

# Worksheet EXPERIMENT – 8

# Name-YANA SRNA

UID-20BCS227

SECTION - 23 B

## Aim:

Explore, visualize, transform and summarize input datasets for building classification models.

## **Requirements:**

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

## **Expected Outcome:**

- Data mining combines concepts, tools, and algorithms from machine learning and statistics for the analysis of very large datasets, so as to gain insights, understanding, and actionable knowledge.
- Rattle uses the Gnome graphical user interface as provided through the RGtk2 package. It runs under various operating systems, including GNU/Linux, Macintosh OS/X, and MS/Windows.

## **Installation Steps**

- Installing R for Windows 1.
  - o Go to https://cran.r-project.org/bin/windows/base/

CHANDIGARH UNIVERSITY Discover, Legrn, Empower.

Prepared By: Divneet Singh Kapoor & Kiran Jot Singh

- o Click on "Download R 4.0.0 for windows"
- 2. Installing RStudio for Windows
  - o Go to https://rstudio.com/products/rstudio/download/
  - Click downloaded file to install (select all default options)

A . A			Project (Norma)	
O United *	-0	Environment History Connections		
1	FUR FUE 1 CT LOT SOURCE * 1	Clobal Research and a local statement a	= Lat • 143	
		Contraction -		
		Environment is emp	ty	
R Sc	ript	-		
		Data		
1:1 (Top Level) :	R Script a			
Console Terminal - Jobs -	-0			
-/> Men and unlesses to mediately the state of a section of the		Dies Diets Deckasse Links Manuar	- 7	
Type 'license()' or 'licence()' for distribution details.			ann L	
tabural language support hus supplies in an English lange				
Natural language support but running in an English local				
R is a collaborative project with many contributors.				
'citation()' on how to cita a an analysis in multication				
Type idem()' for some day	sole	Plottin	g	
'help.start()' for an HTML prowser interface to neip.			Sel.	
Type 'q()' to quit R.				

- 3. Install the following packages.
  - I. list\_of\_packages <- c("tibble", "bitops", "magrittr", "stringi", "XML", "stringr", "Hmisc", "R6", "scales", "lazyeval", "ggplot2", "corrplot", "RGtk2", "cairoDevice", "rattle")
  - II. install.packages(list\_of\_packages, repos="https://cloud.rproject.org/", dependencies = TRUE)
  - III. # Additional packages install.packages(c("tidyselect", "doBy", "ellipse", "mlbench", "amap", "ggdendro", "fpc", "randomForest", "DAAG", "arules"), dependencies = T, type = "binary")

IV. Install Rattle library(rattle) rattle()



· @	R Data Miner - [Rattle] 📃 🗆 🗙						
<u>Project</u> Tools <u>S</u> ettings <u>H</u> elp	Rattle Version 2.5.0 togaware.com						
Execute New Open Save	Report Export Quit						
Data Explore Test Transform Cluster Asso	ciate Model Evaluate Log						
Type:  CSV File ARFF ODBC R Dataset RData File Cibrary Corpus Script							
Filename: (None) 🚔 Separator: 📜 🗹 Header							
Sample Percentage: 70 Count:	0 \$\$ Seed: 123 \$\$ View Edit						
🧼 Input   🔴 Ignore Weight Calculator: 🛛	Target: Auto Categoric Numeric						
Welcome to Rattle (rattle.togaware.com).							
Rattle is a free graphical user interface for Data Mining, developed using R. R is a free software environment for statistical computing and graphics. Together they provide a sophisticated environments for data mining, statistical analyses, and data visualisation.							
See the Help menu for extensive support in using Rattle. The Togaware Desktop Data Mining Survival Guide includes Rattle documentation and is available from datamining.togaware.com							
Rattle is licensed under the GNU General Public License, Version 2. Rattle comes with ABSOLUTELY NO WARRANTY. See Help -> About for details.							
Rattle version 2.5.0 Copyright (C) 2006-2009 Togaware Pty Ltd Rattle is a registered trademark of Togaware Pty Ltd							
To Begin: Choose the Type of data source, sp	ecifically identify the source, then click the Execute button.						

