

****Data Science And Bussiness Analytics Intern At TheSparksFoundation****
(<https://www.thesparksfoundationsingapore.org/>)

GRIPJAN21

Author: Rajat Kumar

Task 1:-Predicting Using Supervised ML

Problem Stepment:- What Will be predicted score of a student if she/he studies for 9.25 hours/day

****First Step:- Importing Standard ML libraries****

```
In [116... import pandas as pd

In [117... import numpy as np

In [118... import matplotlib.pyplot as plt
%matplotlib inline
```

****Step second:- Reading Data set given by The Sparks Foundation****

```
In [119... data = pd.read_csv("https://raw.githubusercontent.com/AdiPersonalWorks/Random/master/student_scores%20-%20student_scores.csv")
data.head()

Out[119...
  Hours  Scores
0    2.5     21
1    5.1     47
2    3.2     27
3    8.5     75
4    3.5     30
```

****Step third:- Visualization of data using scater Plot****

```
In [120... plt.scatter(x=data['Hours'],y=data['Scores'])
plt.title("Hours vs Scores (simple linear regression)")
plt.xlabel("Hours")
plt.ylabel("Scores")
plt.show()
```



****Step fourth:- Splitting Data variables into target and features Here my feature variable will be number of study hours and target variable is scores of students.****

```
In [121... x=data.iloc[:, :-1].values
y=data.iloc[:, 1].values
print("x values:",x)
print("y values:",y)

x values: [[2.5]
[5.1]
[3.2]
[8.5]
[3.5]
[1.5]
[9.2]
[5.5]
[8.3]
[2.7]
[7.7]
[5.9]
[4.5]
[3.3]
[1.1]
[8.9]
[2.5]
[1.9]
[6.1]
[7.4]
[2.7]
[4.8]
[3.8]
[6.9]
[7.8]]
y values: [21 47 27 75 30 20 88 60 81 25 85 62 41 42 17 95 30 24 67 69 30 54 35 76
86]
```

****Step five:- split data set into train and test sets my text size is 0.3.****

```
In [122... from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test = train_test_split(x,y,test_size = 0.3,random_state=0)
print(x_train.shape)
print(x_test.shape)

(17, 1)
(8, 1)
```

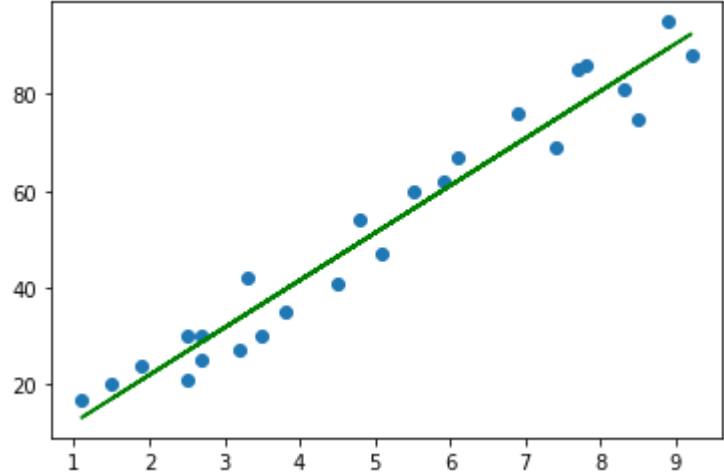
****Step six:- Now i am training simple linear regression model****

```
In [123... from sklearn.linear_model import LinearRegression
regressor=LinearRegression()
regressor.fit(x_train,y_train)

Out[123... LinearRegression()
```

****Step seventh:- Visualization the line of fit after linear regression.****

```
In [124... plt.scatter(x,y)
plt.plot(x,regressor.predict(x),color= 'green' )
plt.show()
```



****Step Eight:- Predicting score for 9.25 hrs/day of study****

```
In [125... pred=regressor.predict([[9.25]])
print("prediction of score for 9.25 is :",pred[0])

prediction of score for 9.25 is : 92.91505723477056
```

****Step nine:- Predicting value for other set value****

```
In [126... y_pred = regressor.predict(x_test)
print("predicted:",y_pred)
print("Actual",y_test)

predicted: [17.05366541 33.69422878 74.80620886 26.8422321 60.12335883 39.56736879
20.96909209 78.72163554]
Actual [20 27 69 30 62 35 24 86]
```

****Step Final:- Evaluating the my model****

```
In [127... from sklearn.metrics import r2_score
print("R-squared score:",r2_score(y_test,y_pred))

R-squared score: 0.9568211104435257
```

****Thank you****

```
In [ ]:
```