Prepare a classification model using SVM for salary data

**Ans :**

**R Code :**

## Support Vector Machines

########## Salary Data Set #########

# divide into training and test data

salary\_train <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Support Vector Machines\\SalaryData\_Train(1).csv')

salary\_test <- read.csv('D:\\Data Science\\Excelr\\Assignments\\Assignment\\Support Vector Machines\\SalaryData\_Test(1).csv')

##Training a model on the data ----

# begin by training a simple linear SVM

library(kernlab)

salary\_classifier <- ksvm(Salary ~ ., data = salary\_train,

kernel = "vanilladot")

## Evaluating model performance ----

# predictions on testing dataset

salary\_predictions <- predict(salary\_classifier, salary\_test)

head(salary\_predictions)

#table(salary\_predictions, salary\_test$letter)

agreement <- salary\_predictions == salary\_test$Salary

prop.table(table(agreement))

## Improving model performance ----

salary\_classifier\_rbf <- ksvm(Salary ~ ., data = salary\_train, kernel = "rbfdot")

salary\_predictions\_rbf <- predict(salary\_classifier\_rbf, salary\_test)

head(salary\_predictions\_rbf)

agreement\_rbf <- salary\_predictions\_rbf == salary\_test$Salary

table(agreement\_rbf)

prop.table(table(agreement\_rbf))

**Results :**

> salary\_classifier <- ksvm(Salary ~ ., data = salary\_train,

+ kernel = "vanilladot")

Setting default kernel parameters

> ## Evaluating model performance ----

> # predictions on testing dataset

> salary\_predictions <- predict(salary\_classifier, salary\_test)

> head(salary\_predictions)

[1] <=50K <=50K <=50K >50K <=50K >50K

Levels: <=50K >50K

> #table(letter\_predictions, letters\_test$letter)

> agreement <- salary\_predictions == salary\_test$Salary

> prop.table(table(agreement))

agreement

FALSE TRUE

0.1537185 0.8462815

> ## Improving model performance ----

> salary\_classifier\_rbf <- ksvm(Salary ~ ., data = salary\_train, kernel = "rbfdot")

> salary\_predictions\_rbf <- predict(salary\_classifier\_rbf, salary\_test)

> head(salary\_predictions\_rbf)

[1] <=50K <=50K <=50K >50K <=50K >50K

Levels: <=50K >50K

> agreement\_rbf <- salary\_predictions\_rbf == salary\_test$Salary

> table(agreement\_rbf)

agreement\_rbf

FALSE TRUE

2196 12864

> prop.table(table(agreement\_rbf))

agreement\_rbf

FALSE TRUE

0.1458167 0.8541833

**Inference :**

Getting good results in kernel rbfdot.