**“Experiment 3.2”**

1. **Aim:**

Study of Regression Analysis using R Programming.

1. **Theory:**

**Linear Regression:** It is a commonly used type of predictive analysis. It is a statistical approach for modeling the relationship between a dependent variable and a given set of independent variables.

**There are two types of linear regression:**

* Simple Linear Regression
* Multiple Linear Regression

If there is only one input variable (x), then such linear regression is called **simple linear regression**. And if there is more than one input variable, then such linear regression is called **multiple linear regression**.

1. **Solution**/**Code:**

library(caTools)

library(ggplot2)

dataset = read.csv('F:/sem6/DM Lab/Salary\_Data.csv')

split = sample.split(dataset$Salary, SplitRatio = 0.7)

trainingset = subset(dataset, split == TRUE)

testset = subset(dataset, split == FALSE)

lm.r= lm(formula = Salary ~ YearsExperience,data = trainingset)

coef(lm.r)

ypred = predict(lm.r, newdata = testset)

ggplot() +

geom\_point(aes(x = trainingset$YearsExperience, y = trainingset$Salary), colour='red') +

geom\_line(aes(x = trainingset$YearsExperience,

y = predict(lm.r, newdata = trainingset)), colour = 'blue') +

ggtitle('Salary vs Experience (Training set)') +

xlab('Years of experience') + ylab('Salary')

ggplot() +

geom\_point(aes(x = testset$YearsExperience, y = testset$Salary), colour = 'red') +

geom\_line(aes(x = trainingset$YearsExperience,

y = predict(lm.r, newdata = trainingset)), colour = 'blue') +

ggtitle('Salary vs Experience (Test set)') +

xlab('Years of experience') + ylab('Salary')

1. **Result/Output:**



