SPARKS FOUNDATION PROJECT DEC.2022

By. PANGULURI V V SRINIVASA PRANAVA SAI

TASK 1 Problem Statement

Predicting the performance of a student based on the no. of study hours, what will be the predicted score if the student studies for 9.25 hrs/day?

```
#importing libraries
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```
#loading dataset
df = pd.read_csv("http://bit.ly/w-data")
df.head()
```

	Hours	Scores
0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30

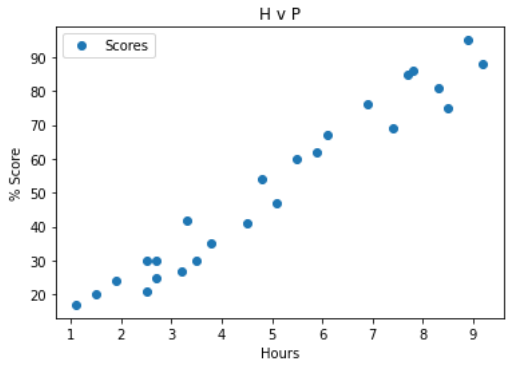
```
#checking the datatypes
df.dtypes
```

```
Hours     float64
Scores      int64
dtype: object
```

```
#shape of the data
df.shape
```

```
(25, 2)
```

```
#plotting
df.plot(x = 'Hours', y = 'Scores', style = 'o')
plt.title('H v P')
plt.xlabel('Hours')
plt.ylabel('% Score')
plt.show()
```




```
#declaring variables
X = df.iloc[:, :-1].values
y = df.iloc[:, 1].values
```

```
print(X.shape, y.shape)
```

```
(25, 1) (25,)
```

```
#splitting dataset into train and test set
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size = 0.2, random_state=0)
```

```
#LinearRegression
from sklearn.linear_model import LinearRegression
regressor = LinearRegression()
regressor.fit(X_train, y_train)
```

```
LinearRegression()
```

```
y_pred = regressor.predict(X_test)
```

```
print(X_test)
```

```
[[1.5]
 [3.2]
 [7.4]
 [2.5]
 [5.9]]
```

```
#Comparing actual and predicted values
df = pd.DataFrame({'Actual': y_test, 'Predicted': y_pred})
df
```

	Actual	Predicted
0	20	16.884145
1	27	33.732261
2	69	75.357018
3	30	26.794801
4	62	60.491033

```
#checking the model accuracy
print(regressor.score(X_test, y_test))
```

```
0.9454906892105354
```

```
#performance metrics
from sklearn.metrics import mean_absolute_error
from sklearn.metrics import mean_squared_error
print(mean_absolute_error(y_test, y_pred))
print(mean_squared_error(y_test, y_pred))
```

```
4.183859899002982
21.598769307217456
```

```
#predicting the required model
print(regressor.predict([[9.25]]))
```

```
[93.69173249]
```

✓ 0s completed at 19:47