EXPERIMENT – 7

PREDICTING MODEL - LINEAR REGRESSION

Aim:

To perform salary prediction model using Linear Regression

Procedure:

- Upload the given dataset
- Import all the necessities
- Read through the dataset and make it as dataframe
- Through sklearn train the model
- Test the model

Program:

```
from google.colab import files
uploaded=files.upload()
import numpy as np
import pandas as pd
file=next(iter(uploaded))
df=pd.read_csv(file)
df
```

| | YearsExperience | Salary |
|---|---|---|
| 0 | 1.1 | 39343 |
| 1 | 1.3 | 46205 |
| 2 | 1.5 | 37731 |
| 3 | 2.0 | 43525 |
| 4 | 2.2 | 39891 |
| 5 | 2.9 | 56642 |
| 6 | 3.0 | 60150 |
| 7 | 3.2 | 54445 |
| 8 | 3.2 | 64445 |
| 9 | 3.7 | 57189 |
| 10 | 3.9 | 63218 |
| 11 | 4.0 | 55794 |
| 12 | 4.0 | 56957 |
| 13 | 4.1 | 57081 |
| 14 | 4.5 | 61111 |
| 15 | 4.9 | 67938 |

| | | |
|---|---|---|
| 16 | 5.1 | 66029 |
| 17 | 5.3 | 83088 |
| 18 | 5.9 | 81363 |
| 19 | 6.0 | 93940 |
| 20 | 6.8 | 91738 |
| 21 | 7.1 | 98273 |
| 22 | 7.9 | 101302 |
| 23 | 8.2 | 113812 |
| 24 | 8.7 | 109431 |
| 25 | 9.0 | 105582 |
| 26 | 9.5 | 116969 |
| 27 | 9.6 | 112635 |
| 28 | 10.3 | 122391 |
| 29 | 10.5 | 121872 |

```
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
```

```
df.dropna(inplace=True)
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 30 entries, 0 to 29
Data columns (total 2 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   YearsExperience  30 non-null     float64
 1   Salary           30 non-null     int64  
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
```

```
df.describe()
```

| | YearsExperience | Salary |
|---|---|---|
| count | 30.000000 | 30.000000 |
| mean | 5.313333 | 76003.000000 |
| std | 2.837888 | 27414.429785 |
| min | 1.100000 | 37731.000000 |
| 25% | 3.200000 | 56720.750000 |
| 50% | 4.700000 | 65237.000000 |
| 75% | 7.700000 | 100544.750000 |
| max | 10.500000 | 122391.000000 |

```
features=df.iloc[:,[0]].values
labels=df.iloc[:,[1]].values
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(features,labels,test_size=0.2,random_state=42)
from sklearn.linear_model import LinearRegression
model=LinearRegression()
model.fit(x_train,y_train)
```

```
LinearRegression()
```

```
model.score(x_train,y_train)
```

```
0.9645401573418146
```

```
model.score(x_test,y_test)
```

```
0.9024461774180497
```

```
model.coef_
```

```
array([[9423.81532303]])
```

```
model.intercept_
```

```
array([25321.58301178])
```

```
filename = list(uploaded.keys())[0]
import pickle
pickle.dump(model,open(filename,'wb'))
model=pickle.load(open(filename,'rb'))
yr_of_exp=float(input("Enter years of experience: "))
yr_of_exp_NP=np.array([[yr_of_exp]])
Salary=model.predict(yr_of_exp_NP)
```

```
Enter years of experience: 34
```

```
print("Estimated Salary for {} years of experience is {}: " .format(yr_of_exp,Salary))
```

```
Estimated Salary for 34.0 years of experience is [[345731.30399483]]:
```

Result:

Thus the python program for predicting model using Linear Regression is executed and verified