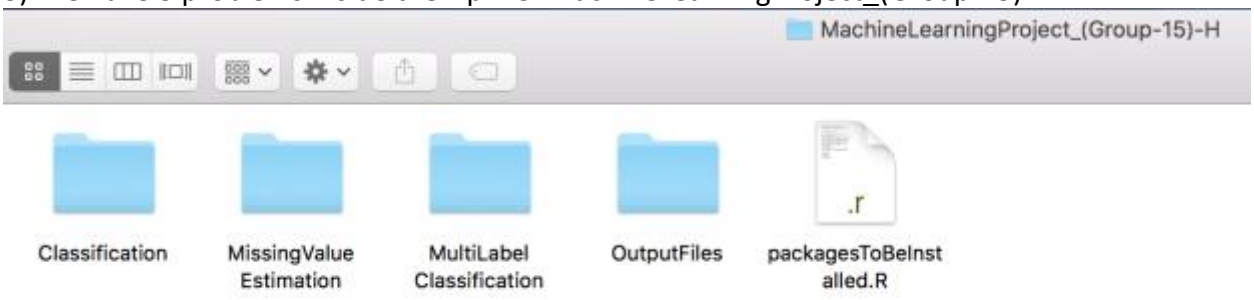
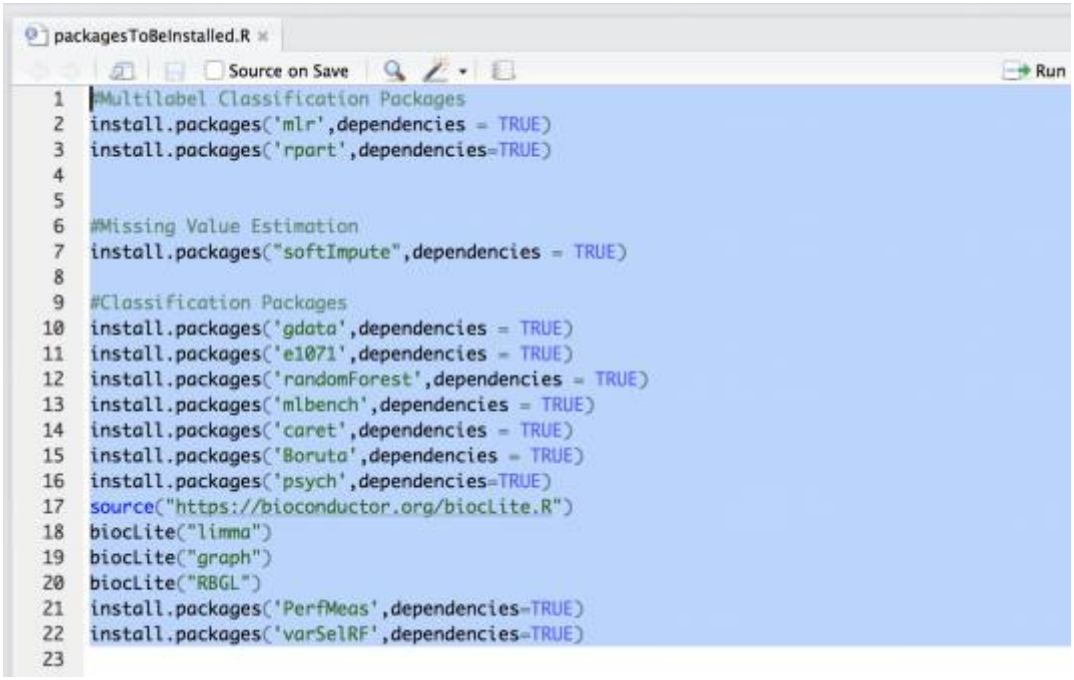
Instructions for Package Installation and Running the code

1) Install R Studio for executing our project.
2) If you have downloaded from the dropbox link, you will see the following folder structure.

3) We have 3 problems inside the zip file 'MachineLearningProject_(Group-15)'



4) Open a new R Studio window, Install the dependencies that are needed for our project which is mentioned in the separate file called – '**packagesToBeInstalled.R**', select all the lines and RUN the file first.