### 4. Then:

- I. Click on the **Execute** button;
- II. Click on **Yes** within the resulting popup;
- III. Click on the **Model** tab;
- IV. Click on the **Execute** button.
- 5. Open the RStudio window and execute the command library (rattle),rattle() to get the R Data Miner window.



RStudio e Edit Code View Plots Session Build Debug Profile Tools Help	- 0		
🔹 🚳 🤡 📲 🔚 👘 🚱 to file/function 👘 🔛 🔹 Addins 🔹	Project: (Nor		
Console Terminal × Jobs ×	Environment History Connections Tutorial		
-10	🞯 🕞 🖙 Import Dataset 🔹 🎻		
version 4.0.3 (2020-10-10) "Bunny-Wunnies Freak Out"	Global Environment • Q		
opyright (C) 2020 The R Foundation for Statistical Computing	Data		
is free software and comes with ABSOLUTELY NO WARRANTY. 'ou are welcome to redistribute it under certain conditions. ype 'license()' or 'licence()' for distribution details. Natural language support but running in an English locale			
is a collaborative project with many contributors.			
ype 'contributors()' for more information and citation()' on how to cite R or R packages in publications.	Files Plots Packages Help Viewer		
<pre>ype 'demo()' for some demos, 'help()' for on-line help, or help.start()' for an HTML browser interface to help. ype 'q()' to quit R.</pre>	on in)   ⊅ Zoom   ≧ Export +   O		
<pre>library(rattle) oading required package: tibble oading required package: bitops attle: A free graphical interface for data science with R. ersion 5.4.0 Copyright (c) 2006-2020 Togaware Pty Ltd. ype 'rattle()' to shake, rattle, and roll your data. rattle() oading required package: Rotk2 worch went outside hinges. Try setting notch=FALSE. worch went outside hinges. Try setting notch=FALSE. worch went outside hinges. Try setting notch=FALSE.</pre>			
oton went outside ninges. Try setting noton=FALSE.			
he following objects are masked from 'package:base':			
former and unler			
lonmacipear, onica			
Sill R Data Miner - [Rattle (mtcars)] Project Tools Settings Help	Rattle Version 540 taaan		
Datz Explore Test Transform Cluster Associate Model Evaluate Log			
Source: O File O ARFF O ODBC O R Dataset O RData File   Library O Corpus O Script			
Data Name: mtcars:datasets:Motor Trend Car Road Tests			
Partition 70/15/15 Seed: 42 View Edit			
Input      Ignore Weight Calculator.     Input     Ignore Weight Calculator.     Input     Input			
No Variable Data Tune Innut Tarraet Rick Ident Innore Weinht Commant			
1 mpg Numeric			
2 cyl Numeric			
2         cyl         Numeric (a)         (b)         (c)         (			
1         may immediate         0         <			
2         cyl         Numeric         Image: 2         Image: 2         Image: 2           3         disp         Numeric         Image: 2         Image: 2         Image: 2           4         hp         Numeric         Image: 2         Image: 2         Image: 2           5         drat         Numeric         Image: 2         Image: 2			
2       cyl       Numeric       Image: 2       Image: 2         3       disp       Numeric       Image: 2       Image: 2         4       hp       Numeric       Image: 2       Image: 2         5       drat       Numeric       Image: 2       Image: 2         6       wt       Numeric       Image: 2       Image: 2			
2       cyl       Numeric (a)       C       C       C       Unique: 3         3       disp       Numeric (a)       C       C       C       Unique: 3         4       hp       Numeric (a)       C       C       C       Unique: 22         5       drat       Numeric (a)       C       C       C       Unique: 22         6       vt       Numeric (a)       C       C       C       Unique: 23         7       geec       Numeric (a)       C       C       C       Unique: 30			
2       cyl       Numeric       Image: Constraint Co			
2       cyl       Numeric       Image: 2       Image: 3         3       disp       Numeric       Image: 2       Image: 2         4       hp       Numeric       Image: 2       Image: 2         5       drat       Numeric       Image: 2       Image: 2         6       wt       Numeric       Image: 2       Image: 2         7       gsc       Numeric       Image: 2       Image: 2         8       vs       Numeric       Image: 2       Image: 2         9       am       Numeric       Image: 2       Image: 2			
2 cyl       Numeric       0       0       0       0       Unique: 3         3 disp       Numeric       0       0       0       0       Unique: 27         4 hp       Numeric       0       0       0       0       Unique: 22         5 drat       Numeric       0       0       0       0       Unique: 22         6 wt       Numeric       0       0       0       0       Unique: 29         7 qsec       Numeric       0       0       0       Unique: 21         8 vs       Numeric       0       0       0       Unique: 2         9 am       Numeric       0       0       0       Unique: 3         10 gear       Numeric       0       0       0       Unique: 3			

