Data Mining Prof. Arindam Dasgupta Department of Computer Science and Engineering Indian Institute of Technology, Kharagpur

Lecture – 43 Tutorial

Hello. Welcome to data mining with R course. I am going to present how to apply the R programming language in application of data mining, I am a research scholar in compressors department, my name is Arindam Dasgupta.

(Refer Slide Time: 00:40)

	Introduction
• The comp • Ir • R • It • F • R	e R is a functional language for statistical butation and graphics nitially developed by Ihaka and Gentleman lich of Statistical functions which are indispensable for data mining is a scripting language ree to download under the GNU General Public license uns on all platforms like Windows, Mac OS X and Linux
۰The	re are lots of packages related to data mining are
availa	able
• Fr • Cl • Cl • Ti	equent patterns and associations rules assification and predictions uster Analysis me Series data mining
	Data Mining With R

Now, one of what is R? Basically, R is a programming language, it is basically a functional language for statistical computation and graphics, initially, it is developed by Ihaka and gentleman, it is a riched with lots of statistical functions which are very essential for data mining applications. Basically, it is a scripting language and this tool can be downloaded freely from under the GNU license and you can execute run or run the R programming in any platform such as windows MAC, OS, X and Linux because it supports all the platforms. R is very beautiful language because it is basically functional language each functions are already implemented in the R library and we just use that functions for our pass passes.

There are lots of packages data mining available data mining packages are available these packages are frequent pattern analysis and rules as described in the theoretical class and classification prediction algorithms cluster analysis packages, there are time series data mining packages there are lots of packages available just you have to download the package in real time and just apply the functions on that packages.

(Refer Slide Time: 02:23)



Now, how to get the R tool? To get the R tool, we have to go to the this website is called this website is called cran project org bin windows or base just go to this link at first, then download the R according to your platform in my case, these windows platform that is why I have downloaded R win 32 exe after downloading the software, we have to install the software, it is very easy to install just click on the R exe icon, I will show you the this is the this is the R software, you just double click on; this icon and just install by applying the next I have already installed here.

(Refer Slide Time: 03:48)



After that after installing the R studio, we have to install another software application that is basically intergrated development environment it is also free and open source id, it is called R studio, it is very user friendly, it provides an user friendly interface to execute your R programming language. I will show you the how to what is the interface of the R studio. These are interface of R studio just I am basically, this part of the top left one part of this area is used to write the programming language R programming language the bottom part is the basically the console of R console of R executing the R script.

Basically, this is the editor and these are R console and right hand side will display, this part top right hand side this part will display the local variables the which are just defined in the R terminal and the bottom of portion you can get the what is this is the files which are in this computer and any plot will be displayed here.

(Refer Slide Time: 05:19)

nsole Terminal ×	5	Environment	History	Connecti	ions_ 🗆
MiningData/ 🔅		er 🔒 🖻	Import Datase	et = 🛛 🥑	≣ List
		🔒 Global Env	ronment •	Q,	
		En	vironment is	empty	
		Files Plots	Packages	Help	Vinner
		New Folde	r O Deleti	e 🍺 Rer	name
		A 1	lame		
		693	51330-graph	ics-wallpr	apers.jpg
		🗆 📁 Cus	tom Office Te	mplates	
		Document.rtf			
		U D Pres	entation1.pp	e	

(Refer Slide Time: 05:28)

🔹 🖓 💣 • 🔚 🗐 🗇 👘 Go to file/function 🔤 🕅 • Addins •			🙁 Pro	ect: (None)
Console Terminal × C/MiningData/ A	5 /	nvironment	History Conner Import Dataset •	tions
		Global Env	ironment • Q,	
		Er	wironment is empty	
	F	iles Plots	Packages Help	Vinser
	R	Install	Update Q	
		Name	Description	V
	U	ser Library		
		arules	Mining Association Rules and Frequent Itemsets	1.6- ⊚ 0
		asse	Easy Pre and Post Assertions	0.2.(()
		BH	Boost C++ Header Files	1.66 () 1
		bindr	Parametrized Active Bindings	0.1.1 ()
		bind	An 'Rcpp' Interface to Active Bindings	0.2 ()

These are package package list; these are which can be used in my program in our program and it the help if any R resource is needed then you can help get help from this link.

And now I want to show you some basic R program at first to click file create a 1 R script new file, I have created one new file and I have just put the script here each scripts are suppose first line in first line of the script, it is.

(Refer Slide Time: 06:31)

Until	fed1* ×	- 7	Enviro	onment	History	Connecti	
			-		moost Datase		= the
		Ge Kull 1 Ge 1 Ge Source • 1 G			mport Datase		= USI
23	}		G	obal Envir	onment •	4	
24	hw.fl()						
25	#Set Working directory						
27	setwd("C: /MiningData")			Env	ironment is	empty	
28	Sector Crymininguaca						
29	#Clear Workspace						
30	rm(list=ls())		Files	Plots	Packages	Help	Vinner
31			ON	w Folder	O Delete	- Re	name (
32	# Clear Console:			line	a a a a a a a a a a a a a a a a a a a		ingenting in the
33	cat("\014")		- 14	Home			
34	freed on fills			A N	ime		
35	#read csv Tile	1	U	6935	1330-graphi	cs-wallpi	apers.jpg
30	Salaries		0	Custo	om Office Te	mplates	
38	summary(Salaries)	1		Deau	ment of		
38:18	(Top Level) :	R Script 0	0		mencru		
	The second se		0	greti			
Console	Terminal ×	= 0		Prese	ntation1.ppt		
C:/Mini	ngData/ 🕫		0	🗐 R			
•							

