Linear Regression

Exp : 07

Date: 23-09-2025

Aim:

To implement and evaluate a **Simple Linear Regression** model using the Scikit-learn library to predict an employee's **Salary** based on their **Years of Experience** and save the trained model for future use.

Algorithm:

1. **Load and Clean Data:** Load data, inspect info, describe stats, and remove nulls (dropna).

2. **Prepare Data:** Separate features (X: Years Experience) and label (Y: Salary).

3. **Split Data:** Split the data into training and testing sets (train_test_split).

4. **Train Model:** Initialize and fit the LinearRegression model.

5. **Evaluate Model:** Calculate the model score (R-squared) for training and testing sets.

6. **Get Parameters:** Retrieve the learned coefficient and intercept.

7. **Model Persistence:** Save the model using pickle.dump and load it back using pickle.load.

8. **Predict:** Use the loaded model to predict the salary for a new input value.

Code:

```python
import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
import pickle
df=pd.read_csv('Salary_data.csv')
df.info()
df.dropna(inplace=True)
df.info()
df.describe()
features=df.iloc[:, [0]].values
label=df.iloc[:, [1]].values
x_train, x_test, y_train, y_test=train_test_split(features, label, test_size=0.2, random_state=0)
model=LinearRegression()
model.fit(x_train, y_train)
model.score(x_train,y_train)
model.score(x_test,y_test)
model.coef_
model.intercept_
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

RangeIndex: 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'

RangeIndex: 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

Years Experience 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

0.9603182547438908

0.9184170849214232

array([[9281.30847068]])

array([27166.73682891])

Enter Years of Experience: 44

Estimated Salary for 44.0 years of experience is [[435544.30953887]]:

Result:

The Simple Linear Regression model was successfully implemented to predict Salary based on Years of Experience. [cite_start]The model achieved a high R-squared score, **0.960** on the training data [cite: 401] [cite_start]and **0.918** on the test data[cite: 403], indicating a strong fit. [cite_start]The learned regression equation is **Salary = 9281.31 * (Years of Experience) + 27166.74** [cite: 406, 408]. The trained model was successfully saved and loaded using the pickle module, demonstrating model persistence, and was used to make a new prediction.