Worksheet 3.1

Student Name:Yatin **UID: 19bcs2545**
Branch: BE-CSE **Section/Group: CSE 11 /b**
Semester: 5th **Date of Performance: 11 November 2021**

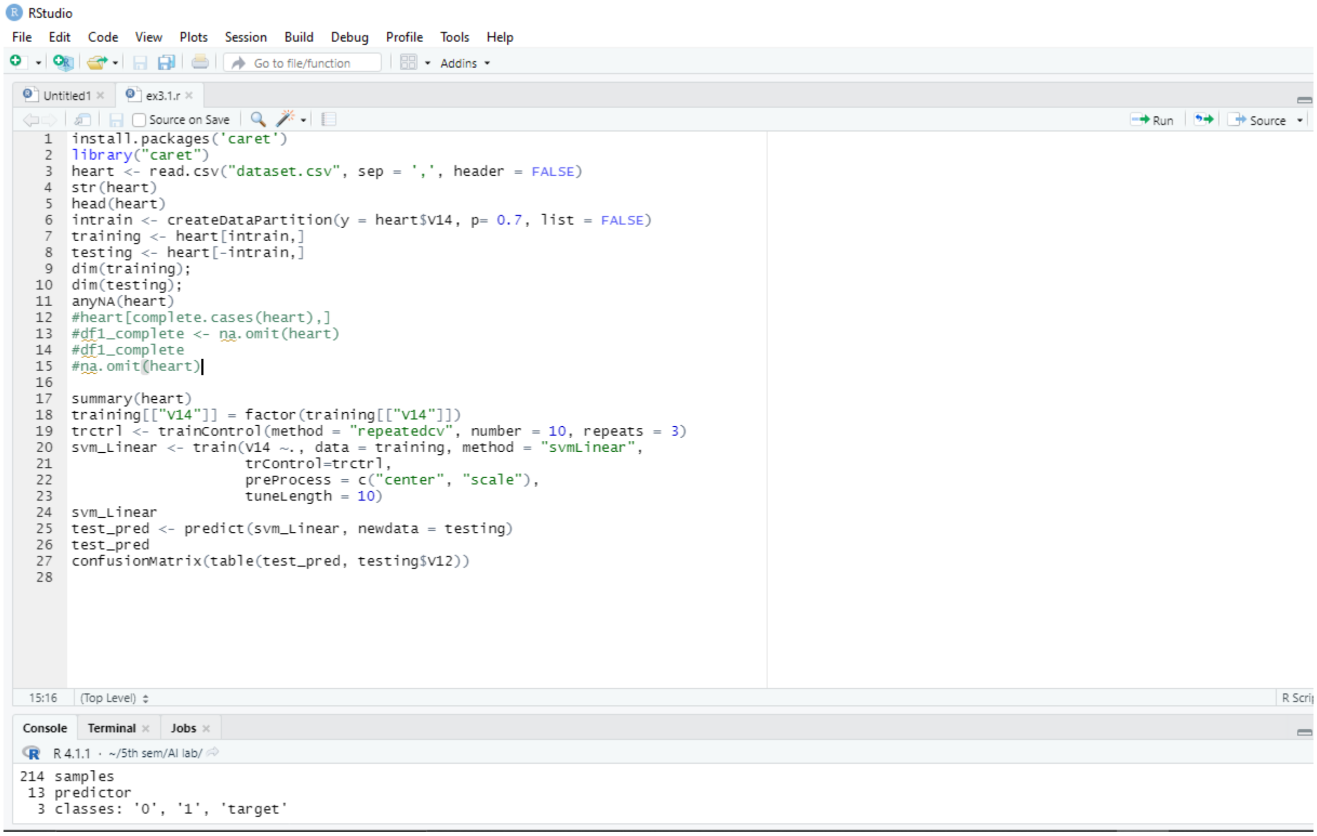
Subject Name: AIML lab **Subject Code: CSP-303**

1. **Aim/Overview of the practical:**
   IMPORT A HEART DISEASE DATASET. TRAIN AND TEST FOR PREDICTING WHETHER A PATIENT IS SUFFERING FROM ANY HEART DISEASE OR NOT.
   Note: if you don't find the said dataset you can proceed with any other dataset as well.

