Aim:

To develop a Simple Linear Regression model that predicts an employee's salary based on their years of experience using the Salary dataset.

Procedure:

1. Import libraries — Import necessary packages such as numpy, pandas, and sklearn.
2. Load dataset — Read the dataset (Salary_data.csv) into a DataFrame using pd.read_csv().
3. Inspect data — Use df.info() and df.describe() to view dataset details.
4. Clean data — Remove null values using df.dropna(inplace=True).
5. Define features and labels —
   Features → YearsExperience
   Labels → Salary
6. Split dataset — Use train_test_split() to divide data into training and test sets.
7. Train model — Fit a LinearRegression() model using training data.
8. Evaluate model — Print the training and test scores, coefficient, and intercept.
9. Save and load model — Use the pickle module to store and reload the model.
10. Predict salary — Take user input for years of experience and predict salary using the trained model.

In [8]:
```python
#Experiment No. 7
import numpy as np
import pandas as pd

# Load dataset
df = pd.read_csv("Salary_data.csv")
df

# Check info
df.info()

# Remove null values
df.dropna(inplace=True)
df.info()

# Describe dataset
df.describe()

# Features and Labels
features = df.iloc[:, [0]].values
label = df.iloc[:, [1]].values

# Split dataset
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(features, label, test_size=0

# Train model
from sklearn.linear_model import LinearRegression
model = LinearRegression()
model.fit(x_train, y_train)

# Model evaluation
print("Training score:", model.score(x_train, y_train))
print("Test score:", model.score(x_test, y_test))
print("Coefficient:", model.coef_)
print("Intercept:", model.intercept_)

# Save model
import pickle
pickle.dump(model, open('SalaryPred.model', 'wb'))

# Load model
model = pickle.load(open('SalaryPred.model', 'rb'))

# Prediction
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, Sal
```

```
<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
<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
Training score: 0.9411949620562126
Test score: 0.988169515729126
Coefficient: [[9312.57512673]]
Intercept: [26780.09915063]
Estimated Salary for 5.0 years of experience is [[73342.97478427]]:
```

In [ ]:

Result:

The linear regression model successfully predicts the salary based on years of experience.