

## Worksheet

### EXPERIMENT – 9

Name- YANA SRIVASTAVA

UID-20BCS2279

SECTION – 23 B

#### **Aim:**

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

#### **Requirements:**

- (i) Rattle
- (ii) R Data Miner.

#### **Expected Outcome:**

Rattle specifically uses a simple tab-based concept for the user interface, capturing a work flow through the data mining process with a tab for each stage.

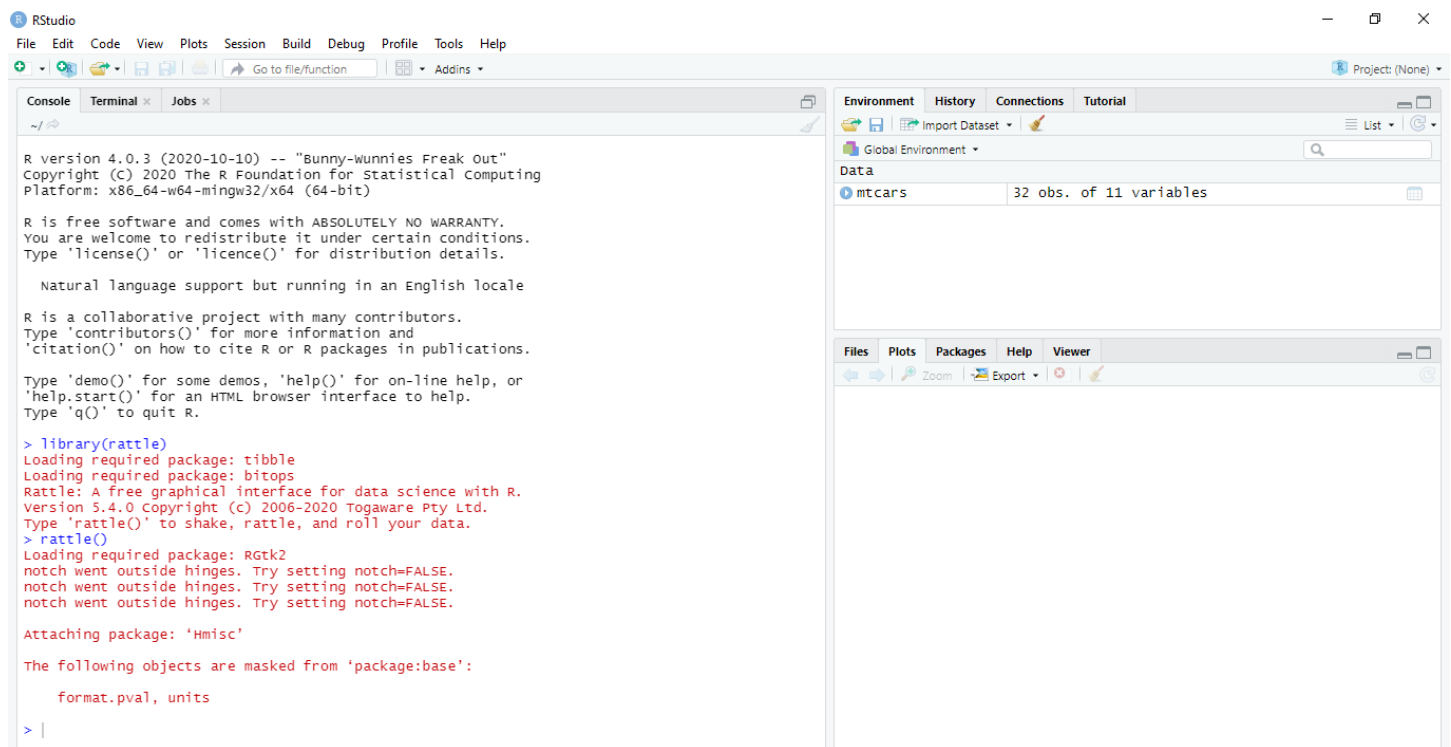
- A typical work flow progresses from the left most tab (the Data tab) to the right most tab (the Log tab).
- For any tab the idea is for the user to configure the available options and then to click the Execute button (or F2) to perform the appropriate task.
  - The status bar at the base of the window will indicate when the action is complete.
- Once we have processed our data, we are ready to build a model and with Rattle we can build the model with just a few mouse clicks. Using a

