The Spark Foundation

Supervised Learning

Task 1

- predicted score if a student studies for 9.25 hrs/day

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:
```
data = pd.read_csv('C:/Users/ADMIN\OneDrive/Desktop/spark intenship/Task 1-linear regression/studytym_vs_marks.csv'
, header = 0)
```

In [3]:
```
data.head()
```

Out[3]:

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

In [4]:
```
data.describe()
```

Out[4]:

	Hours	Scores
count	25.000000	25.000000
mean	5.012000	51.520000
std	2.525094	25.244009
min	1.100000	17.000000
25%	2.700000	30.000000
50%	4.800000	47.000000
75%	7.400000	75.000000
max	9.200000	95.000000

In [5]:
```
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 25 entries, 0 to 24
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   Hours   25 non-null     float64
 1   Scores  25 non-null     int64  
dtypes: float64(1), int64(1)
memory usage: 464.0 bytes
```

In [6]:
```
X = data.iloc[: , 0:1].values
Y = data.iloc[: , 1].values
```

Using Train_Test_Split Package we split dataset into two parts

In [7]:
```
from sklearn.model_selection import train_test_split
X_train , X_test , Y_train , Y_test = train_test_split(X , Y , test_size = 0.20 , random_state = 0)
```

Import Linear Regression package from sklearn.linear_model

In [9]:
```
from sklearn.linear_model import LinearRegression
```

In [10]:
```
model = LinearRegression()
```

In [11]:
```
model.fit(X_train , Y_train)
```

Out[11]:
```
LinearRegression(copy_X=True, fit_intercept=True, n_jobs=None, normalize=False)
```

In [12]:
```
Y_predict = model.predict(X_test)
```

In [14]:
```
Y_predict
```

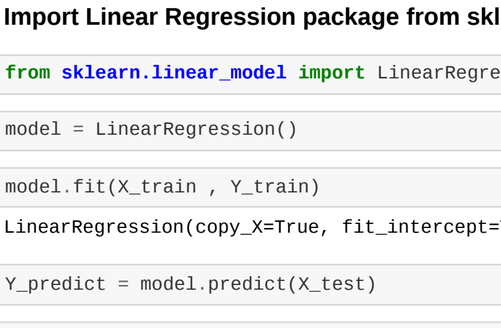
Out[14]:
```
array([17.01102227, 33.82419397, 75.36261815, 26.90112327, 60.52746666])
```

using matplotlib draw Scatter plot

In [15]:
```
plt.scatter(X_train , Y_train , color = 'red')
plt.plot(X_train , model.predict(X_train) , color = 'blue')
plt.xlab('Score of student')
plt.plot('Hours study by a student')
plt.show()
```

```
---------------------------------------------------------------------------
AttributeError                            Traceback (most recent call last)
<ipython-input-15-b14e14e95e8d> in <module>
      1 plt.scatter(X_train , Y_train , color = 'red')
      2 plt.plot(X_train , model.predict(X_train) , color = 'blue')
----> 3 plt.xlab('Score of student')
      4 plt.plot('Hours study by a student')
      5 plt.show()

AttributeError: module 'matplotlib.pyplot' has no attribute 'xlab'
```



1. redicted score if a student studies for 9.25 hrs/day

In [16]:
```
model.predict([[9.25]])
```

Out[16]:
```
array([93.659305])
```

In [20]:
```
model.predict([[9.25]])[0]
```

Out[20]:
```
93.65930499805776
```

In [ ]:
```

```