	...	...	...	...
1055	69316.796889	6.300409	7.873576	4.28	24448.211461
3244	74727.245159	6.070088	5.585787	2.49	27372.305555
259	63183.889742	5.570562	7.541807	5.08	58691.200665
178	78379.412239	5.985697	5.681834	3.20	31357.388214
2784	65554.842521	5.092064	7.982764	6.47	39239.838253

4000 rows × 5 columns

```
[23]: linearmodel.fit(X_train,y_train) # training is taking place
```

```
[23]: LinearRegression()
```

```
[24]: test_pred=linearmodel.predict(X_test)
      train_pred=linearmodel.predict(X_train)
```

```
[25]: pd.DataFrame(test_pred,columns=["Predictions"]).head()
```

```
[25]:    Predictions
    0    1.438564e+06
    1    1.161466e+06
    2    1.460026e+06
    3    6.110389e+05
    4    1.387686e+06
```

```
[26]:
```

USA housing prediction - Linear Regressi...   Draft saved

File   Edit   View   Run   Add-ons   Help

+   🗑   ✂   ⧉   📋   ▷   ▷▷   Run All   Code ▾   ● Draft Session (8m)   ⏻   ↻   ⋮

```python
from sklearn.metrics import r2_score,mean_absolute_error,mean_squared_error
```

" Test prediction evaluation"

```python
[29]:   r_squared=r2_score(test_pred,y_test)
        print("R2 Score:", r_squared)
```

R2 Score: 0.9046796597914799

```python
[30]:   linearmodel.score(X_test,y_test)
```

[30]:   0.9140423945227004

```python
[31]:   mae=mean_absolute_error(y_test,test_pred)
        print("Mean Absolute Error (MAE):", mae)
```

Mean Absolute Error (MAE): 81023.44047681554

Models   ⌃

[+ Add Models]

No models added
Add a Kaggle model

Notebook options   ⌄

Schedule a notebook to run   ⌄

Code Help   ⌃

🔍 Find code help

USA housing prediction - Linear Regressi...   Draft saved

File   Edit   View   Run   Add-ons   Help

+   🗑   ✂   ⧉   📋   ▷   ▷▷   Run All   Code ▾   ● Draft Session (8m)   ⏻   ↻   ⋮

```python
        (np.dot(X_test,coef.transpose())+intercept)
```

```
[40]:  array([[1438564.40798403],
               [1161466.12056832],
               [1460026.18587627],
               [ 611038.86113306],
               [1387686.04009869],
               [ 513080.36497973],
               [1224117.61290795],
               [ 779130.42049552],
               [1212908.8601377 ],
               [1257728.14421955],
               [1107375.33113647],
               [1424933.36272296],
               [1880330.09111944],
               [1293467.06572774],
               [1227695.07801086],
               [1028919.6230989 ],
               [1324845.72239494],
               [1548918.57610611],
               [ 883897.17590746],
               [1096242.94629345],
               [1331480.18483404],
               [1584136.21941831],
               [1067960.96083832],
               [1118698.5715939 ],
               [2082407.37139755],
               [1026686.38365648],
               [1153551.2773749 ],
               [1568470.1086974 ],
               [1076947.13448431],
               [1330879.82644885],
               [1750747.42422742],
               [1014812.62871642],
               [1801926.55600264],
               [1620939.46498849],
               [1423668.23622426],
               [1100698.55612707],
               [1541929.98626004],
```

Models   ⌃

[+ Add Models]

No models added
Add a Kaggle model

Notebook options   ⌄

Schedule a notebook to run   ⌄

Code Help   ⌃

🔍 Find code help

File Edit View Run Add-ons Help

```
[42]: train_pred

[42]: array([[1269923.18564459],
             [1379662.72602339],
             [1714251.56302348],
             ...,
             [1457345.63232965],
             [1214536.60784077],
             [1188131.44764333]])

[43]: (np.dot(X_train,coef.transpose())+intercept)

[43]: array([[1269923.18564459],
             [1379662.72602339],
             [1714251.56302348],
             ...,
             [1457345.63232965],
             [1214536.60784077],
             [1188131.44764333]])

[44]: houses.corr()
```

[44]:	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price
Avg. Area Income	1.000000	-0.002007	-0.011032	0.019788	-0.016234	0.639734
Avg. Area House Age	-0.002007	1.000000	-0.009428	0.006149	-0.018743	0.452543
Avg. Area Number of Rooms	-0.011032	-0.009428	1.000000	0.462695	0.002040	0.335664

File Edit View Run Add-ons Help

```
[45]: houses.columns

[45]: Index(['Avg. Area Income', 'Avg. Area House Age', 'Avg. Area Number of Rooms',
             'Avg. Area Number of Bedrooms', 'Area Population', 'Price', 'Address'],
            dtype='object')

[46]: houses.head(7)
```

[46]:	Avg. Area Income	Avg. Area House Age	Avg. Area Number of Rooms	Avg. Area Number of Bedrooms	Area Population	Price	Address
0	79545.458574	5.682861	7.009188	4.09	23086.800503	1.059034e+06	208 Michael Ferry Apt. 674\nLaurabury, NE 3701...
1	79248.642455	6.002900	6.730821	3.09	40173.072174	1.505891e+06	188 Johnson Views Suite 079\nLake Kathleen, CA...
2	61287.067179	5.865890	8.512727	5.13	36882.159400	1.058988e+06	9127 Elizabeth Stravenue\nDanieltown, WI 06482...
3	63345.240046	7.188236	5.586729	3.26	34310.242831	1.260617e+06	USS Barnett\nFPO AP 44820
4	59982.197226	5.040555	7.839388	4.23	26354.109472	6.309435e+05	USNS Raymond\nFPO AE 09386
5	80175.754159	4.988408	6.104512	4.04	26748.428425	1.068138e+06	06039 Jennifer Islands Apt. 443\nTracyport, KS...
6	64698.463428	6.025336	8.147760	3.41	60828.249085	1.502056e+06	4759 Daniel Shoals Suite 442\nNguyenburgh, CO ...

CONCLUSION:

In conclusion, training and testing a machine learning model is a systematic and iterative process that involves collecting and preparing data, selecting an appropriate model, training and fine-tuning the model, and evaluating its performance. It is a critical step in creating predictive models for a wide range of tasks, from classification and regression to clustering and recommendation systems.