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

sample dataset that someone else has already prepared for us, in Rattle we simply:

## Steps

Open the RStudio window and execute the command library (rattle), rattle() to get the R Data Miner window.



```
RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Go to file/function Addins
Project: (None)
Environment History Connections Tutorial
Global Environment
Data
mtcars 32 obs. of 11 variables
Files Plots Packages Help Viewer
Zoom Export

~/
R version 4.0.3 (2020-10-10) -- "Bunny-wunnies Freak out"
Copyright (C) 2020 The R Foundation for Statistical Computing
Platform: x86_64-w64-mingw32/x64 (64-bit)

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> library(rattle)
Loading required package: tibble
Loading required package: bitops
Rattle: A free graphical interface for data science with R.
Version 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd.
Type 'rattle()' to shake, rattle, and roll your data.
> rattle()
Loading required package: RGtk2
notch went outside hinges. Try setting notch=FALSE.
notch went outside hinges. Try setting notch=FALSE.
notch went outside hinges. Try setting notch=FALSE.

Attaching package: 'Hmisc'

The following objects are masked from 'package:base':

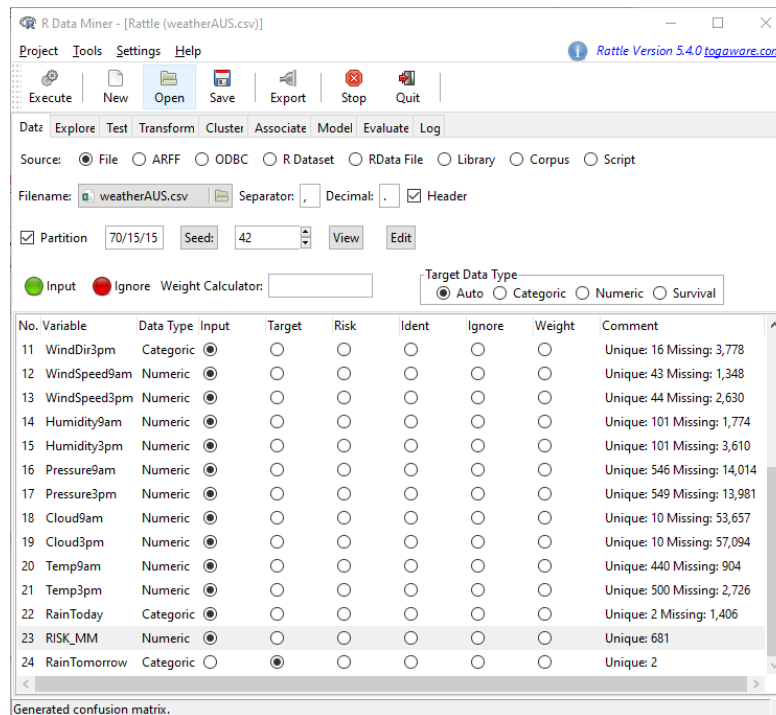
  format.pval, units

> |
```

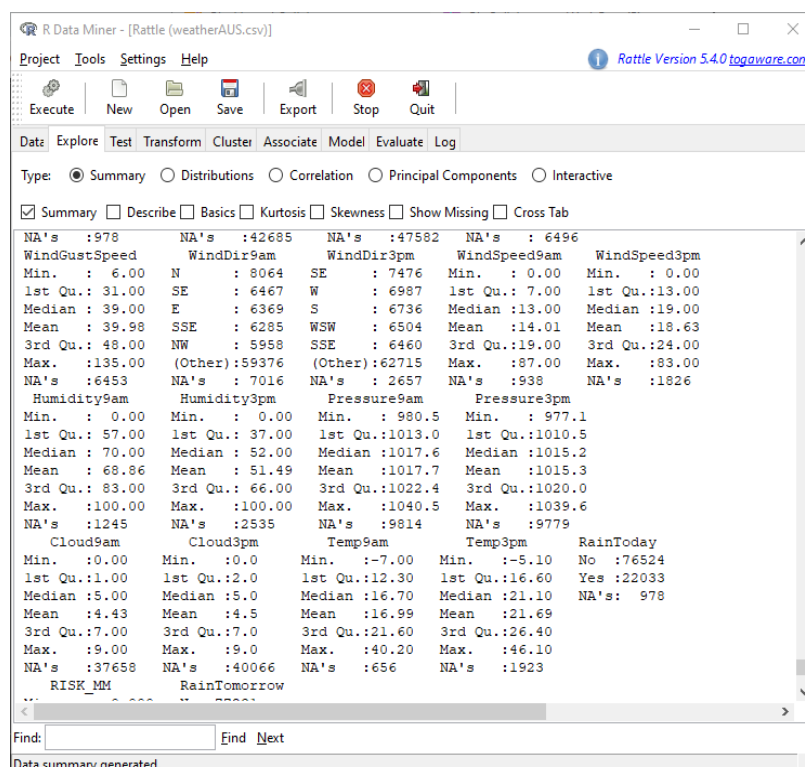
- After the R Data Miner window opens, download [weatherAUS.csv](#) data file from any browser and under the filename tab, locate and open the file. The data will be visible
- Now select Rain Tomorrow as target. After selecting the target, execute the data. Every time we make changes like changing the

