**HOME TASK**

**EXPERIMENT – 9**

**GROUP NUMBER: 3**

|  |  |  |
| --- | --- | --- |
| **S. No** | **Student Name** | **UID** |
| **1** | Sahul Kumar Parida | **20BCS4919** |

**Aim:**

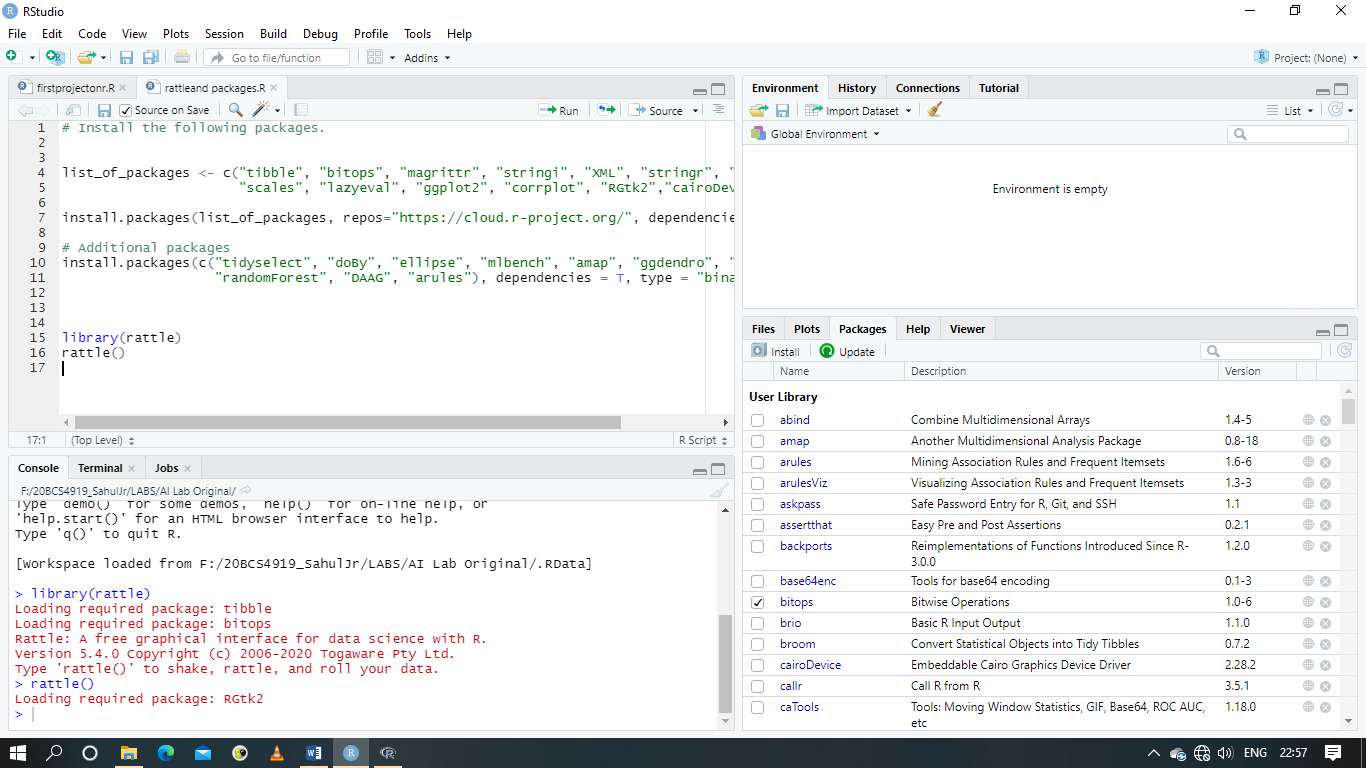
**Predict whether or not it will rain tomorrow by training a binary classification model.**