(Refer Slide Time: 06:33)

nees Re Cais les hes Sees het Delay hets hes hes 1 • Oli Cais les hes Sees het Delay hets hes hes 1 • Oli Cais les hes Sees het Delay hets hes hes	R Project (None)
© Unobled1 × □□ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □ □	Environment History Connection
<pre>1 x < 2 2 y <- exp(-x²/2)/sqrt(2*pi) 3 x 4 y 5 6 #Vectorization of multiple data 7 x <- c(4, 7, 73, 5, 76, 2, 80)</pre>	Global Environment Q Environment is empty
<pre>8 x <- sqrt(y) 9 10 #plot collection or vector 1 x <- c(-2, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2) 12 v 13 y <- exp(-x^2/2)/sqrt(2*pi) 14 y 15 plot(x, y) 16 lines(x, y) 17</pre>	Files Pickages Help Vim/EF New Folder Delete Rename Image: Constraint of the second se
1:7 (lop Level) 2 R Script : Console Terminal ×	greti
C/MiningData/ /2	

It is showing x left arrow two here x is the object basically, not is exactly like very well, basically, it is a object it stores the very value 2 in the x variable.

(Refer Slide Time: 07:16)



Suppose I am to execute this in R console. Now, I want to execute this after executing this value two is has been assigned to object x, I want to view this value just type x here and write which showing the value of x is 2 ok.

Then I want to assign one expression, I am executing directly from my source script by clicking here just you have to just put the cursor here and click run, it will execute a particular command you see these executing this particular command is written here from this portion to this portion and now I want to view; what is the value of y, this is the value of y, you see, this is a basically expression e to the power minus x square by 2 and by sqrt 2 star pi, if I want to know that what is the value of pi you just type the pi value it is constant; it is already defined in R and then what is the;

Now, I have to show how to store the multiple datas into a particular object suppose I am to executes this v these vectorization, here see first bracket start within the bracket, there are lots of values 4 7; these are the values different types of values here; basically it it is a it sees for collection it is a collection of datas; this data has been stored into the object v.

(Refer Slide Time: 09:54)



Then after then after storing the value into the v, I want to display the value of v into the x into x. Now see now each of the value has been square rooted. Now check the value of x here, it is showing the each value is square rooted; that means, the vectorization on multiple data and the square root function has been applied in each of the data into the collections.

(Refer Slide Time: 10:49)



Now, I want to plot how to plot multiple vector data; suppose, first I store a collection of data into the x variable these the collection of data, then I will display the x variable, then

I apply this formula into each of data set here in x then plot, then I will display the value of y, then I plot the data set.

(Refer Slide Time: 11:47)



Now, start now it is already stored, then it is displaying the value of x object. Now, it is displaying, it is computing the values of y object, then it is displaying the values of y objects, I want to plot x data of x and y just for plotting the data of x and y just use plot within x axis and y axis.

(Refer Slide Time: 12:23)

Untitled1* ×			Environment	History	Connections
🗢 🔄 🔄 Source on Save 🔍 🎢 🛛 🗌	Run 🐤 Source •	2	e 8 10	Import Datase	e - 1 🗶 🗏 1
0 #plot collection or vector			Global Em	vironment •	2
1 x < - c(-2, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2)			Values		
2 X 3 y - eve(-v/3/3)/cost(30ni)		1.1	varues	num [1.5]	1 4 7 23
4 v		- 11-	v	num [1.0]	1 -2 -1 5
5 plot(x, y)		- 1-	*	num [1:0]	0 054 0
6 lines(x, y)			у	num [1:3	, 0.034 0
7			Files Plots	Packages	Help View
8 #User-defined functions				7	
9 nw.ri <- function()	86		1 1 1 P	200m 1 200	xport • •
CL (Top Level) =	R Sc	npt =			
sole Terminal ×					
MiningData/ 🔅		1			
2.000000 2.645751 4.847680 8.729261 8.944272			_		
plot collection or vector			- 30	0	0
<pre>c <- c(-2, -1.5, -1.0, -0.5, 0, 0.5, 1.0, 1.5, 2)</pre>			> 0 -	0	0
20 15 10 05 00 05 10 15 20			-	0	0
-2.0 -1.5 -1.0 -0.5 0.0 0.5 1.0 1.5 2.0			8 -		
<- exp(-x+2/2)/341 ((2-p1)			o		
0.05399097 0.12951760 0.24197072 0.35206533 0.39894	228 0.35206533 0.24197072			-2 -1 0	1 2
0.12951760 0.05399097					

Now, it is plotting here.

(Refer Slide Time: 12:30)



It is showing; these are x axis and these are the y axis, I want to know I want to connect these points with line just typed lines there is a command lines. Now it is connected.

(Refer Slide Time: 13:00)



These are the very basic implementation of plotting which will be used in data mining application.

Now, we can define functions in R programming; these are the structure just write a function name here and then define it is a function the function name is this and function definition, I just generate use one object hw, then type insert one string here in this object

and display the string and this is the this is the calling of function. Now at first I have to define the function. Now function is defined. Now I want to call the function.

(Refer Slide Time: 14:08)

D Route for fail fail fail for the finite half faile halfs for the		1012
♥ • 🎕 🚰 • 🔒 🔛 🚔 🔿 Go to file/function 🔡 🖾 • Addins •	•	Project: (None)
Unitied1* ×		Environment History Connections
19 hw.t1 <- function()	Being 10 10 source 11 -	Global Environment • Q
20 * (w <- "Hello world" 22 hw 23 } 24 5 hw.fl()		v num [1:5] 4 7 23 x num [1:9] -2 -1.5 y num [1:9] 0.054 0 Functions
26 #Set Working directory 28 setwd("C:/MiningData") 23. [(bplewe):	R Script ©	Files Plots Packages Help Vinner
Console Terminal × C/MiningData/ 🕫	-0	
<pre>+ } Warning message: In doTryCatch(return(expr), name, parentenv, handler) : invalid graphics state >#User-defined functions > hw.fl <- function() + { + hw <- "Hello World" + hw } ></pre>		× 000 -2 -1 0 1 2 x
0 0 🖄 0 0 0 0 0 0		

Here it is showing the hello world after by calling the function.

