Experiment:-1

Objective:- Linear Regration

```
import numpy as np
import pandas as pd
```

```
from google.colab import drive
drive.mount('/content/drive/')
```

```
Mounted at /content/drive/
```

```
series = pd.read_csv('/content/drive/MyDrive/annual_csv.csv',index_col = 'Date')
```

```
series
```

Date	Price
1950-12	34.720
1951-12	34.660
1952-12	34.790
1953-12	34.850
1954-12	35.040
...	...
2015-12	1068.317
2016-12	1152.165
2017-12	1265.674
2018-12	1249.887
2019-12	1480.025

70 rows × 1 columns

```
series.head()
```

Date	Price
1950-12	34.72
1951-12	34.66
1952-12	34.79
1953-12	34.85
1954-12	35.04

```
series.tail()
```

```
                Price
Date
```

```
trainingSize = int(len(series) * 0.75)
```

```
2018-12  1152.165
```

```
testingSize = int(len(series) - trainingSize)
```

```
trainingSize, testingSize
```

```
(52, 18)
```

```
train,test = series[0:trainingSize],series[trainingSize:len(series)]
```

```
len(test), len(train)
```

```
(18, 52)
```

```
from os import set_inheritable
def create_dataset(series, time_steps = 3):
  Xs, Ys = [], []
  for i in range(len(series) - time_steps):
    v = series.iloc[i:(i + time_steps)].values
    Xs.append(v)
    Ys.append(series['Price'].iloc[i+time_steps])
  return np.array(Xs),np.array(Ys)
```

```
series.shape
```

```
(70, 1)
```

```
time_steps = 3
X_train, Y_train = create_dataset(train, time_steps)
X_test, Y_test = create_dataset(test, time_steps)

print(X_train.shape, Y_train.shape)
```

```
(49, 3, 1) (49,)
```

```
x_tr = X_train.reshape(len(X_train),3)
x_t = X_test.reshape(len(X_test),3)
print(x_tr.shape)
```

```
(49, 3)
```

```
print(X_train)
```

```
[[[ 34.72 ]
  [ 34.66 ]
  [ 34.79 ]]

 [[ 34.66 ]
  [ 34.79 ]
  [ 34.85 ]]

 [[ 34.79 ]
  [ 34.85 ]
  [ 35.04 ]]

 [[ 34.85 ]
  [ 35.04 ]
  [ 34.97 ]]

 [[ 35.04 ]
  [ 34.97 ]
  [ 34.9  ]]

 [[ 34.97 ]
  [ 34.9  ]
  [ 34.99 ]]

 [[ 34.9  ]
  [ 34.99 ]
```

```
         [ 35.09 ]]

        [[ 34.99 ]
         [ 35.09 ]
         [ 35.05 ]]

        [[ 35.09 ]
         [ 35.05 ]
         [ 35.54 ]]

        [[ 35.05 ]
         [ 35.54 ]
         [ 35.15 ]]

        [[ 35.54 ]
         [ 35.15 ]
         [ 35.08 ]]

        [[ 35.15 ]
         [ 35.08 ]
         [ 35.08 ]]

        [[ 35.08 ]
         [ 35.08 ]
         [ 35.12 ]]

        [[ 35.08 ]
         [ 35.12 ]
         [ 35.13 ]]

        [[ 35.12 ]
         [ 35.13 ]
```

```
print(Y_train)
```

```
     [ 34.85   35.04   34.97   34.9    34.99   35.09   35.05   35.54   35.15
       35.08   35.08   35.12   35.13   35.18   35.19   41.113  35.189  37.434
       43.455  63.779 106.236 183.683 139.279 133.674 160.48  207.895 463.666
      596.712 410.119 444.776 388.06  319.622 321.985 391.595 487.079 419.248
      409.655 378.161 361.875 334.657 383.243 379.48  387.445 369.338 288.776
      291.357 283.743 271.892 275.992]
```

```
X_test
```

```
     array([[[ 333.3  ],
             [ 407.674],
             [ 442.974]],

            [[ 407.674],
             [ 442.974],
             [ 509.423]],

            [[ 442.974],
             [ 509.423],
             [ 629.513]],

            [[ 509.423],
             [ 629.513],
             [ 803.618]],

            [[ 629.513],
             [ 803.618],
             [ 819.94 ]],

            [[ 803.618],
             [ 819.94 ],
             [1135.012]],

            [[ 819.94 ],
             [1135.012],
             [1393.512]],

            [[1135.012],
             [1393.512],
             [1652.725]],

            [[1393.512],
             [1652.725],
             [1687.342]],

            [[1652.725],
             [1687.342],
             [1221.588]],
```

```
             [[1687.342],
              [1221.588],
              [1200.44 ]],

             [[1221.588],
              [1200.44 ],
              [1068.317]],

             [[1200.44 ],
              [1068.317],
              [1152.165]],

             [[1068.317],
              [1152.165],
              [1265.674]],

             [[1152.165],
```

```
from sklearn.linear_model import LinearRegression
reg1 = LinearRegression()
reg1.fit(x_tr, Y_train)
y_pred = reg1.predict(x_t)
```

```
y_pred
```

```
array([ 425.24878102,  499.21195843,  614.7460528 ,  776.8956425 ,
        757.00982513, 1127.57060386, 1307.89266184, 1560.03761342,
       1555.61827349, 1077.90051716, 1217.71508902,  998.82312162,
       1131.7006029 , 1205.02246018, 1164.02035864])
```

```
from sklearn.metrics import mean_squared_error, r2_score
rmse_lr = np.sqrt(mean_squared_error(Y_test, y_pred))
r2_lr = r2_score(Y_test, y_pred)
print(rmse_lr, r2_lr)
```

```
216.41959383878694 0.5761964443390513
```