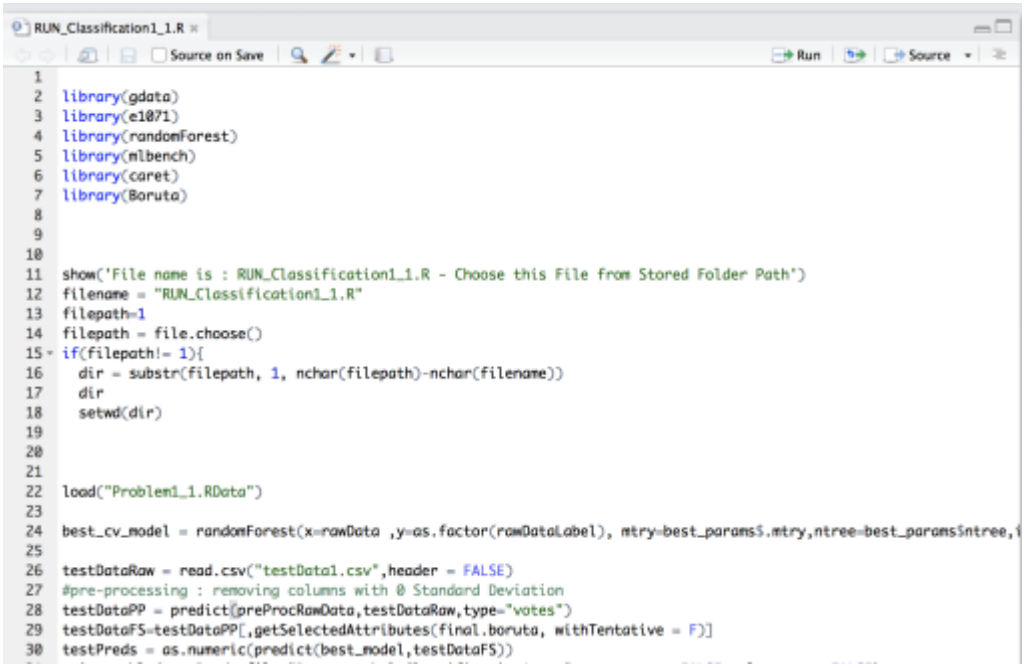


It takes time to install all packages. This is the first Step.

Classification Problem

Steps to Run the **Classification** Problems

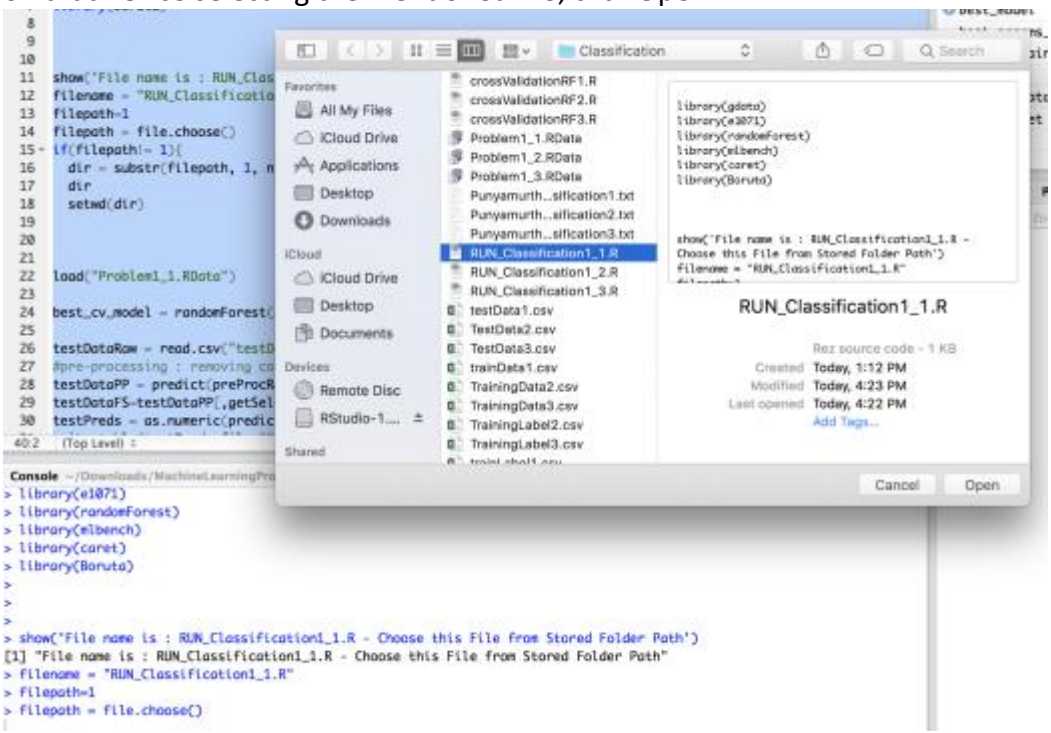
Open **RUN_Classification1_1.R** file



```
RUN_Classification1_1.R
Source on Save | Run | Source

library(gdata)
library(e1071)
library(randomForest)
library(mlbench)
library(caret)
library(Boruta)



show('File name is : RUN_Classification1_1.R - Choose this File from Stored Folder Path')
filename = "RUN_Classification1_1.R"
filepath=1
filepath = file.choose()
if(filepath!= 1){
  dir = substr(filepath, 1, nchar(filepath)-nchar(filename))
  dir
  setwd(dir)



load("Problem1_1.RData")

best_cv_model = randomForest(x=rawData ,y=as.factor(rawDataLabel), mtry=best_params$.mtry,ntree=best_params$ntree,

testDataRaw = read.csv("testData1.csv",header = FALSE)
#pre-processing : removing columns with 0 Standard Deviation
testDataPP = predict(preProcRawData,testDataRaw,type="votes")
testDataFS=testDataPP[,getSelectedAttributes(final.boruta, withTentative = F)]
testPreds = as.numeric(predict(best_model,testDataFS))
```