2. **The task to be done/ Which logistics used:**

   Understanding linear regression and how to explore data set, use and importance of SVM.

3. Steps for experiment/practical/Code:



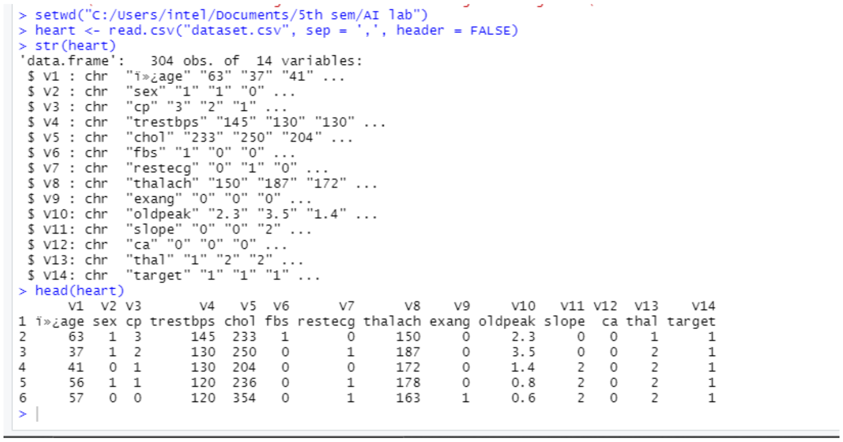
```r
install.packages('caret')
library("caret")
heart <- read.csv("dataset.csv", sep = ',', header = FALSE)
str(heart)
head(heart)
intrain <- createDataPartition(y = heart$V14, p= 0.7, list = FALSE)
training <- heart[intrain,]
testing <- heart[-intrain,]
dim(training);
dim(testing);
anyNA(heart)
#heart[complete.cases(heart),]
#df1_complete <- na.omit(heart)
#df1_complete
#na.omit(heart)

summary(heart)
training[["V14"]] = factor(training[["V14"]])
trctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
svm_Linear <- train(V14 ~., data = training, method = "svmLinear",
                    trControl=trctrl,
                    preProcess = c("center", "scale"),
                    tuneLength = 10)
svm_Linear
test_pred <- predict(svm_Linear, newdata = testing)
test_pred
confusionMatrix(table(test_pred, testing$V12))
```

```
214 samples
 13 predictor
  3 classes: '0', '1', 'target'
```

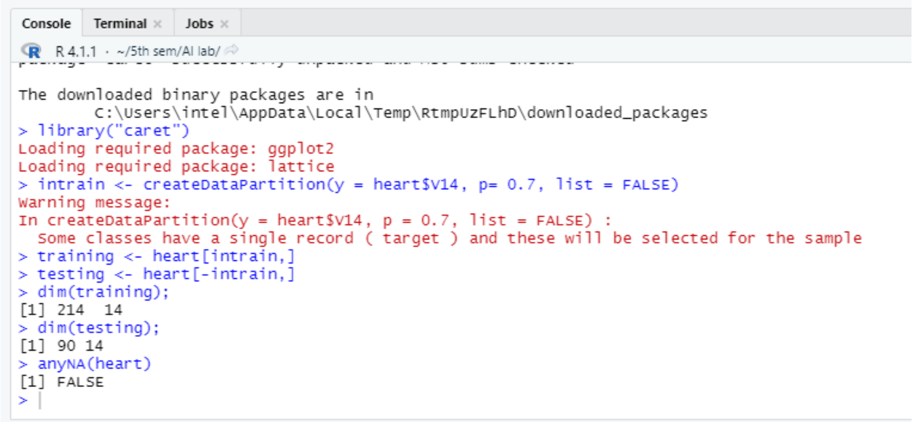
4. Observations/Discussions/ Complexity Analysis:

In this program we explored a data set: heart disease data set. We trained and test the data set to predict whether a patient is suffering from any heart disease or not.