- 6. After the R Data Miner window opens, from the library of data sets available, select the mtcars:datasets: Motor Trend Car Road Test and execute the data.
- 7. Now select miles per gallon (mpg) as a target. Under the explore tab, select distribution from type and tick on boxplot for mpg and wt, execute the selection. The graphs will be visible on the RStudio.





8. Again keeping miles per gallon (mpg) as a target. Under the explore tab, select distribution from type and tick on histogram for mpg and wt, execute the selection. The graphs will be visible on the RStudio.



 9. For the summary statistics for mpg and wt, under the data tab select mpg and wt as inputs and ignore the rest of the variables.
 Execute the data. Under the explore tab, select summary from type and execute the selection.

rrepa	reu by: D	ivneet i	Singi	кароог		i Jot Sing		
Project T	niner - [Rattle (mtcar: ools Settings Hel	s)] D					Rattle Version 54.0 togaware.com	
Execute	New Open	Save	Export S	🔕 🖏 Stop Quit			•	
Data Explore Test Transform Cluster Associate Model Evaluate Log								
Source: O File O ARFF O ODBC O R Dataset O RData File O Library O Corpus O Script								
Data Name: mtcars:datasets:Motor Trend Car Road Tests 🗸								
Partition 70/15/15 Seed: 42 C View Edit								
🔵 Input	🛑 Ignore Weigh	t Calculator:			Target Data Type Auto () Ca	tegoric 🔿 Num	ic 🔿 Survival	
No. Variabl	e Data Type Input	Target	Risk	ldent lg	nore Weight	Comment	^	
1 mpg	Numeric 🔘	0	0	0 (	0 0	Unique: 25		
2 cyl	Numeric 🔘	0	0	0		Unique: 3		
3 disp	Numeric 🔘	0	0	0		Unique: 27		
4 hp	Numeric 🔘	0	0	0		Unique: 22		
5 drat	Numeric ()	0	0	0		Unique: 22		
6 wt	Numeric 💿	0	0	0 (		Unique: 29		
7 qsec	Numeric ()	0	0	0 0		Unique: 30		
8 vs	Numeric ()	0	0	0 0		Unique: 2		
9 am	Numeric ()	0	0	0 0		Unique: 2		
10 gear	Numeric ()	0	0	0		Unique: 3		
٢							×	
R Data N	Ainer - [Rattle (mtcar	rs)]						
		ip		8			Ruttle Version JKo <u>toguwaretor</u>	
Execute	New Open	Save	Export	Stop Qui				
Type:	Summary O Dis	tributions			Log	Interactive		
Summ	arv 🗌 Describe 🗌	Basics 🗌 Ku	rtosis 🗌 Ske	wness 🗌 Sho	w Missing 🗌 Cro	ss Tab		
Below we	summarise th	e dataset					A	
The data	a is limited t	o the tra	ining dat	taset.				
Data fra	ume:crs\$datase	t[crs\$tra	in, c(crs	s\$input, c	rs\$risk, crs	\$target)]	2 observations and 2 variables Maximum # NAs:0	
Storage mpg double wt double								
For the simple distribution tables below the 1st and 3rd Qu. refer to the first and third quartiles, indicating that 25% of the observations have values of that variable which are less than or greater than (respectively) the value listed.								
mg Min. lst Qu. Median Mean 3rd Qu. Max.	eg :10.40 Min. :15.05 lst :18.95 Medi :18.94 Mean :21.48 3rd :32.40 Max.	wt :1.935 Qu.:2.660 an :3.440 :3.425 Qu.:3.825 :5.424						

**CHANDIGAR** UNIVERSITY

## **CONCLUSION :**

Find:

Rattle timestamp: 2020-10-30 19:40:19 DELL

<u>Find</u> Next

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.