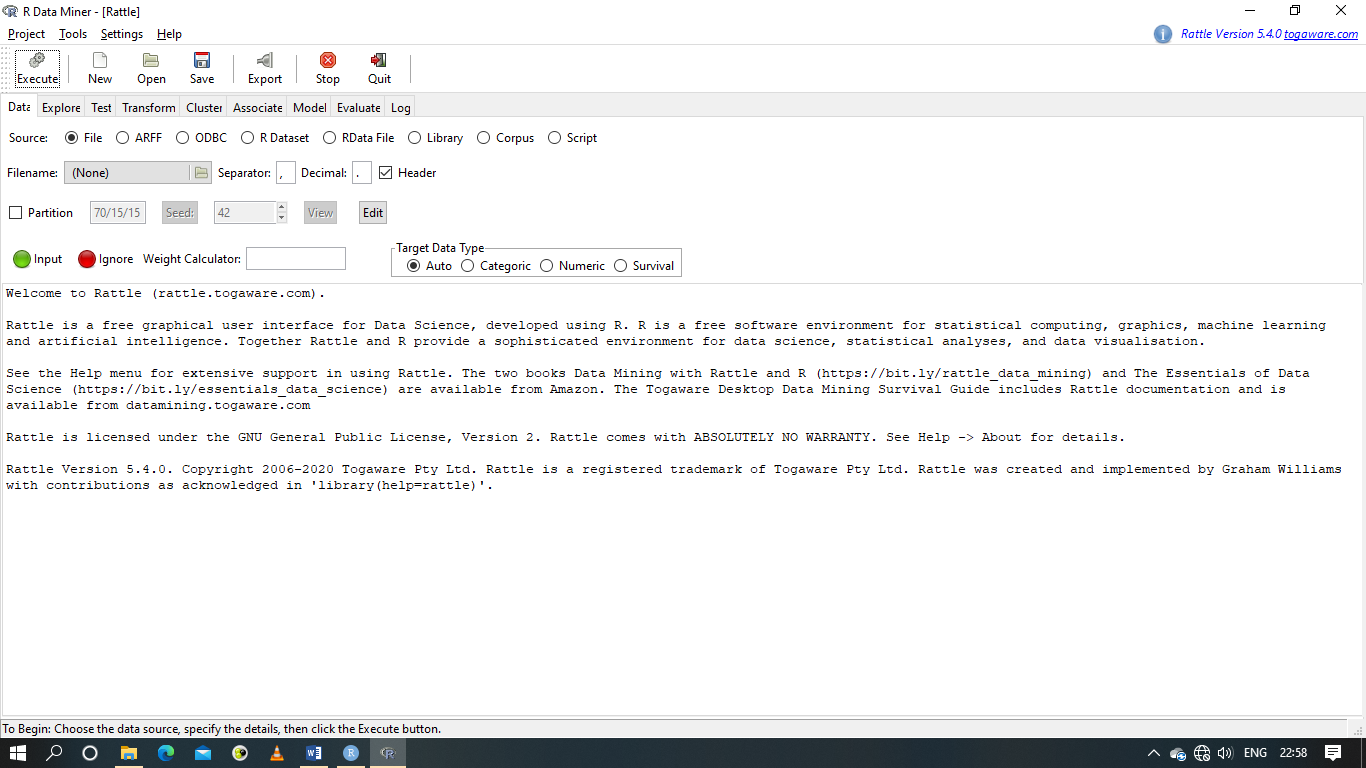
**Requirement:**

PC with internet connectivity, Rattle, R for Windows (Version 4.0.0), RStudio for Windows (Version 1.2.5042) and R Data Miner – [Rattle].