(Refer Slide Time: 14:21)

Untitle	d1* ×	-0	Environment	History	Connectio	ons_ C
	🔊 🗧 🖸 Source on Save 🔍 🧨 🔹 📋	👄 Run 🐤 🕞 Source 🔹 🛎	C 8 8	Import Datas	et = 🧹	≣ List
21	hw <- "Hello World"	1	Global En	vironment •	0.	
22	hw			num [1.5	1 4 7	22
23			V	num [1:5	1 4 7	1 5
25	w, f1()		X	num [1:9	1 0 05	1.5
26		1	y	num [1:9	1 0.05	4 0
27	Set Working directory	1	Functions	,		
28 1	setwd("C:/MiningData")		Files Plots	Packages	Help	Vins
30 4	Clear Workspace		000	Zoom -2	Emort •	0
27:1	(Top Level) \$	R Script ©		Estern -	captors	
		_				
onsole	Ierminal ×	= 0				
2/Minin	gData/ 🕫					
doTi inva #User hw.f: { hw hw	ryCatch(return(expr), name, parentenv, handler) : id graphics state r-defined functions l <- function() <- "Hello World"		y 0.05 0.30	-2 -1 (8 8 8	2
hw.f	10	1				
11 "H	allo World"					

Now, we have learned how to store the values into the objects.

(Refer Slide Time: 14:26)



And how to plot data, then I want to access one data file for data mining, you have to access multiple data files or a single file for the analysis purpose for that we have to set one working directory in warren in in my working directory I have stored multiple data files in c drive, I have created one folder data mining data and within these folder I have stored lots of csv files for analysis these are the csv files. Now, I want to retrieve the csv files through R programming.

(Refer Slide Time: 15:28)



At first, I have to set the working directory because my data files are stored in mining data folder under c drive. Now I execute this instruction.

(Refer Slide Time: 15:59)

• 🤫 💣 • 🔒 🔒 🦾 🍂 Go to file/function		Project: (None)
Untitled1* ×	- 0	Environment History Connections
💷 🖉 📄 Source on Save 🔍 🎢 🛛 🗌	🔿 Ryp 🐤 🔿 Source 🔹 ≷	😅 🔒 🐨 Import Dataset 🔹 🥑 🗏 List
25 hw.f1()		Global Environment • Q.
26		in 11.51 4 7 33
27 #Set Working directory		v num [1:5] 4 / 23
<pre>28 setwd("C:/MiningData")</pre>		x num [1:9] -2 -1.5
29		y num [1:9] 0.054 0
30 FCTear Workspace		Functions
31 rm(list=ls())		
32 33 # Clean Canceler	U. U.	Files Plots Packages Help Vinner
34 cat("\014")		💿 🌼 🔎 Zoom 🖓 Export • 🔘 🔞
30:1 (lop Level) \$	R Script \$	
	_	
onsole Terminal ×	= 0	
:/MiningData/ 🔅		
#User-defined functions		-
hw.f1 <- function()		8 8 9
(NOT & A
hw <- "Hello World"		
hw		8 - 2
1 00		
hw.fl()		2 1 0 1 2
Ij Hello world		-2 -1 0 1 2
"Set working directory		
second c./mininguaca /	1	x

Then I have to clear my workspace these workspace these clear these variables and for that you have to use rm list equals to ls now execute this query now my environment is cleared after that clear my this console these R console.

(Refer Slide Time: 16:28)

			Troject (Nor	
Untitled1* ×	-0	Environment	History Connections	
📖 🖉 🔚 🖸 Source on Save 🛛 🔍 🎢 📲 🗍	🔿 Run 📑 📑 Source 🔹 🛎	💣 🔒 🖙	Import Dataset 🔹 🔏 🗏 Li	
27 #Set Working directory		🜗 Global Envi	ironment • Q	
4:12 (Top Level) \$	R Script @			
onsole Terminal ×	-0	E.	uireenent is enerty	
/MiningData/ 🖗		Environment is empty		
hw <- "Hello World"				
hw		Files Plots	Packages Help Vinn	
}			Zoom - Export - 0	
invalid graphics state				

Clear this console, you have to use this command is cat slash 0 1 4 within string, just execute this command for clearing your console. Now my console is cleared now.

Now, second type we have to read the there is a file salaries dot csv in my data mining directory we have to read this file.

(Refer Slide Time: 16:31)

Life Code	Vew Pats Service Build Debug Politik	Tools Hala							0.0
. 0	8 🕣 B B 🖨 📥	Go to file/function	🛛 • Addins •					😮 Proje	ct: (None)
0 Unti	sled1* ×				-	Environment	History	Connecti	ions_ 🗆
	🖉 🔒 🗌 Source on S	ave 🔍 🎢 • 🗌		🔿 Run 🐤 📑 Source	• ≥	🕝 🔒 🖻	Import Data	vet • 🧃	🗏 List •
27 28 29	#Set Working dire setwd("C:/MiningDa	ctory ata")				👊 Global Env	ironment •	Q,	
30 31 32 33	<pre>#Clear workspace rm(list=ls()) # Clear Console:</pre>					Er	vironment i	s empty	
34 35 36 37 38	<pre>cat("\014") #read csv file Salaries=read.csv Salaries</pre>	("Salaries.csv")]	Files Plots	Packages Zoom -2	Help Export •	Vinser O
34:12	(Top Level) \$				R Script ©				
Console	Terminal ×	9							
C:/Min	ingData/ 🔟								
+ } > hw. [1] " > #Se > set > #Cl > rm(f1() Hello World" t Working director wd("C:/MiningData" ear Workspace list=ls())	y)							

Through this command there is inbuilt function read dot csv.

