

```
In [21]: # Predict Student Marks on the Basis of Number of Hours Studied

# Load the Dataset
import pandas as pd
data = pd.read_csv('C:\\Users\\Apeh\\Desktop\\CODE\\DATASET\\Grade_Set_1.csv')
data
```

```
Out[21]:
```

	Hours_Studied	Test_Grade	Status	Result
0	2	57	fail	D
1	3	66	fail	D
2	4	73	pass	C
3	5	76	pass	C
4	6	79	pass	C
5	7	81	pass	B
6	8	90	pass	B
7	9	96	pass	A
8	10	100	pass	A

```
In [22]: # Explore Data
data.columns
```

```
Out[22]: Index(['Hours_Studied', 'Test_Grade', 'Status', 'Result'], dtype='object')
```

```
In [23]: data.shape
```

```
Out[23]: (9, 4)
```

```
In [24]: # Missing Value?
data.isnull().sum()
```

```
Out[24]: Hours_Studied    0
Test_Grade              0
Status                  0
Result                  0
dtype: int64
```

```
In [25]: # Convert the Categorical Value
from sklearn.preprocessing import LabelBinarizer
lb = LabelBinarizer()
data.Status = lb.fit_transform(data.Status)
```

```
In [26]: data
```

```
Out[26]:
```

	Hours_Studied	Test_Grade	Status	Result
--	---------------	------------	--------	--------

	Hours_Studied	Test_Grade	Status	Result
0	2	57	0	D
1	3	66	0	D
2	4	73	1	C
3	5	76	1	C
4	6	79	1	C
5	7	81	1	B
6	8	90	1	B
7	9	96	1	A
8	10	100	1	A

```
In [27]: # Dependent & Independent Variables  
import numpy as np  
x = data.Hours_Studied.values  
x = x.reshape(9,1)
```

```
In [28]: x.shape  
x
```

```
Out[28]: array([[ 2],  
                [ 3],  
                [ 4],  
                [ 5],  
                [ 6],  
                [ 7],  
                [ 8],  
                [ 9],  
                [10]], dtype=int64)
```

```
In [29]: y = data.Test_Grade.values  
y
```

```
Out[29]: array([ 57,  66,  73,  76,  79,  81,  90,  96, 100], dtype=int64)
```

```
In [30]: # Train the Dataset  
from sklearn.linear_model import LinearRegression  
lin_reg = LinearRegression()
```

```
In [31]: lin_reg.fit(x,y)
```

```
Out[31]: LinearRegression()
```

```
In [32]: pred_val = lin_reg.predict(x)  
pred_val
```

```
Out[32]: array([59.71111111, 64.72777778, 69.74444444, 74.76111111, 79.77777778,
      84.79444444, 89.81111111, 94.82777778, 99.84444444])
```

```
In [34]: # Compare the predicted val with original val
data['predicted_values'] = pred_val
data[['Hours_Studied', 'predicted_values', 'Test_Grade']]
```

```
Out[34]:
```

	Hours_Studied	predicted_values	Test_Grade
0	2	59.711111	57
1	3	64.727778	66
2	4	69.744444	73
3	5	74.761111	76
4	6	79.777778	79
5	7	84.794444	81
6	8	89.811111	90
7	9	94.827778	96
8	10	99.844444	100

```
In [35]: # Evaluate Model Performance
from sklearn.metrics import r2_score
accuracy = r2_score(y, pred_val)
print('Accuracy : ', accuracy)
```

Accuracy : 0.9757431074095347

```
In [36]: # Final Prediction
hrs = float(input('How many hours have you studied? : '))
marks = lin_reg.predict([[hrs]])
print('Student who studies for', hrs,
      'hours will going to score', marks, 'marks.')
```

How many hours have you studied? : 4.5
Student who studies for 4.5 hours will going to score [72.25277778] marks.