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

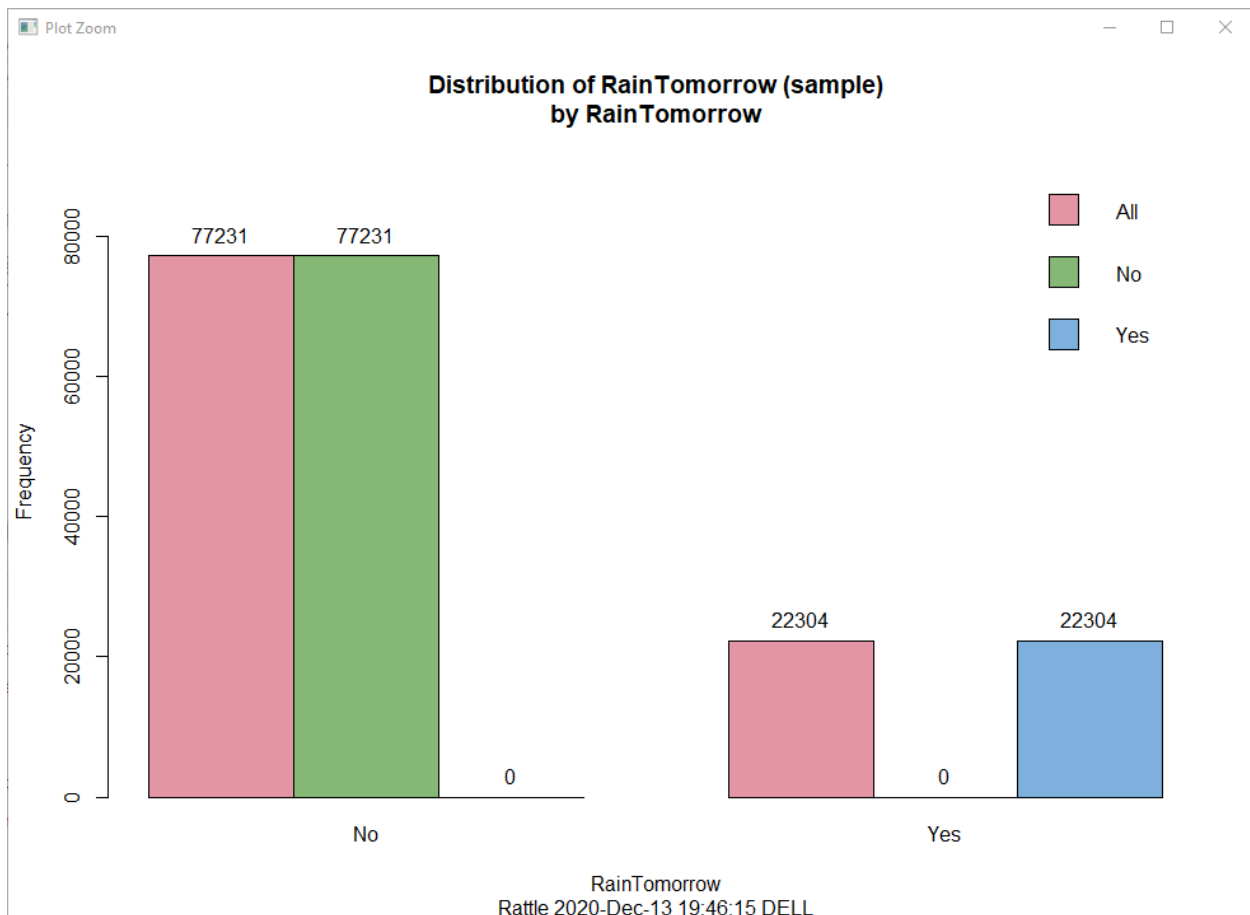
targets, inputs, risks etc. execute the data so that the changes are implied.