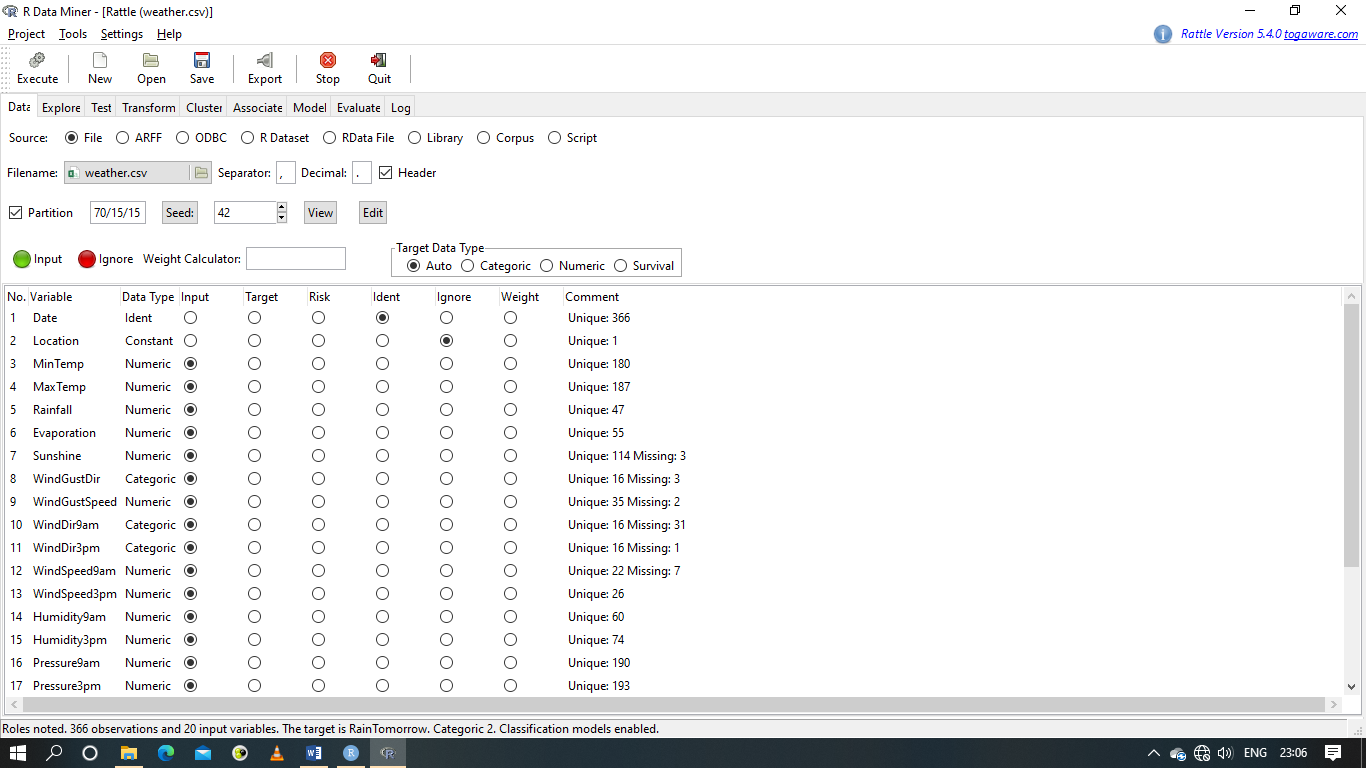
**Command**:

1. First we have to open a file named rattleand packages using RStudio Application.
2. Then we have to install the packages "tibble", "bitops", "magrittr”, “stringi”, “XML", "stringr”, “Hmisc", "R6”, “scales”, “lazyeval”, “ggplot2”, “corrplot", "RGtk2", "cairoDevice", "rattle", tidyselect", "doBy", "ellipse", "mlbench", "amap", "ggdendro", "fpc", "randomForest", "DAAG", "arules" in order to get the R Data Miner – [Rattle] so as to get the desired output.