(Refer Slide Time: 16:44)

Untit	led1* ×	Environment History Conn			
	🙇 🗧 🖸 Source on Save 🔍 🗡 🛛 📋	😁 Run 🔭 📑 Source 🔹 🖹	😅 📊 🐨 Import Dataset 🔹 🔏 📃 List		
22	110		Global Environment • Q		
24	1				
25	hw. f1()				
26			Environment is empty		
27	#Set Working directory		Environment is empty		
28	<pre>setwd("C:/MininoData")</pre>				
29	,				
30	#clear Workspace		Files Plots Packages Help Vinn		
31	rm(list=ls())		Com Pinnet - 0		
32			the second second second second		
33	# Clear Console:				
34	cat("\014")				
35	1				
36	Fread csv file	1			
37	Salaries=read.csv("Salaries.csv")				
38	Salaries				
39	summary(Salaries)				
36:1	(Top Level) \$	R Script ©			
nsole	Terminal ×				
/Mini	ngData/ 🔅				

It is meant for accessing the csv file and after executing this command entire data will be stored in salary objects this salary object now salary objects is.

(Refer Slide Time: 17:21)



These salary objects you see destroying the tables.

(Refer Slide Time: 17:25)

Unoted	1.×	Salaries	×					-0	Environment History Connections
	0 I Y	Filter						q	🥶 🔚 💷 Import Dataset 🔹 🗶 📃 Li
×		rank	discipline	yrs.since.phd	yrs.service	sex	salary		Global Environment • Q
1	1	Prof	8	19	18	Male	139750	0	Salari 307 obs. of 7 v
2	2	Prof	В	20	16	Male	173200		• Salari 597 005. 01 7 4
3	3	AsstProf	В	4	3	Male	79750		
4	4	Prof	8	45	39	Male	115000		-
5	5	Prof	В	40	41	Male	141500		Files Plots Packages Help Vinn
6	6	AssocProf	В	6	6	Male	97000		toport • 1
7	7	Prof	8	30	23	Male	175000		
8	8	Prof	В	45	45	Male	147765		
9	9	Prof	В	21	20	Male	119250		
10	10	Prof	8	18	18	Female	129000		
owing 1 to	11 of 35	07 entries							
onsole	Termina	l×						-0	
/Mining	Data/ 🕫								
#reau	CSV I	i ie	"Salaries	("va)					

Now, I want to display the salary objects in my R console.

(Refer Slide Time: 17:28)

Untitle	d1" ×	Salaries	×					-3	Environment History Connections
	217	Filter						Q.	💣 🔒 🖙 Import Dataset 🔹 🥑 🗏 Lis
^	x °	rank	discipline	yrs.since.phd	yrs.service	sex :	salary		Global Environment • Q,
1	1	Prof	в	19	18	Male	139750	ő	Data
2	2	Prof	8	20	16	Male	173200		O Salari 397 obs. of 7 v
3	3	AsstProf	в	4	3	Male	79750		
4	4	Prof	8	45	39	Male	115000		
5	5	Prof	В	40	41	Male	141500		Files Plots Packages Help Vinn
6	6	AssocProf	в	6	6	Male	97000		💷 💷 🎢 Zoom 🖓 Export • 🖓
7	7	Prof	В	30	23	Male	175000		
8	8	Prof	В	45	45	Male	147765		
9	9	Prof	В	21	20	Male	119250		
10	10	Prof	8	18	18	Female	129000		
11	11	AssocProf	8	12	8	Male	119800		
12	12	AsstProf	8	7	2	Male	79800		
13	13	AsstProf	8	1	1	Male	77700		
14	14	AsstProf	В	2	0	Male	78000		
wing 1	to 11 of 35	97 entries							

(Refer Slide Time: 17:48)

Ander Die der der Ante Ante Ante Ante Ante Ante		lol4
I • • • • • • • • • • • • • • • • • • •		Project: (None)
Untitled1* ×	-0	Environment History Connections
(all) a Gource on Save Q / - C	🗝 Run 🐤 🖙 Source 🔹 ≷	🞯 🔒 🖙 Import Dataset 🔹 🥖 🗏 List
1/ 18 #User-defined functions		Global Environment • Q
19 hw.fl <- function() 20 - {		Data
21 hw <- "Hello world" 22 hw		Salari 557 005. 01 7 V
23 3		
25 hw.fl() 26		Files Plots Packages Help Vinne
27 #Set Working directory		
<pre>28 setwd("C:/MiningData") 29</pre>		
30 #Clear Workspace		
31 rm(list=ls()) 32		
<pre>33 # Clear Console: 34 cat("\014")</pre>		
35 36 Bread cay file		
37 Salaries=read.csv("Salaries.csv")		
38 Balaries		
<pre>39 summary(Salaries)</pre>		
38:1 (Top Level) \$	R Script ©	
Console	80	
		a

(Refer Slide Time: 18:04)

• 🚳 🐨 • 🔚 🛃 📼 🦯 🏘 Go to file/function	Project: (Non
wree enolo Terminal × 'MinimpOsts/ # # # # # # # # # # # # #	Environment History Connections Image: State of the

(Refer Slide Time: 18:13)

ource							80	Environment History Connections
onsole	Terminal ×						-6	Global Environment • Q
C:/Mining	Data/							Data
22 122	Prof	A	32	32	Male	124309		Calari 307 abs of 7 v
23 123	Prof	A	24	22	Male	97262		U Salari 397 005. 01 7 V
24 124	AssocProf	A	25	22	Female	62884		
25 125	Prof	A	24	22	Male	96614		
26 126	Prof	A	54	49	Male	78162		
27 127	Prof	A	28	26	Male	155500		Files Plots Packages Help Vinne
28 128	AsstProf	A	2	0	Female	72500		Com Parate Q
9 129	Prof	A	32	30	Male	113278		the second restriction of the
30 130	AsstProf	A	4	2	Male	73000		
31 131	AssocProf	A	11	9	Male	83001		
32 132	Prof	A	56	57	Male	76840		
33 133	AssocProf	A	10	8	Female	77500		
34 134	AsstProf	A	3	1	Female	72500		
35 135	Prof	A	35	25	Male	168635		
6 136	Prof	A	20	18	Male	136000		
7 137	Prof	A	16	14	Male	108262		
8 138	Prof	A	17	14	Male	105668		
39 139	AssocProf	A	10	7	Male	73877		
0 140	Prof	A	21	18	Male	152664		
1 141	AssocProf	A	14	8	Male	100102		
2 142	AssocProf	A	15	10	Male	81500		