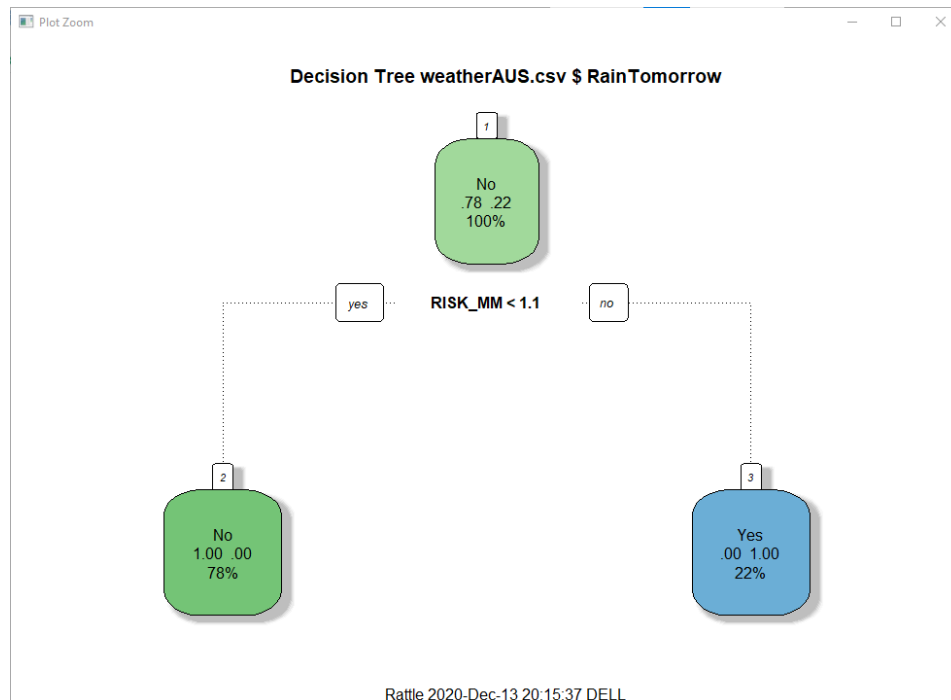
- For the summary statistics, under the data tab select the required variables as inputs. Execute the data. Under the explore tab, select summary from type and execute the selection. The summary of the given data is :