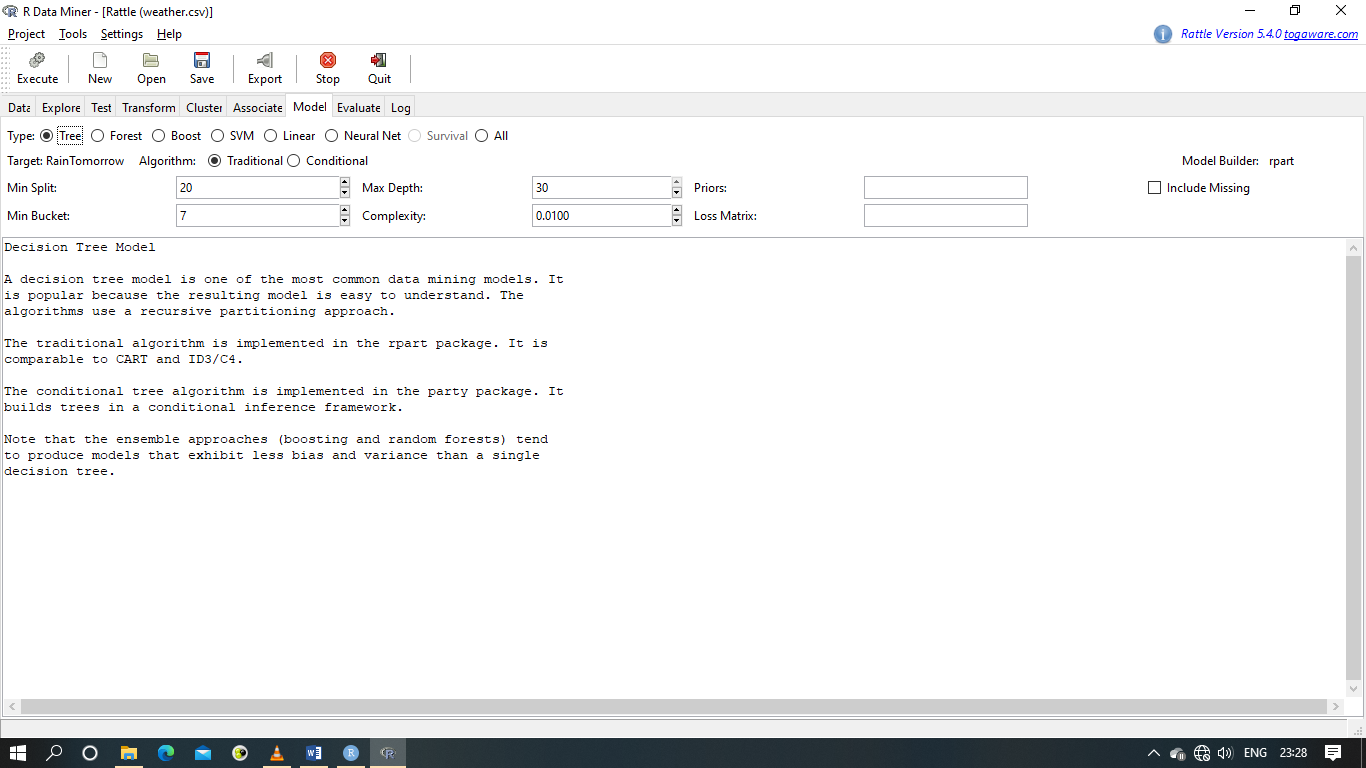




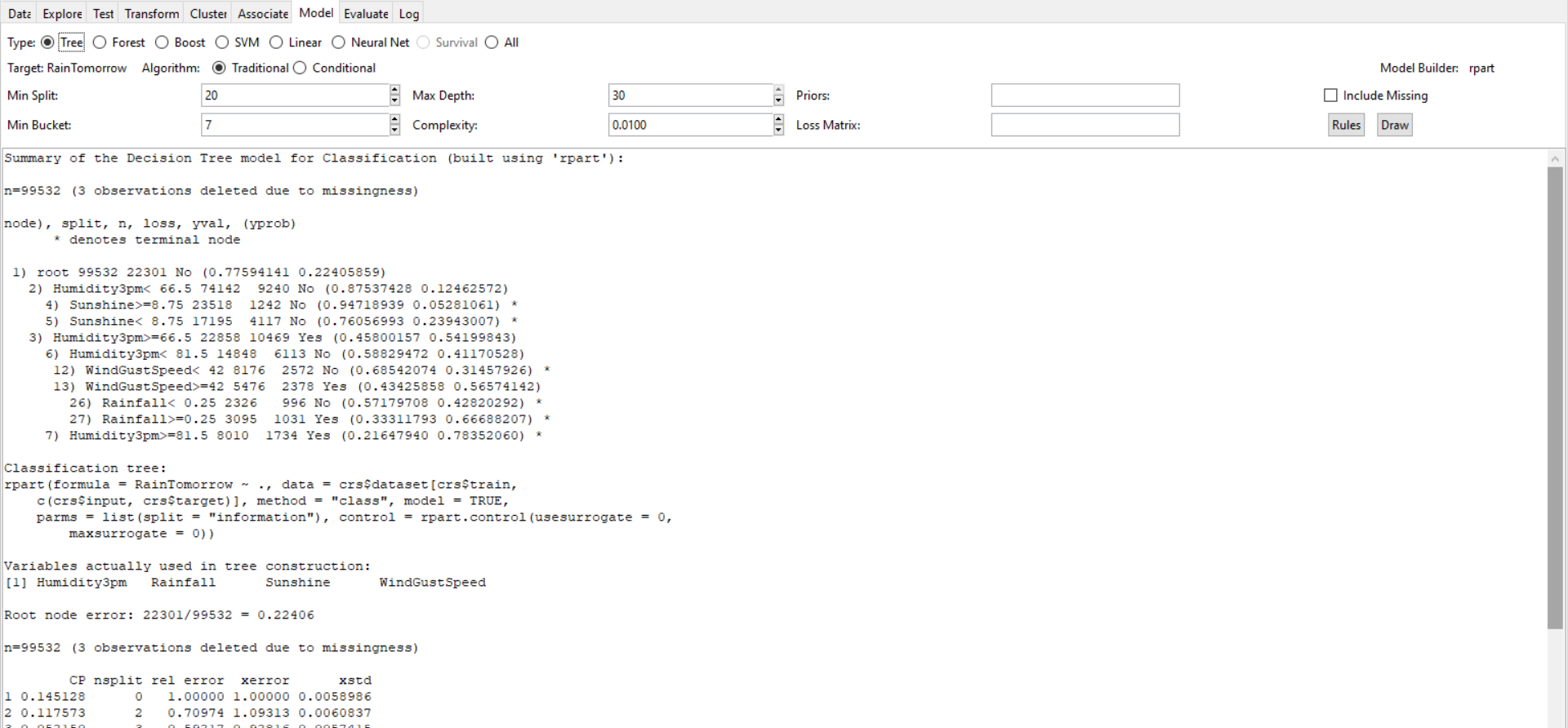
1. We have to open the file named weatherAUS.csv from R Data Miner – [Rattle].



1. In the Data tab, we have to Ignore the four values namely Date, Location, Rain Today and Risk\_MM and also we have to set the Target to Rain Tomorrow. Next we have to execute the file.
2. Next we have to go the Model tab, set the Type to Tree and execute the file.



**Expected output:**

******

Summary of the Decision Tree model for Classification (built using 'rpart'):

n=99532 (3 observations deleted due to missingness)

node), split, n, loss, yval, (yprob)

\* denotes terminal node

1) root 99532 22301 No (0.77594141 0.22405859)

2) Humidity3pm< 66.5 74142 9240 No (0.87537428 0.12462572)

4) Sunshine>=8.75 23518 1242 No (0.94718939 0.05281061) \*

5) Sunshine< 8.75 17195 4117 No (0.76056993 0.23943007) \*

3) Humidity3pm>=66.5 22858 10469 Yes (0.45800157 0.54199843)

6) Humidity3pm< 81.5 14848 6113 No (0.58829472 0.41170528)

12) WindGustSpeed< 42 8176 2572 No (0.68542074 0.31457926) \*

13) WindGustSpeed>=42 5476 2378 Yes (0.43425858 0.56574142)

26) Rainfall< 0.25 2326 996 No (0.57179708 0.42820292) \*

27) Rainfall>=0.25 3095 1031 Yes (0.33311793 0.66688207) \*

7) Humidity3pm>=81.5 8010 1734 Yes (0.21647940 0.78352060) \*

Classification tree:

rpart(formula = RainTomorrow ~ ., data = crs$dataset[crs$train,

c(crs$input, crs$target)], method = "class", model = TRUE,

parms = list(split = "information"), control = rpart.control(usesurrogate = 0,

maxsurrogate = 0))

Variables actually used in tree construction:

[1] Humidity3pm Rainfall Sunshine WindGustSpeed

Root node error: 22301/99532 = 0.22406

n=99532 (3 observations deleted due to missingness)

