rm(list = ls())

setwd("C:\\Users\\ti1969\\Downloads")

library("readxl")

library(class)

ds<--read\_excel("Final Data\_Report 3.xlsx")

head(ds)

ds= ds\*-1

ds$`Financial Risk` <- ifelse(ds$`Financial Risk`== 2, 1, 0)

ds$`Financial Risk` = as.factor(ds$`Financial Risk`)

set.seed(123)

trainingRowIndex <- sample(1:nrow(ds), 0.5\*nrow(ds))

train.x=ds[trainingRowIndex,c(1:7,9)]

train.y=ds[trainingRowIndex,8]

test.x=ds[-trainingRowIndex,c(1:7,9)]

test.y=ds[-trainingRowIndex,8]

test.x = test.x[-1,]

test.y = test.y[-1]

knn.model<-knn(train.x,test.x,train.y,k=10, prob=T)

table(knn.model,test.y)

mean(knn.model==test.y)

library(pROC)

roc(test.y, attributes(knn.model)$prob)

plot(roc(test.y, attributes(knn.model)$prob),

print.thres = T,

print.auc=T)

acc<-rep(0,50)

for (knum in 1:50) {

knn.fit<-knn(train.x,test.x,train.y,k=knum)

acc[knum]<-mean(knn.fit==test.y)

}

print(paste("Highest accuracy of",as.character(acc[which.max(acc)]),"at k =",as.character(which.max(acc) )))

plot(1:50,acc,type="p",xlab="k",ylab="Test Acc", pretty=0)

knn.model<-knn(train.x,test.x,train.y,k=7, prob=T)

table(knn.model,test.y)

mean(knn.model==test.y)

library(pROC)

roc(test.y, attributes(knn.model)$prob)

plot(roc(test.y, attributes(knn.model)$prob),

print.thres = T,

print.auc=T)