Expt No: 7
Date: 08-09-25

Linear Regression

Aim:

To build a linear regression model using Python to predict the salary of an employee based on their years of experience using the dataset salary_data.csv.

Algorithm:

1. Start
2. Import required libraries (numpy, pandas and sklearn) for data handling and modeling.
3. Load the dataset and read the CSV file.
4. Explore and clean the data
   - Use info() and describe() to understand the dataset.
   - Handle missing values using dropna()
5. Prepare the data
6. Split the dataset
7. Train the model
8. Evaluate the model
9. Save and reload the model
10. Make predictions
11. Display result and save.

Program:

```
import numpy as np
import pandas as pd
df = pd.read_csv("Salary_data.csv")
df
df.info()
df.dropna(inplace=True)
df.info()
df.describe()
features = df.iloc[:,[0]].values
label = df.iloc[:,[1]].values
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)
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)
model.score(x_train, y_train)
model.score(x_test, y_test)
model.coef_
model.intercept_
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)
print("Estimated salary for {} years of experience is {}
.format(yr_of_exp, Salary))
```

Output :

```
<class 'pandas.core.frame.DataFrame'>
Range Index: 30 entries, 0 to 29
Data columns (total 2 columns):
```

#	Column	Non-Null Count	Dtype
0	Years Experience	30 non-null	float64
1	Salary	30 non-null	int64

```
dtypes: float64(1), int64(1)
memory usage: 612.0 bytes
<class 'pandas.core.frame.DataFrame'>
Range Index: 30 entries, 0 to 29
```

Sala

Result:

The Linear Regression model was successfully implemented to predict an employee's salary based on years of experience.