**DECISION TREE & RANDOM FOREST**

#DECISION TREE

getwd()

setwd("C:\\Users\\10ani\\Desktop\\DS\_classes")

hr\_data=read.csv("HR\_DATA.csv")

colnames(hr\_data)

dim(hr\_data)

str(hr\_data)

corr=cor(hr\_data[,-c(4,5,6,8,20)])

corr=round(cor(hr\_data[,-c(4,5,6,8,20)]),2)

corr

write.csv(corr,"corr.csv")

set.seed(5467)

hr\_data$Attrition\_txt=NULL

ind=(sample(nrow(hr\_data),nrow(hr\_data)\*0.8))

train=hr\_data[ind,]

test=hr\_data[-ind,]

#install.packages('rpart')

library(rpart)

#install.packages('rpart.plot')

library(rpart.plot)

#install.packages('FSelector') ################################

library(FSelector)

#install.packages('pROC')

library(pROC)

fit1=rpart(Attrition~. ,data=train,method = "class")

rpart.plot(fit1)

rpart.rules(fit1)

fit2=rpart(Attrition~. ,data = train,parms = list(split="information"),control = rpart.control(minsplit = 20,minbucket = 10,maxdepth = 3),method="class")

rpart.plot(fit2)

sum(train$Attrition)/nrow(train) ##############PURPOSE#######################

library(caret)

varImp(fit2) #variable importance

information.gain(attrition~. ,data=train) ####################################

tree\_pred=predict(fit2,test,type="class")

table=table(actual=test$Attrition,prediction=tree\_pred)

247/294

**#roc**

train\_pred=predict(fit2,train,type="prob")

train\_pred

train\_pred=predict(fit2,train,type="prob")[,2]

train\_pred

trainROC=roc(response=train$Attrition,predictor = train\_pred,plot = TRUE,auc = TRUE)

plot(trainROC)

trainROC$auc

**RANDOM FOREST**

#install.packages('randomForest')

library(randomForest)

train1<-train

train1$Att<-ifelse(train1$Attrition==1,"Yes","No")

train1$Attrition<-NULL

test$Att<-ifelse(test$Attrition==1,"Yes","No")

test$Attrition<-NULL

train1$Att<-as.factor(train1$Att)

fit3= randomForest(Att~. ,data = train1, ntree=1000)

rfpred<-predict(fit3,test)

table(Actual=test$Att,predicted=rfpred)