

# REPORT

Princy Gautam(B20BB051)

---

## Problem 1

A.

- We imported different libraries necessary for calculations and plotting of graph.
- We assigned the name data to read a csv file.
- With the help of matplotlib we plotted a scatter plot of the data frame.

B.

1.

- split the data into train and test set in the ratio of 80:20
- Implement the linear regression model and fit the model on `x_train` and `y_train`.
- Predict `y` from `x_test`.
- Label `x` and `y` axes.
- Plot the graph which would be straight line.

2.

- Calculate the slope and intercept of the graph plotted in B.1. using `model.coef_` and `model.intercept` and then print them.

3.

- Estimate mean.
- Estimate variance.
- Estimate covariance.
- And coefficients by defining functions.
- Plot values.
- Plot line values of `x` and `y`.
- Plot line.
- Plot data points.
- Label `a` and `y` axes and implement the graph.

4.

- Implement both the graphs of (1) and (3) in a single scatter plot and observe the difference.

## **Problem 2**

A.

- Import necessary libraries.
- Read the csv file using pandas and assign it a name.
- Since gender is a nominal data, so we use nominal encoder to convert the categorical data into numerical data.
- This assigns as follows:
- Female: 0
- Male: 1
- Drop the Gender column and concatenate the nominal encoded columns in the original data frame and name it as data\_new to differentiate it from the original data frame.
- Split the dataset into training and test set in the ratio of 80:20

B.

- Import *LogisticRegression* from *sklearn.linear\_model*.
- Fit X\_train and y\_train.
- Predict the values of y for X\_test.

C.

- Now we will create confusion matrix to check the accuracy of classification.
- Import *metrics* and *confusion matrix* from sklearn library.
- Print confusion matrix giving it two parameters, mainly y\_test and y\_pred.
- Make a plot of the confusion matrix of y\_test and y\_pred.
- Print accuracy, precision and recall of the plot.