

# TSF TASK 1 - GRIP MARCH'21

ADITYA AMBWANI

## Prediction Using Supervised ML

**Aim:-** To predict the percentage of students based on number of study hours.

The Sparks Foundation GRIP MARCH'21

```
In [7]: # importing libraries
import numpy as np
import pandas as pan
import matplotlib.pyplot as mat
```

```
In [8]: # Reading Data
data = pan.read_csv('https://raw.githubusercontent.com/AdiPersonalWorks/Random/master/s
```

```
In [9]: # Calculating number of rows and coloumns
data.shape
```

Out[9]: (25, 2)

```
In [10]: # No. of rows = 25
# No. of coloumns = 2
# Displaying dataset in tabular form
data.head(25)
```

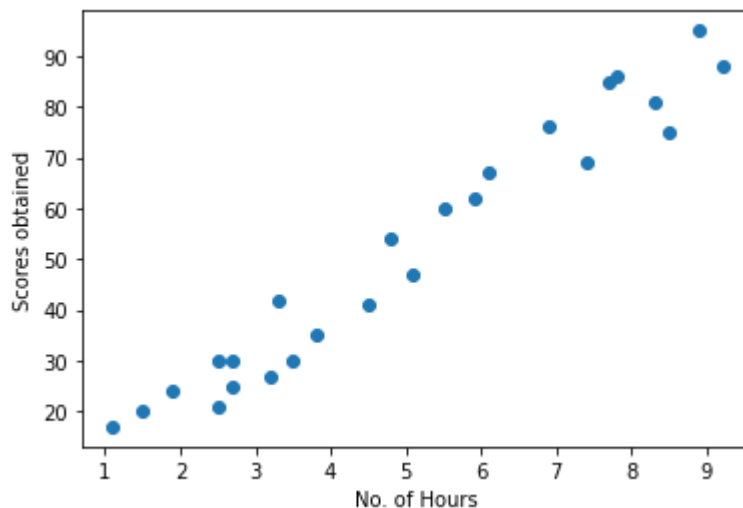
Out[10]:

	Hours	Scores
--	-------	--------

0	2.5	21
1	5.1	47
2	3.2	27
3	8.5	75
4	3.5	30
5	1.5	20
6	9.2	88
7	5.5	60
8	8.3	81
9	2.7	25
10	7.7	85
11	5.9	62

	Hours	Scores
12	4.5	41
13	3.3	42
14	1.1	17
15	8.9	95
16	2.5	30
17	1.9	24
18	6.1	67
19	7.4	69
20	2.7	30
21	4.8	54
22	3.8	35
23	6.9	76
24	7.8	86

```
In [11]: # Plotting the data on a graph
# Scatter plot
mat.scatter(x='Hours', y='Scores', data=data)
mat.xlabel('No. of Hours')
mat.ylabel('Scores obtained')
mat.show()
```



RESULT:- As the no. of hours increase, the marks obtained by student increase. So study more!!!

```
In [22]: x=data.iloc[:, :-1]
y=data.iloc[:, 1]
```

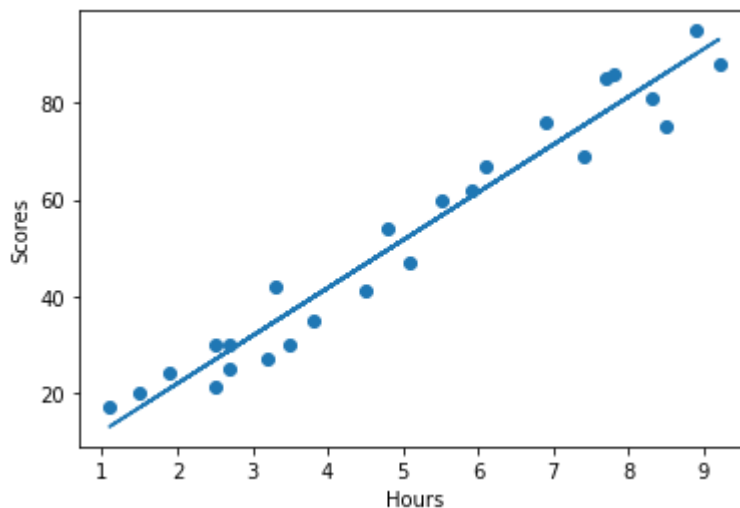
```
In [31]: # Splitting our data set into training and testing
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(x, y, test_size = 0.2, random_state

# Linear regression on our trained dataset
```

```
In [32]: from sklearn.linear_model import LinearRegression
lr = LinearRegression()
lr.fit(x_train,y_train)
```

Out[32]: LinearRegression()

```
In [35]: # Plotting linear regression on our actual plot
mat.scatter(x,y)
mat.plot(x,lr.coef_*x +lr.intercept_)
mat.xlabel('Hours')
mat.ylabel('Scores')
mat.show()
```



```
In [36]: #Prediction of the scores obtained
y_pd = lr.predict(x_test)
```

```
In [37]: #Comparison
y_pdfinal = pan.DataFrame({'Actual Scores':y_test,'Predicted Scores':y_pd})
y_pdfinal.head()
```

```
Out[37]:
```

	Actual Scores	Predicted Scores
5	20	16.884145
2	27	33.732261
19	69	75.357018
16	30	26.794801
11	62	60.491033

```
In [39]: # Evaluating the trained dataset on giving hours studied and obtainig the marks
# Hours studied = 9.25/day
hr = [[9.25]]
marks = lr.predict(hr)
res = (marks[0])
print(res)
```

93.69173248737538

RESULT:- The score obtained is 93.69 if a student studies 9.25hrs/day

In [ ]: