GO_STP_5856 - Ashwin S

Assignment-8

```python
import pandas as pd
import numpy as np
from sklearn import preprocessing
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error, mean_absolute_error, r2_score, accuracy_sc
```

```python
df = pd.read_csv("/content/50_Startups.csv")
```

```python
df.head()
```

	R&D Spend	Administration	Marketing Spend	State	Profit
0	165349.20	136897.80	471784.10	New York	192261.83
1	162597.70	151377.59	443898.53	California	191792.06
2	153441.51	101145.55	407934.54	Florida	191050.39
3	144372.41	118671.85	383199.62	New York	182901.99
4	142107.34	91391.77	366168.42	Florida	166187.94

```python
df.tail()
```

	R&D Spend	Administration	Marketing Spend	State	Profit
45	1000.23	124153.04	1903.93	New York	64926.08
46	1315.46	115816.21	297114.46	Florida	49490.75
47	0.00	135426.92	0.00	California	42559.73
48	542.05	51743.15	0.00	New York	35673.41
49	0.00	116983.80	45173.06	California	14681.40

```python
df.isnull().sum()
```

```
R&D Spend          0
Administration     0
Marketing Spend    0
State              0
Profit             0
dtype: int64
```

```
lae=LabelEncoder()
df["State"] = lae.fit_transform(df["State"].astype(str))


x = df[["R&D Spend","Administration","Marketing Spend","State"]]
y = df[["Profit"]]


X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2, random_state=0)
regressor = LinearRegression()
regressor.fit(X_train, y_train)
print("Training complete.")
```

```
    Training complete.
```

```
print("Training Score: ",regressor.score(X_train, y_train))
print("Test Score: ",regressor.score(X_test, y_test))
```

```
    Training Score:  0.9500009880362248
    Test Score:  0.9386861070938134
```

```
n1 = float(input("Enter the R&D Spend: "))
n2 = float(input("Enter the Administration Spend: "))
n3 = float(input("Enter the Marketing Spend: "))
n4 = int(input("Enter the State: "))
own_pred = regressor.predict([[n1,n2,n3,n4]])
print("The Profit based on the given information will be around: {:.2f}".format(own_pred[0
```

```
    Enter the R&D Spend: 165349.20
    Enter the Administration Spend: 136897.80
    Enter the Marketing Spend: 471784.10
    Enter the State: 2
    The Profit based on the given information will be around: 192416.49
```

```
print("Mean Squared Error: ",mean_squared_error(y_test, y_pred))
print("Mean Absolute Error: ",mean_absolute_error(y_test, y_pred))
print("Root Mean Squared Error: ",np.sqrt(mean_absolute_error(y_test, y_pred)))
print("R - 2: ",r2_score(y_test, y_pred))
```

```
    Mean Squared Error:  78413822.17201337
    Mean Absolute Error:  7400.616305036963
    Root Mean Squared Error:  86.02683479610862
    R - 2:  0.9386861070938134
```

0s completed at 6:19 PM