Now, it is showing in R console suppose, I want to summarize this salary summarize means, it will display the; suppose, there is a very well called discipline in this very well, there are 2 values A and B, it contains 181 values of A and 2 16 values of B. Similarly, these are basically nominal value and the numeric values in numeric values, it is showing the what is the minimum value what is the maximum value what is a quartile what is the median value? Mean value of each attributes in salary the minimum salary is this maximum salary is this and mean salary is this and it is it is showing all the summary information summary information of each attribute.

(Refer Slide Time: 19:38)



Now, I will leave the basics of R programming which are needed to needed in any time any type of determining algorithms enough.

Now, at first I will discuss about how to use apriori algorithm for accessing association rules from a data set basically apriori algorithm, we have learned it finds the frequent data items with minimum support in I will show, I have a grocery data set and from that grocery data set, we just access the rules with support 0.006 and confidence 2.5 after generating the rules, we will store the rules into a csv file and for apriori algorithm there are the input is one file it is called candidates itemset and output is the frequent itemset ok.

To execute apriori algorithm we have to use one package it is very essential it is called arules to download this package you just go to at right hand side click on install. (Refer Slide Time: 21:24)



And type arules here arules here, it is displaying I selected the arules here. Now in install the arules packets from this terminal, after it is already installed in my system after installing the library installing the package, sorry, we have to include the library in R program for including library, just write library and type the package name and now I am including the library.

(Refer Slide Time: 22:24)



It is included; it is included smoothly

(Refer Slide Time: 22:52)

		4.0	trus fruit																		
	50																				
A		0					-				M	 0	9		ų	Y			- 10	-14	- 16
trus tructuemo fo	or warding	rereatly so	ups																		
obical is holling	comee																				
hole milk																					
p fruit ysgurt her veg whole r	oream of	te neal sp ie long life	teads bakery p	tube																	
hole mil butter da/buns	yopet	rice	strativ	e cleaner																	
her yes LMI mil	a relation	relation	he biguage (appetizer)																	
steel plants																					
hole rol cereals																					
opical frother w	and the late	a bottled	e chocola	64																	
trus fruit tropical	trubole n	albutter	and	viget	for	bottled	dahes .														
ref				1																	
nkfurte rolly/by	reads																				
cken tropical	hut																				
ther sugar	fruit/ve	pi reuspa	iers .																		
it/vegetable jui																					
Raged fruit/vey	petables																				
colume																					
cialty bar																					
ter vegetables																					
ther mil pastry																					
ole milk																					
opical fr cream d	N process	ec deterge	rt newspa	pers																	
pical fr root veg	pr other ve	g hozen d	e rolty/by	reflow	sweet s	pr salty and	Cwaffes	candy	bathroo	m deaner											
Ceo a carried																					
1																					
rafa stantan	PE 9008	and the second	· · · ·																		
er vegetatives																					
en presona	mart/ve	fe caused	n newspa	de nobbe	4 ords																
of several	a sources	a special	hand		a subline																
only may a	d courses	in spears	Dutter	a rigera	a napers																
vege occur ve	other of	or prover ag	i and	Aland an	a sector	shears a	damer														
d annes	determ																				
ty unde	- and																				
Amerialia																					
and hear	-																				
visited other up	a whole a	theread in																			
a fruit puriellar	a neurola	Lars.																			
uses indu/bu	reads	canned	w special	h Inhospin	a bass																
scal fr mot ver	ar whole a	siyoput.	domasi	to brown b	ripathy	new	cereals	10744	soda	wattes	anty										
ties yogut		1.40				-					1										
	eries	10																			

Then I have to read the read the grocery csv file, the grocery csv file is like that these are grocery csv file.

(Refer Slide Time: 22:53)

AL HOME INDUT	PAGE LA	YOUT	FORM	ULAS	0A1A	10.0	. V	EW	KSTAT																			Sprin
* I N .	1. 6	citrus (vit.																									
A 1	1	ç		p				Q.	н	1				м	N	0	q	1	1	1	ų.	v	w	x	v	1	AA	
citrus fruit	semi-fo	(app	14	4110	Ø5																							
tropical fruit	yoput	coffee																										
whole milk																												
pip fruit	yipri	(ream	(he me	al spre	ads .																							
other vegetables	whole r	doorde	se kon	glifeb	akery p	oduct																						
whole milk	butter	yapri	rice		shrative	deater																						
rolls/burs																												
other vegetables	UH-mi	rola/b	un bei	tied by	i hquor ()	opetice	1																					
potted plants																												
whole mile	cereals																											
tropical fruit	other ve	giahital	ire bot	tied w	chocola	te .																						
citrus fruit	tropical	t whole	nibd	ber .	eard	vier	60	1	bottled -	i dahas																		
beef																												
Arankhurter	nh/h	shoe's																										
chicken	topical	hut																										
butter	1100	fruit/e	-		10																							
hut/vegetable luice	1		-																									
parkaged fruit/ungetables																												
chocolate																												
specialty har																												
other vegetables																												
butter mile	andre .																											
whole mile	in the second																											
Internal Profit		-	-	-	-																							
Interest front	and see	- star				and the set	-		-	-	and.	hatberry	m danaser															
hotfied water	canned.		4				- 11		10.1 1.0		and .																	
and a set		-																										
Address	a da da a	and a	1.00	-																								
sites and the	10.00		-17																									
other regelations	-	Paral.																										
prown press	1000	in and the	de cas		reaspe	percept	1.04																					
And the second s	ceverage .			ioary .	and a second	-																						
ranouller near	other ve	funda	-		Decisies.	* 1994	****	8.05																				
root vegetaties		faces.		- and a	india.	direct of	_			-																		
port	terres	or an a	4		a, dia	1. 810. 1	******	·	#1+#1+¥	ceaner																		
beer	faber.	ortery	•**																									
party	1004																											
analiseBarage Ince																												
canned beer																												
root vegetables	other ve	Fault	mil del	nert.																								
citrus truit	zwiebal	newsy	apers.																									
ranada	ren/bu	spote	Car	ned by	special	y ranopp	det				-		-															
tropical truit	root veg	ratole	nil yo	pet .	conest	K Brown	public bet	W.	rds.	cereals	10744	soda	wattes	candy														
berries	YIPH																											
r r aracarlas																	11											
							-						-	-	-				-		-					-	-	-
			_	_	_	_	_										_											
A 140	6		6			5	11																					

