EXPERIMENT NO: 7

Salary Prediction using Linear Regression

Aim:

To build and train a linear regression model that predicts an employee's salary based on their years of experience.

Algorithm:

1. **Load Data: Import the salary dataset using pandas.**
2. **Clean Data: Check for missing values and remove null entries.**
3. **Split Data: Separate features (YearsExperience) and labels (Salary), then divide into training and testing sets.**
4. **Train Model: Fit a LinearRegression model using the training data.**
5. **Evaluate Model: Calculate model accuracy using training and testing scores.**
6. **Model Parameters: Display slope (coef_) and intercept (intercept_).**
7. **Save & Load Model: Save the trained model using pickle and reload it for use.**
8. **Prediction: Take user input for years of experience and predict the corresponding salary.**

Program:

[22]:
```
import numpy as np
import pandas as pd
df=pd.read_csv("C:/Users/vijay/Downloads/Salary_data.csv")
df
```

[22]:

	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

[23]:
```
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
```

[24]:
```
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
```

[25]:
```
df.describe()
```

[25]:

	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

[26]:
```
features=df.iloc[:,[0]].values
label=df.iloc[:,[1]].values
```

[27]:
```
from sklearn.model_selection import train_test_split
x_train,x_test,y_train,y_test=train_test_split(features,label,test_size=0.2,random_state=42)
```

[28]:
```
from sklearn.linear_model import LinearRegression
model=LinearRegression()
```

```
[29]: model.fit(x_train,y_train)
      model.score(x_train,y_train)
      model.score(x_test,y_test)

[29]: 0.9024461774180497

[30]: model.coef_

[30]: array([[9423.81532303]])

[31]: model.intercept_

[31]: array([25321.58301178])

[32]: import pickle
      pickle.dump(model,open('SalaryPred.model','wb'))
      model=pickle.load(open('SalaryPred.model','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:  44
```

Result:

The trained linear regression model accurately predicts salary based on years of experience. The model is saved for future use, and upon entering a number of years, it outputs the estimated salary value.