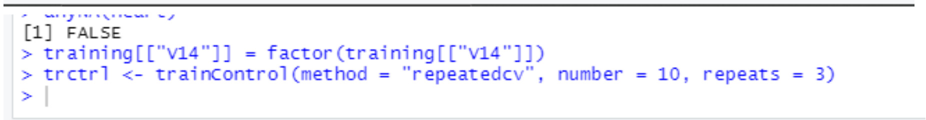
5. Result/Output/Writing Summary:



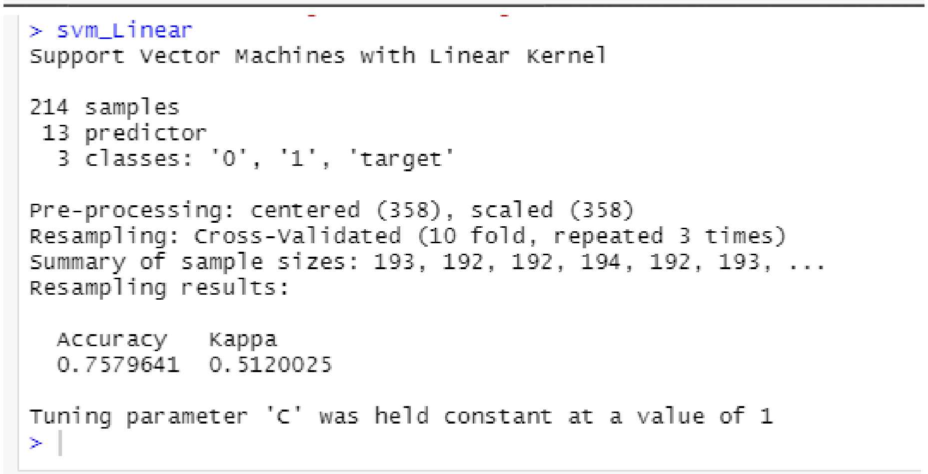
```
> setwd("C:/Users/intel/Documents/5th sem/AI lab")
> heart <- read.csv("dataset.csv", sep = ',', header = FALSE)
> str(heart)
'data.frame':	304 obs. of  14 variables:
 $ V1 : chr  "ï»¿age" "63" "37" "41" ...
 $ V2 : chr  "sex" "1" "1" "0" ...
 $ V3 : chr  "cp" "3" "2" "1" ...
 $ V4 : chr  "trestbps" "145" "130" "130" ...
 $ V5 : chr  "chol" "233" "250" "204" ...
 $ V6 : chr  "fbs" "1" "0" "0" ...
 $ V7 : chr  "restecg" "0" "1" "0" ...
 $ V8 : chr  "thalach" "150" "187" "172" ...
 $ V9 : chr  "exang" "0" "0" "0" ...
 $ V10: chr  "oldpeak" "2.3" "3.5" "1.4" ...
 $ V11: chr  "slope" "0" "0" "2" ...
 $ V12: chr  "ca" "0" "0" "0" ...
 $ V13: chr  "thal" "1" "2" "2" ...
 $ V14: chr  "target" "1" "1" "1" ...
> head(heart)
      V1  V2 V3       V4   V5  V6      V7      V8    V9     V10   V11 V12  V13    V14
1 ï»¿age sex cp trestbps chol fbs restecg thalach exang oldpeak slope  ca thal target
2     63   1  3      145  233   1       0     150     0     2.3     0   0    1      1
3     37   1  2      130  250   0       1     187     0     3.5     0   0    2      1
4     41   0  1      130  204   0       0     172     0     1.4     2   0    2      1
5     56   1  1      120  236   0       1     178     0     0.8     2   0    2      1
6     57   0  0      120  354   0       1     163     1     0.6     2   0    2      1
> 
```