These are the items which are bought together these are transaction informations it contains the transactions of items this citrus fruit semi finished bread tropical food yogurt coffee are brought together in next transaction, only whole milk in next transaction pip fruit yogurt cream cheese these are the datasets my I have to find the item sets which are bought together next in most of time most frequently. Now I have to read the data set. Now dataset is readed the dataset is stored into the grocery.

(Refer Slide Time: 24:29)

+ 9	🛛 🚰 • 🔒 🔂 🚍 🌧 Go to file/function 🔤 🔯 • Addins	•				🚯 Proj	ect: (None
Untitle	d1*×	-0	Envir	onment	History	Connec	tions_ (
	🔊 🗧 🖸 Source on Save 🔍 🎢 🛛 📋	- Reg 🐤 🔿 Source 🔹 🖻	@ 1		Import Datas	et = 🚽	(= u
15 i	inspect(groceries[1:5])		G G	obal Envi	ronment •	Q.	
12:1	(Top Level) \$	R Script ©	Data				
Console	Terminal ×	-0	0 gr	ocer	Formal o	lass	tr 0
C:/Mining	gData/ 🗇						
 rm(1) #Set setwork #inst libra oading Attachi The following 	<pre>ist=is() working directory d("C:/winingpata") tall.packages("arules") ay required package: Matrix ing package: 'arules' llowing objects are masked from 'package:base': arguitate write</pre>		Files In User 1	Plots stall (Name Library arules	Packages Update Description Mining Association and Frequei Itemsets	Help Q Rules	V 1.6- © 0
#inst	tall.packages("arules")			asse	Easy Pre and Assertions	i Post	0.2.0 ()
libra groce	ary(arules) eries <- read.transactions("groceries.csv", sep :	= ",",rm.duplicates = T)		BH	Boost C++ I Files	Header	1.66 () 1
				bindr	Parametrize Active Bindi	d ngs	0.1.10
				bind	An 'Rcpp' In	terface	0.2 0

Then next it is showing the summary of the data set.

(Refer Slide Time: 24:53)



It contains 9835 rows and 169 columns because somebody in transaction, there are they have bought 169 items that is why it is showing here these are the most frequent items whole milk vegetables rolls and soda; these are more most frequently items and these are the size of distributions transactions.

(Refer Slide Time: 25:48)

🔹 🥸 📽 🔹 🚼 🛃 📥 🛛 🔺 Go to file/function 👘 🛛 🖾 🔹 Add	ns •			0	👂 Proje	ct: (None)
United1*×	-6	Envir	onment	History (Connect	ions_ [
<pre>17 #plot the frequency of items 18 itemFrequencyPlot(groceries, support = 0.1) 19 itemFrequencyPlot(groceries, topN = 20) 20</pre>		Data	ocer	Import Dataset ironment • C	ass t	= Ust
21 #default settings result in zero rules learned 22 apriori(groceries) 23 24		Eller	Distr	Buchunger	Hala	Man
<pre>25 #set better support and confidence levels to le 26 groceryrules <- apriori(groceries, parameter = 27</pre>	arn more rules list(support = 0.006, confiden	D In	stall (Update C Description	neip D,	V
<pre>28 #look at the first three rules 29 inspect(groceryrules[1:3]) 30 31 32 3 #Tenroving model performance sorting procerv rules</pre>	iles hy lift	User I	Library arules	Mining Association Ri and Frequent	ules	1.6- © 0
<pre>34 inspect(sort(groceryrules, by = "lift")[1:5]) 35</pre>			asse	Easy Pre and I Assertions	Post	0.2.(()
<pre>35 #Finding subsets of rules containing any berry 37 berryrules <- subset(groceryrules, items %in% " 38 inspect(berryrules)</pre>	'berries")	0	BH	Boost C++ He Files	eader	1.66 () 1
39 · · · · · · · · · · · · · · · · · · ·	R Script ©		bindr	Parametrized Active Binding	gs	0.1.10
onsole	60		bind	An 'Rcpp' Inte to Active Bind	erface dings	0.2 ()

(Refer Slide Time: 25:49)