while executing, there appears a dialog box to choose the file, Kindly choose the same file '**RUN_Classification1_1.R'** from your workspace , This is to **set Directory for the working**.

This is very important because all the data related files and code workspace are indexed based on that. Once selecting the mentioned file, click Open .



```
Classification
Favorites
All My Files
iCloud Drive
Applications
Desktop
Downloads
iCloud
iCloud Drive
Desktop
Documents
Devices
Remote Disc
RStudio-1...
Shared

crossValidationRF1.R
crossValidationRF2.R
crossValidationRF3.R
Problem1_1.RData
Problem1_2.RData
Problem1_3.RData
Punyamurth...sification1.txt
Punyamurth...sification2.txt
Punyamurth...sification3.txt
RUN_Classification1_1.R
RUN_Classification1_2.R
RUN_Classification1_3.R
testData1.csv
TestData2.csv
TestData3.csv
trainData1.csv
TrainingData2.csv
TrainingData3.csv
TrainingLabel2.csv
TrainingLabel3.csv

RUN_Classification1_1.R
Rez source code - 1 KB
Created Today, 1:12 PM
Modified Today, 4:23 PM
Last opened Today, 4:22 PM
Add Tags...

Cancel
Open

Console
> library(e1071)
> library(randomForest)
> library(mlbench)
> library(caret)
> library(Boruta)
>
>
>
> show('File name is : RUN_Classification1_1.R - Choose this File from Stored Folder Path')
[1] "File name is : RUN_Classification1_1.R - Choose this File from Stored Folder Path"
> filename = "RUN_Classification1_1.R"
> filepath=1
> filepath = file.choose()
```

Program will run and Output will be generated in the same place where the program is opened in the following text file – '**PunyamurthulaClassification1.txt**'

Follow the same process for the Classification problem 2,3 in the following file names,
Program name: RUN_Classification1_2.R -
output File Name: - PunyamurthulaClassification2.txt

Program name: RUN_Classification1_3.R -
output File Name: - PunyamurthulaClassification3.txt.

All the above program files are modeled in a way that it gives the answer for our saved workspace. If you want to check the full running of the same code with Cross validation. Follow the below steps,
Select the following files,
Program Name: crossValidationRF1.R
crossValidationRF2.R
crossValidationRF3.R
These Program takes around 30 minutes to run and produce the output.

Missing Value Estimation:

Go to folder which contains the file, Run the following files,
Program Name:
RUN_Missing_Value_Estimation_1.R
RUN_Missing_Value_Estimation_2.R
RUN_Missing_Value_Estimation_3.R

Output will be in
PunyamurthulaMissingResult1.txt
PunyamurthulaMissingResult2.txt
PunyamurthulaMissingResult3.txt

Multi Label Classification:
Go to the folder which contains the file, Run the following file,

Program Name:
MultiLabelClassification.R

Output File:
SrinivasanMultClassification.txt

NOTE:
Kindly after running the program select the appropriate file while running, this sets the working directory for each problem. We have tested this in 4 different system. During each running of a program, File Needs to be chosen.