**Assignment 10.2**

1. Use the below given data set

Data Set

2. Perform the below given activities:

a. Create classification model using different random forest models

b. Verify model goodness of fit

c. Apply all the model validation techniques

d. Make conclusions

e. Plot importance of variables

|  |
| --- |
| #Answers |
|  | #a |
|  | #reading the dataset |
|  | #using bankloan dataset |
|  | bankloan1 <- read.csv("D:\\BIG DATA\\DATA ANALYTICS WITH R, EXCEL & TABLEAU\\20 TIME SERIES FORECASTING\\bank-additional-full.csv",sep = ";") |
|  | View(bankloan1) |
|  |  |
|  | str(bankloan1) |
|  |  |
|  | # tree |
|  | library(caTools) |
|  | library(tree) |
|  | set.seed(1) |
|  | sam<- sample(x=1: nrow(bankloan1), size = 0.80\*nrow(bankloan1)) |
|  | train1<- bankloan1[sam, ] |
|  | table(bankloan1$default) |
|  | table(train1$default) |
|  |  |
|  | test1<- bankloan1[-sam, ] |
|  | table(test1$default) |
|  |  |
|  | model\_tree1<- tree(default~., data = train1) |
|  | summary(model\_tree1) |
|  |  |
|  | plot(model\_tree1); text(model\_tree1, pretty = 0, cex = 0.75) |
|  |  |
|  | pred\_tree1<- predict(model\_tree1, newdata = test1, |
|  | type = 'class') |
|  |  |
|  | conf\_tree1<- table(test1$default, pred\_tree1) |
|  | conf\_tree1 |
|  |  |
|  | OAA\_tree1<- (conf\_tree1[1,1]+conf\_tree1[2,2])/sum(conf\_tree1) |
|  | OAA\_tree1 |
|  |  |
|  | #~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~~RF |
|  |  |
|  | #random forest |
|  | library(randomForest) |
|  | model\_rf1<- randomForest(default~., data = train1) |
|  | model\_rf1 |
|  |  |
|  | summary(model\_rf1) |
|  |  |
|  | #b),c),d),e) |
|  | #Answers |
|  | #validation |
|  | #interpretation, Accuracy and model goodness of our model |
|  | #verify model goodness of fit |
|  | library(rpart) |
|  | library(caret) |
|  | # define training control |
|  | train\_control<- trainControl(method="cv", number=10) |
|  |  |
|  | # train the model |
|  | #method rf |
|  | model<- train(default~., data=train1, trControl=train\_control, method="rf") |
|  | model |
|  |  |
|  | # make predictions |
|  | predictions<- predict(model,test1) |
|  |  |
|  | # append predictions |
|  | pred<- cbind(test1,predictions) |
|  |  |
|  | # summarize results |
|  | confusionMatrix<- confusionMatrix(pred$predictions,pred$default) |
|  | confusionMatrix |
|  |  |
|  | # define training control |
|  | train\_control<- trainControl(method="cv", number=10) |
|  |  |
|  | # train the model |
|  | # method boosted tree |
|  | model<- train(default~., data=train1, trControl=train\_control, method="bstTree") |
|  | model |
|  |  |
|  | # make predictions |
|  | predictions<- predict(model,test1) |
|  |  |
|  | # append predictions |
|  | pred<- cbind(test1,predictions) |
|  |  |
|  | # summarize results |
|  | confusionMatrix<- confusionMatrix(pred$predictions,pred$churn) |
|  | confusionMatrix |
|  |  |
|  | #how do we create a cross validation scheme |
|  | control <- trainControl(method = 'repeatedcv', |
|  | number = 10, |
|  | repeats = 3) |
|  | seed <-7 |
|  | metric <- 'Accuracy' |
|  | set.seed(seed) |
|  | mtry <- sqrt(ncol(train1)) |
|  | tunegrid <- expand.grid(.mtry=mtry) |
|  | rf\_default <- train(default~., |
|  | data = train1, |
|  | method = 'rf', |
|  | metric = metric, |
|  | tuneGrid = tunegrid, |
|  | trControl = control) |
|  | print(rf\_default) |
|  |  |
|  | #prediction of model\_rf1 |
|  | pred\_rf1<- predict(model\_rf1, test1, type = 'class') |
|  | head(pred\_rf1, 15) |
|  |  |
|  | #interpretation, Accuracy and model goodness of our model |
|  | #verify model goodness of fit |
|  | #summary |
|  | summary(model\_rf1) |
|  |  |
|  | #confusion matrix of model\_rf1 |
|  | conf\_rf1<- table(test1$default, pred\_rf1) |
|  | conf\_rf1 |
|  |  |
|  | #accuracy of model\_rf1 |
|  | OAA\_rf1<- (conf\_rf1[1,1]+conf\_rf1[2,2])/sum(conf\_rf1) |
|  | OAA\_rf1 |
|  |  |
|  | #plotting imp of variance |
|  | library(caret) |
|  |  |
|  | importance(model\_rf1) |
|  | varImp(model\_rf1) |
|  | varImpPlot(model\_rf1, col = 'red') |