		-				
Unt	fied1* ×	=0	Envi	onment	History Connec	tions
	🖉 🔚 🗌 Source on Save 🛛 🔍 🗡 📲	📑 Run 📑 📑 Source 🔹 🖹	2		Import Dataset 🔹 🔰	🕻 🗏 Lis
2	rm(list=ls())		4.0	ilobal Env	ironment • Q	
3	#Set working directory		Data			
4	setwd("C:/MiningData")		0.0	ocar	Formal class	tr 0
678	<pre>#install.packages("arules") library(arules)</pre>		U g	ocer	Pormar crass	Cl in M
9 10	#Data Import		Files	Plots	Packages Help	View
11	<pre>groceries <- read.transactions("groceries.csv", sep</pre>	",",rm.duplicates = T)	01	nstall 🚺	Update Q	
12	summary(groceries)			Name	Description	V
14	#look at the first five transactions		User	Library		
15	inspect(groceries[1:5])			arules	Mining	1.6- 0
16					Association Rules	0
17	#plot the frequency of items				and Frequent	
18	itemFrequencyPlot(groceries, support = 0.1)				Itemsets	
19	<pre>itemFrequencyPlot(groceries, topN = 20)</pre>			asse	Easy Pre and Post	0.2.0
20	#default settings result in zero rules learned				Assertions	
22	apriori(groceries)			BH	Boost C++ Header	1.66 ()
23					Files	1
24	1			bindr	Parametrized	0.1.3 ()
14:1	(Top Level) 0	R Script 0		blad	An 'Renn' Interface	02.0
Consol		80	0	bind	to Active Biodiner	0.2 0

Now, and to plot the frequency of items these are frequency of items.

(Refer Slide Time: 26:01)



(Refer Slide Time: 26:05)



(Refer Slide Time: 27:20)



Item frequency bottled water other vegetables roll buns root vegetables these are the showing the frequency of items and to display frequency of plot this frequency of item top twenty items here is the top 20 items. Frequency of top 20 items, it is basically relative frequency in 0 to 1 scale, the most the whole milk is maximum frequency has a maximum frequency, the domestic is the minimum frequency. Now I apply the apriori. Now, apply now I have applied the apriori algorithm, it has confidence level initially the confidence is 0 8 minimum interval is 00.5 and now no rules has been generated in first iteration.

Now, we have to improve the grocery rules which support 0.6 and confidence 0.25 and minimum length is 2. Now the rules has been created, I have to inspect the rules. Now, it is showing the rules left hand side potted plants whole milk pasta depends on whole milk which support these are support. This is confidence and these are lift value, these are the rules. Now, this showing first 3 rules and to inspect by lift inspect the last last 5 rules by lift operation.

(Refer Slide Time: 29:50)

						Environment	History Connections
Console Terminal ×					-6	Global Envi	Import Dataset • 2
C/MiningData/						Data	
#Improving model per	formance sorting	grocery rules by 1	ift			grocer	Formal class tr. Q
Ths The section of th	rules, by = 11T rhs	t ([1:5]) support	confidence	lift	coun	grocer	Formal class ru Q
1] {herbs}	=> {root vegeta	bles} 0.007015760	0.4312500	3.956477	6	Files Plots	Packages Help Vinni
2] {berries}3] {other vegetables, tropical fruit,	=> {whipped/sou	r cream} 0.009049314	4 0.2721713	3.796886	8	4 4 F	Zoom 2 Export • 3
whole milk} 4] {beef,	=> {root vegeta	bles} 0.007015760	0.4107143	3.768074	6		
other vegetables 5] {other vegetables,	=> {root vegeta	bles} 0.007930859	9 0.4020619	3.688692	7		
tropical fruit}	=> {pip fruit}	0.009456024	0.2634561	3.482649	9		

(Refer Slide Time: 29:51)

	80	Environment History	Connections
control Terminal × C/MiningData/ ∅ 2/MoningData/ ∅ 3/look at the first three rules inspect(groceryrules[1:3]) Ths Ths rhs support confidence lift 2] {posted plants} ⇒ {whole milk} 0.006914082 0.4000000 2] {posted plants} ⇒ {whole milk} 0.007015760 0.4312500 3] (herbs) ⇒ * rhs support confidence * inspect(sort(groceryrules, by = "lift")[1:5]) 1hs rhs * support confidence lift 1] (herbs) ⇒ {root vegetables} 0.007015760 0.4312500 2] (berries) ⇒> {root vegetables} 0.007015760 0.4312500 3.956477	coun 6	Ervironment History	Connections
<pre>3] {other vegetables, tropical fruit, whole milk}</pre>	6		

Now, it is showing herbs; if it is herbs, then root vegetables will be in right hand side, it is berries whipped and sour cream and tropical other vegetables tropical fruit whole milk root vegetables and so on.

(Refer Slide Time: 30:18)



(Refer Slide Time: 30:25)