CP nsplit rel error xerror xstd

1 0.145128 0 1.00000 1.00000 0.0058986

2 0.117573 2 0.70974 1.09313 0.0060837

3 0.052150 3 0.59217 0.92816 0.0057415

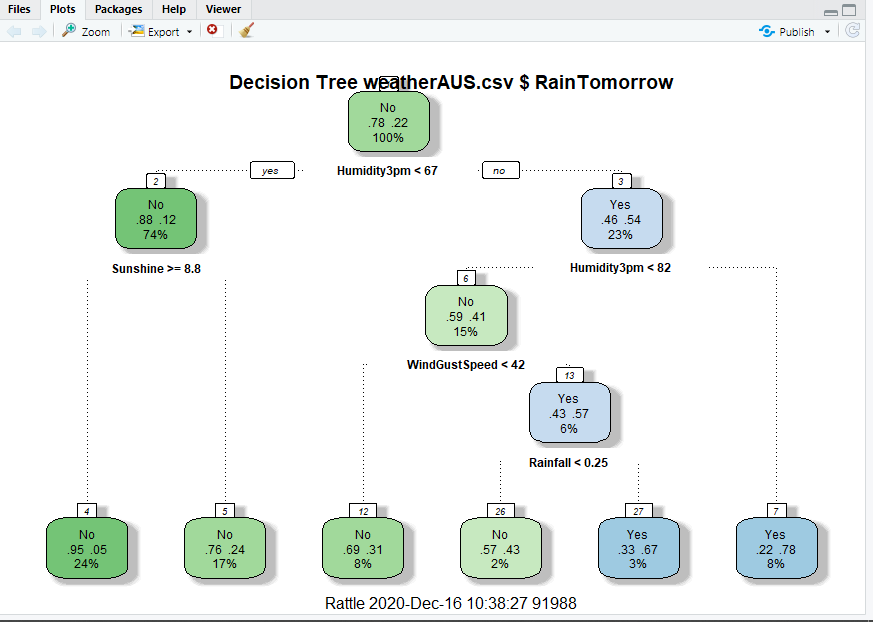
4 0.015739 4 0.54002 0.85669 0.0055715

5 0.010000 5 0.52428 0.82660 0.0054955

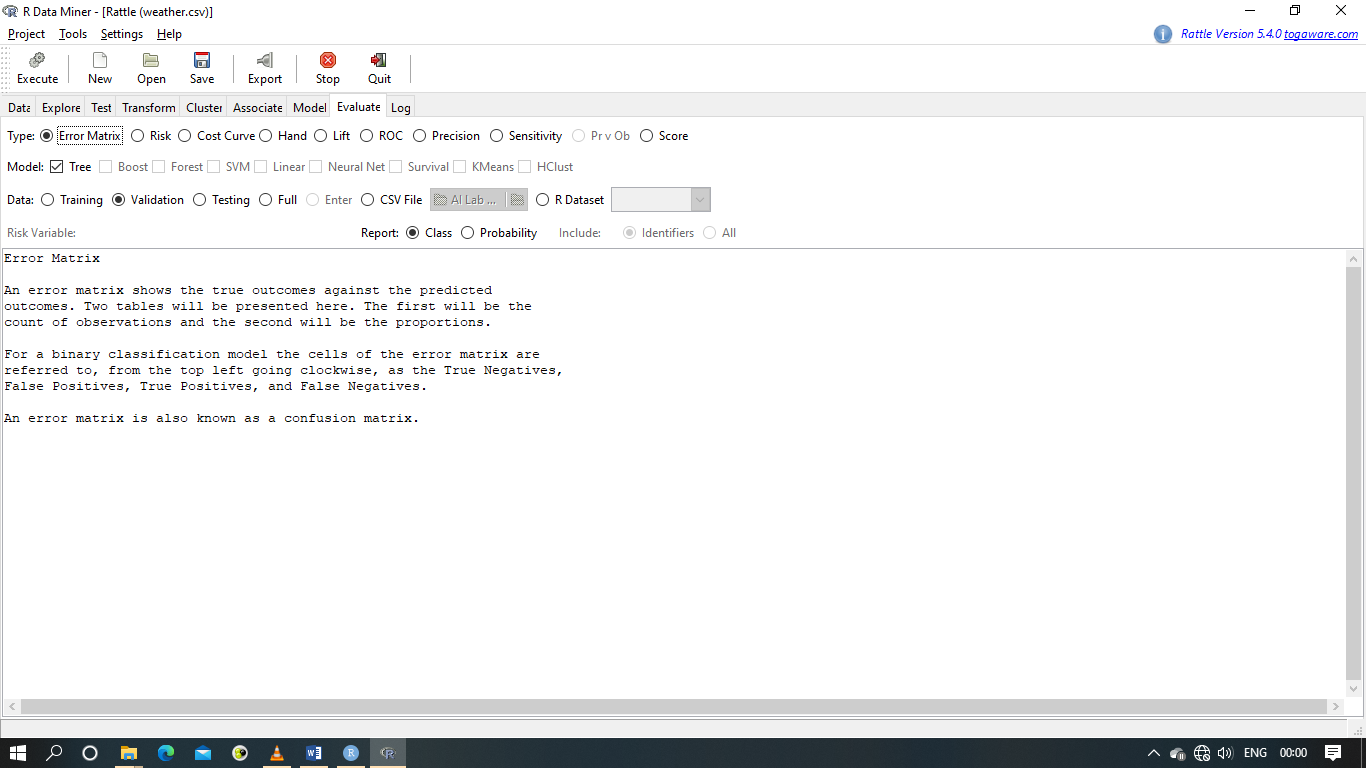
Time taken: 2.69 secs

Rattle timestamp: 2020-12-16 10:33:56 91988

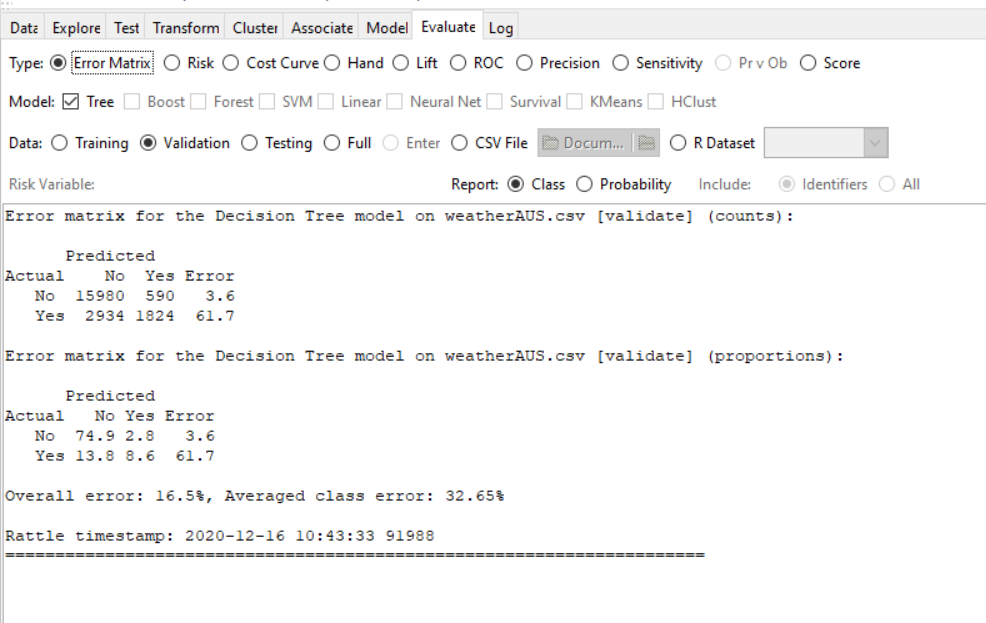
There are two things on the right side of the Model tab namely Rules and Draw. We have to click on Draw on the right side under the Model Tab in order to get the flowchart in the Plotting Region of RStudio Application.



Next we have to go to the Evaluate tab. Set the Type to Error Matrix, Model to Tree and Data to Validation.



Lastly we have to click on execute in order to get the final output.



Error matrix for the Decision Tree model on weatherAUS.csv [validate] (counts):

Predicted

Actual No Yes Error

No 15980 590 3.6

Yes 2934 1824 61.7

Error matrix for the Decision Tree model on weatherAUS.csv [validate] (proportions):

Predicted

Actual No Yes Error

No 74.9 2.8 3.6

Yes 13.8 8.6 61.7

Overall error: 16.5%, Averaged class error: 32.65%

Rattle timestamp: 2020-12-16 10:43:33 91988

======================================================================