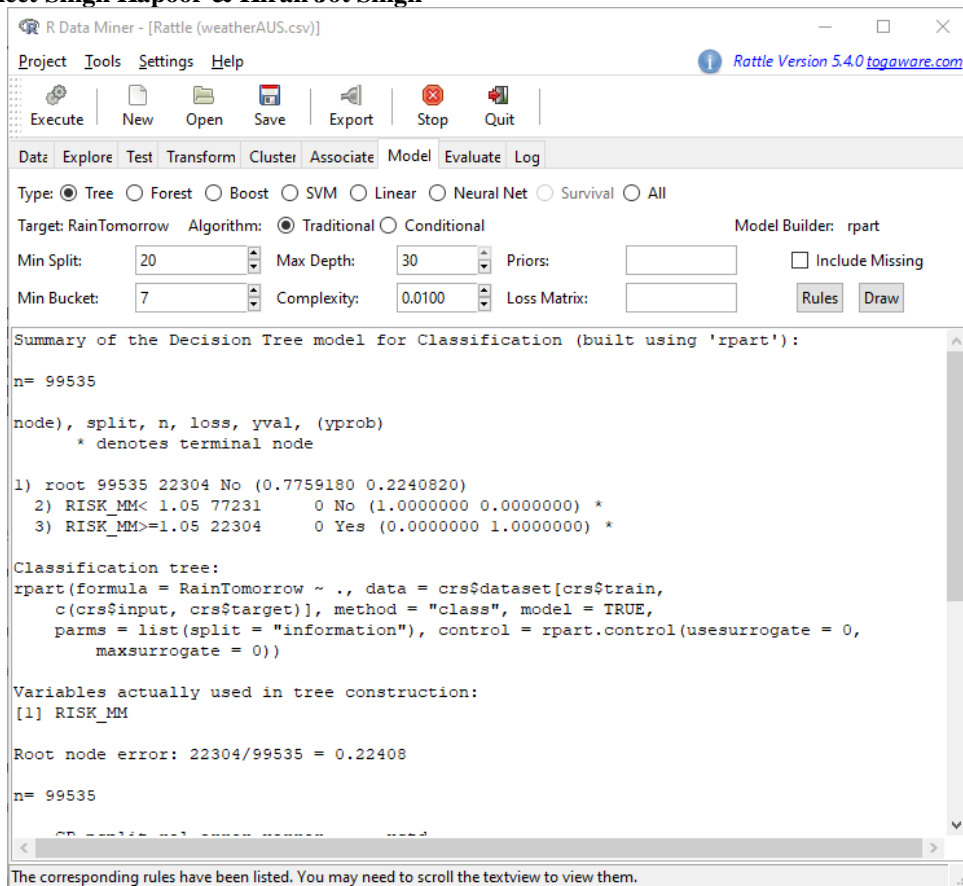
- To get the bar plot for Rain Tomorrow, select Rain Tomorrow as a target. Under the explore tab, select distribution from type and tick on bar plot for Rain tomorrow, execute the selection. The graph will be visible on the RStudio plot area. Click on zoom to view the enlarged plot.



- To get the decision tree for Rain Tomorrow, select Rain Tomorrow as a target. Under the model tab, select tree from type and execute the selection. Two options i.e. Rule and Draw will be visible in the target window, select draw and again execute the selection. The decision tree will be visible in the RStudio plot area. Click on zoom to view the enlarged plot.



- To get the rules for which the decision tree of Rain Tomorrow is made, select Rain Tomorrow as a target. Under the model tab, select tree from type and execute the selection. Two options i.e. Rule and Draw will be visible in the target window, select rule and execute the selection again. The rules will be visible in the R Data miner window.



R Data Miner - [Rattle (weatherAUS.csv)]

Project Tools Settings Help Rattle Version 5.4.0 togaware.com

Execute New Open Save Export Stop Quit

Data Explore Test Transform Cluster Associate Model Evaluate Log

Type:  Tree  Forest  Boost  SVM  Linear  Neural Net  Survival  All

Target: RainTomorrow Algorithm:  Traditional  Conditional Model Builder: rpart

Min Split: 20 Max Depth: 30 Priors:   Include Missing

Min Bucket: 7 Complexity: 0.0100 Loss Matrix:

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

```
n= 99535
node), split, n, loss, yval, (yprob)
  * denotes terminal node
1) root 99535 22304 No (0.7759180 0.2240820)
 2) RISK_MM< 1.05 77231 0 No (1.0000000 0.0000000) *
 3) RISK_MM>=1.05 22304 0 Yes (0.0000000 1.0000000) *
```

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:

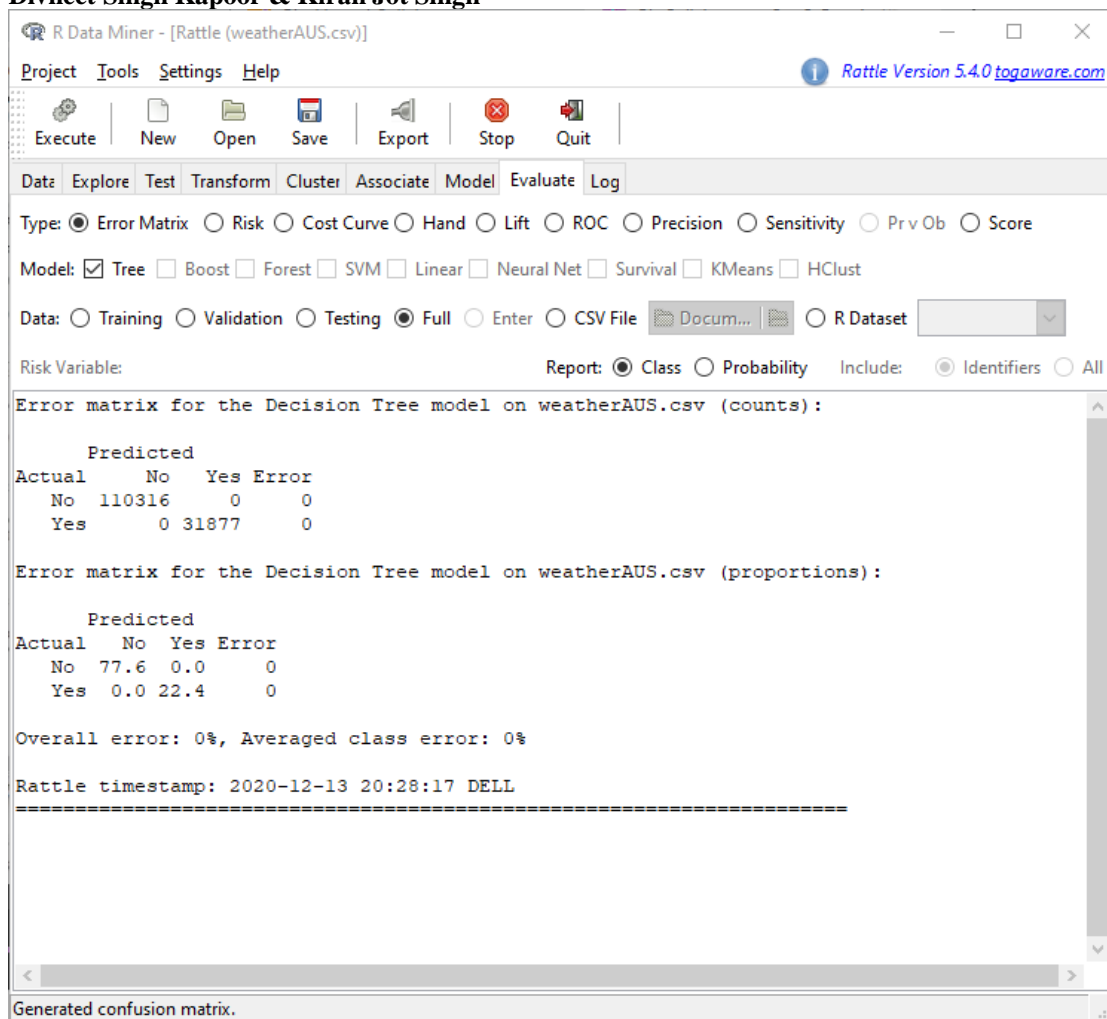
```
[1] RISK_MM
```

Root node error: 22304/99535 = 0.22408

n= 99535

The corresponding rules have been listed. You may need to scroll the textview to view them.

- To check the correctness and errors in the data, under the evaluate tab, select error matrix from type. Also select full from data tab to check errors in whole data and execute the selection. The errors will be visible in the R Data window.



R Data Miner - [Rattle (weatherAUS.csv)]

Project Tools Settings Help Rattle Version 5.4.0 [togaware.com](http://togaware.com)

Execute | New | Open | Save | Export | Stop | Quit

Data | Explore | Test | Transform | Cluster | Associate | Model | Evaluate | Log

Type:  Error Matrix  Risk  Cost Curve  Hand  Lift  ROC  Precision  Sensitivity  Prv Ob  Score

Model:  Tree  Boost  Forest  SVM  Linear  Neural Net  Survival  KMeans  HClust

Data:  Training  Validation  Testing  Full  Enter  CSV File   R Dataset

Risk Variable: Report:  Class  Probability Include:  Identifiers  All

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

		Predicted		
Actual	No	Yes	Error	
	No	110316	0	0
Yes	0	31877	0	

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

		Predicted		
Actual	No	Yes	Error	
	No	77.6	0.0	0
Yes	0.0	22.4	0	

Overall error: 0%, Averaged class error: 0%

Rattle timestamp: 2020-12-13 20:28:17 DELL

---

Generated confusion matrix.

## CONCLUSION:

Rattle continues to undergo development, extending in directions dictated by its actual use in data mining and from suggestions and code offered by its user population.

*Thank You Mam*

**Prepared By: Divneet Singh Kapoor & Kiran Jot Singh**