

# School of Computer Science Engineering and Technology

Course-BTech  
Course Code - CSET211  
Year - Second  
Date - 27/08/2024

Type - AI Core-1  
Course Name - Statistical Machine Learning  
Semester - ODD  
Batch - CSE 3rd Semester

## Lab Assignment - 4 : Simple linear regression with scratch and with inbuilt function

### CO- Mapping

Section	CO1	CO2	CO3	CO4
Section 1: Q1-Q10	√			
Section 2: Q1-Q4	√			

**About Dataset:** Imagine you are a hiring manager in a growing tech company, and you are curious about the impact of experience on employee salaries. You have collected data over the years, recording the number of years each employee has worked and their current salary. The Salary\_Data dataset contains two columns: "YearsExperience" (the number of years an employee has worked) and "Salary" (the corresponding salary). As your company scales, you want to predict how much to offer new hires based on their experience. By analyzing it, you aim to answer questions like: How much does an additional year of experience typically add to someone's salary? Is there a strong, linear relationship, or do other factors play a more significant role?

### Section 1: Simple linear regression with scratch

1. Load the Salary\_data dataset.
2. Display the first 5 rows of the dataset.
3. Check for missing values in the dataset.
4. Replace missing values in the dataset (if any) with the mean of the respective columns.
5. Display basic statistics like mean, median, and standard deviation for both columns.
6. Divide data into 80% train and 20% test split.
7. Create a scatter plot showing the relationship between "YearsExperience" (X-axis) and "Salary" (Y-axis) for train split data.
8. Train a linear regression model without using inbuilt function.
9. Predict the income for the test dataset using the trained model.
10. Calculate the Mean Squared Error (MSE) and R-squared score for the predictions.

## **Section 2: Linear regression using inbuilt function**

1. Train a linear regression model using inbuilt function using train set.
2. Predict the income for the test dataset using the trained model.
3. Calculate the Mean Squared Error (MSE) and R-squared score for the predictions.
4. Compare both outcomes from without inbuilt function (section 1) and using inbuilt function.

**Platform Required:** Anaconda, Editor: Jupyter/Spyder/Pycharm/Google Colab

### **Submission Instructions:**

- Submission required .ipynb file only
- Submission is through LMS only.