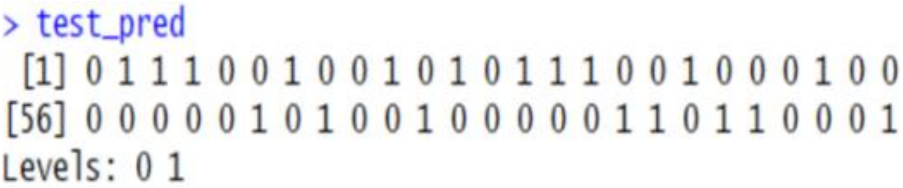
```
The downloaded binary packages are in
        C:\Users\intel\AppData\Local\Temp\RtmpUzFLhD\downloaded_packages
> library("caret")
Loading required package: ggplot2
Loading required package: lattice
> intrain <- createDataPartition(y = heart$V14, p= 0.7, list = FALSE)
Warning message:
In createDataPartition(y = heart$V14, p = 0.7, list = FALSE) :
  Some classes have a single record ( target ) and these will be selected for the sample
> training <- heart[intrain,]
> testing <- heart[-intrain,]
> dim(training);
[1] 214  14
> dim(testing);
[1] 90 14
> anyNA(heart)
[1] FALSE
> 
```



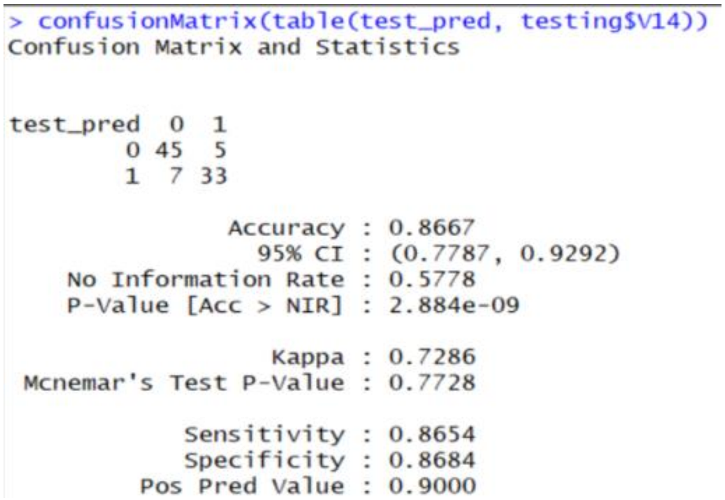
```
[1] FALSE
> training[["V14"]] = factor(training[["V14"]])
> trctrl <- trainControl(method = "repeatedcv", number = 10, repeats = 3)
> 
```



```
> svm_Linear
Support Vector Machines with Linear Kernel

214 samples
 13 predictor
  3 classes: '0', '1', 'target'

Pre-processing: centered (358), scaled (358)
Resampling: Cross-Validated (10 fold, repeated 3 times)
Summary of sample sizes: 193, 192, 192, 194, 192, 193, ...
Resampling results:

  Accuracy   Kappa
  0.7579641  0.5120025

Tuning parameter 'C' was held constant at a value of 1
> 
```



```
> test_pred
 [1] 0 1 1 1 0 0 1 0 0 1 0 1 0 1 1 1 0 0 1 0 0 0 1 0 0
[56] 0 0 0 0 0 1 0 1 0 0 1 0 0 0 0 0 1 1 0 1 1 0 0 0 1
Levels: 0 1
```



```
> confusionMatrix(table(test_pred, testing$V14))
Confusion Matrix and Statistics


test_pred  0  1
        0 45  5
        1  7 33

               Accuracy : 0.8667
                 95% CI : (0.7787, 0.9292)
    No Information Rate : 0.5778
    P-Value [Acc > NIR] : 2.884e-09

                  Kappa : 0.7286
 Mcnemar's Test P-Value : 0.7728

            Sensitivity : 0.8654
            Specificity : 0.8684
         Pos Pred Value : 0.9000
```

Learning outcomes (What I have learnt):

1. We learned about R language.
2. We learned linear regression and use of SVM.
3. We learned how to test and train any data set.

Evaluation Grid (to be created as per the SOP and Assessment guidelines by the faculty :)

Sr. No.	Parameters	Marks Obtained	Maximum Marks
1.			
2.			
3.			