**Tasks for Week-6: K-Nearest Neighbor Classifier**

Understand the following operations/functions on ‘Credit’ data and perform similar operations on ‘Parkinsons’ dataset based on given instructions.

rm(list=ls())

setwd("C:/Users/vitcc/Desktop/Plan-Fall-2020-21/DA-Lab")

data <- read.csv("Credit.csv",stringsAsFactors = T)

data$Cpur<-as.integer(data$Cpur)

data$Prop<-as.integer(data$Prop)

data[,-6] <- scale(data[,-6])

#install.packages("dplyr")

library(dplyr)

data\_TRAIN <-sample\_n(data,900)

data\_TEST <-setdiff(data,data\_TRAIN)

# install.packages("class") # for K-NN

library(class)

knnpredict <- knn(train=data\_TRAIN[,-6],test=data\_TEST[,-6],cl=data\_TRAIN$creditScore, k=5)

# install.packages("caret") # Classification and Regression Training

library(caret)

confusionMatrix(table(knnpredict,data\_TEST$creditScore),positive='good')

Sample Results:

Total number of observations =

Number of training observations =

Number of testing observations =

|  |  |  |  |
| --- | --- | --- | --- |
| K Value | Accuracy | Sensitivity | Specificity |
| 3 |  |  |  |
| 5 |  |  |  |
| 7 |  |  |  |
| . |  |  |  |
| . |  |  |  |