<pre></pre>	ude Edit Code View Plats Session Build Debug Pushe Tools Help					tel
Undeddi* Imvromment History Connectione, 20 #default settings result in zero rules learned Imvromment Imvromment Imvromment 21 #default settings result in zero rules learned Imvromment Imvromment Imvromment 22 apriori(groceries) Imvromment Imvromment Imvromment Imvromment 23 #set better support and confidence levels to learn more rules grocer Formal class tr Imvromment Imvromvromment Imvromvromvromment	🔹 🧐 💣 • 🔒 🛃 📥 🍌 Go to file/function 🔢 🗟 • Ad	ins •			👂 Proje	ct: (Non
Import Data Import Import Data Import Import Data Import I	Untitled1* ×	=	Environment	History	Connecti	ions_ (
<pre>20 21 #default settings result in zero rules learned 22 apriori(groceries) 23 24 25 #set better support and confidence levels to learn more rules 26 groceryrules <- apriori(groceries, parameter = list(support = 0.006, confident 27 29 flook at the first three rules 29 inspect(groceryrules[1:3]) 31 32 31 32 32 #Improving model performance sorting grocery rules by lift 34 inspect[sort(groceryrules, by = "lift")[1:5]) 35 i 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules, items %in% "berries") 39 fariting nubsets of rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nat. 34 % (poteed): 35 meaker 36 meaker 37 % % % % % % % % % % % % % % % % % % %</pre>	💷 🖉 🔒 🖸 Source on Save 🔍 🎢 📲 🚺	Run 😁 🕞 Source 🔹 🗏	e 🔒 🖻	Import Dataset	e - 🖌	≣u
<pre>21 #default settings result in zero rules learned 22 apriori(groceries) 23 24 25 #set better support and confidence levels to learn more rules 26 groceryrules <- apriori(groceries, parameter = list(support = 0.006, confident 27 28 #look at the first three rules 29 inspect(groceryrules[1:3]) 30 31 32 31 #Improving model performance sorting grocery rules by lift 31 inspect[sort(groceryrules, by = "lift")[1:5]) 32 33 44 #riting the rules to a CSV file 34 writing of rules to a CSV file 34 writing the rules to a CSV file 35 #difference file 36 #Gouldent Containing and the rules to a CSV file 37 #server rules and the rules to a CSV file 38 #server rules and the rules to a CSV file 39 #writing the rules to a CSV file 39 #server rules and the rules to a CSV file 39 #server rules and the rules to a CSV file 30 #writing the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 31 #server rules and the rules to a CSV file 32 #server rules and the rules to a CSV file 33 #server rules and the rules to a CSV file 34 #server rules and the rules to a CSV file 35 #server rules and the rules to a CSV file 36 #server rules and the rules to a CSV file 37 #server rules and the rules to a CSV file 38 #server rules and the rules to a CSV file 39 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules and the rules to a CSV file 30 #server rules rules to a CSV file 30 #ser</pre>	20		' 🚺 Global Env	ironment •	2	
<pre>22 apriori(groceries) 23 24 set better support and confidence levels to learn more rules 25 #set better support and confidence levels to learn more rules 26 groceryrules <- apriori(groceries, parameter = list(support = 0.006, confident 27 #look at the first three rules 28 #look at the first three rules 29 inspect(groceryrules[1:3]) 30 31 32 33 #Improving model performance sorting grocery rules by lift 34 inspect(sort(groceryrules, by = "lift")[1:5]) 35 i 36 #finding subsets of rules containing any berry items 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 40 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar; 446 (bplewel): 45 mode Compared Compared</pre>	21 #default settings result in zero rules learned		Data			
<pre>23 24 25 25 26 26 27 26 27 27 28 28 29 29 29 29 29 29 29 20 20 20 20 20 20 20 20 20 20 20 20 20</pre>	22 apriori(groceries)		Ogrocer	Formal cl	ace t	e (
<pre>#set better support and confidence levels to learn more rules forcer_rules <- apriori(groceries, parameter = list(support = 0.006, confidence</pre>	23		O grocer	Formal al	1000 0	1 m 10
grocerycules <- apricing/groceries, parameter = list(Support = 0.006, confident #look at the first three rules inspect(groceryrules[1:3]) #Improving model performance sorting grocery rules by lift inspect(sort(groceryrules, by = "lift")[1:5]) iffinding subsets of rules containing any berry items berryrules <- subset(groceryrules, items %in% "berries") inspect(berryrules) #writing the rules to a CSV file writing the rules to a CSV file mode @ mode	24 25 #set better support and confidence levels to]	earn more rules	grocer	Formal C	lass r	u (
<pre>Flook at the first three rules inspect(groceryrules[13]) #Improving model performance sorting grocery rules by lift inspect(sort(groceryrules, by = "lift")[1:5]) #Improving subsets of rules containing any berry items #finding subsets of rules containing any berry items #finding subsets of rules containing any berry items #finding the rules to a CSV file #write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.na; #de (top Lewel) : #finding the rules to a CSV file #finding the rules to a CSV file #rules containing any berry thems #de (top Lewel) : #finding the rules to a CSV file #finding the ru</pre>	26 grocervrules <- apriori(groceries, parameter =	list(support = 0.006, confider				
<pre>28 #look at the first three rules 9 inspect(groceryrules[1:3]) 31 32 33 33 34 35 35 36 37 berryrules containing any berry items 37 berryrules <- subset(groceryrules, items %in% "berries") 39 39 40 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv",sep = ",", quote = TRUE, row.nar. 446 (toptevent) = R Script = 446 (toptevent) = R Script = 447 (toptevent) = R Script = R</pre>	27	the second s	Files Plots	Packages	Help	View
<pre>29 inspect(groceryrules[1:3]) 30 31 4 #Improving model performance sorting grocery rules by lift 31 inspect[sort(groceryrules, by = "lift"][1:5]) 5 #finding subsets of rules containing any berry items 37 berryrules <= subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 44 writing the rules to a CSV file 44 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar; 446 (bptewell = R Script = R Scr</pre>	28 #look at the first three rules		1000 1100	The state of the s	ment	0
<pre>30 30 31 32 #Improving model performance sorting grocery rules by lift inspect[sort(groceryrules, by = "lift")[1:5]) 35 36 #Finding subsets of rules containing any berry items 37 berryrules, - subset(groceryrules, items %in% "berries") 38 39 47 #Finding the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar; 444 (Top Level) : msde</pre>	<pre>29 inspect(groceryrules[1:3])</pre>		a 917	Zoom I 😬 B	xport •	v
<pre>31 32 #Improving model performance sorting grocery rules by lift 333 #Improving model performance sorting grocery rules by lift 34 353 35 36 47 37 36 47 37 37 38 48 48 5 38 48 5 39 4 48 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5</pre>	30		0			
<pre>24 #Improving model performance sorting grocery rules by lift 34 inspect[sort(groceryrules, by = "lift")[1:5]) 35 i 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 39 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar; 446 (bptevel): msole</pre>	31					
31 Support[growing mode performance softing growing mode performode performode performance softing growing mode perform	32 #Tenroving model performance sorting grocery r	ulas by lift				
<pre>33 i i construction of rules containing any berry items 34 ffinding subsets of rules containing any berry items 35 berryrules - subset(groceryrules, items %in% "berries") 36 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar; 446 (Top Level) : R Scopt : 55 msde 55</pre>	<pre>34 inspect(sort(arocervrules, by = "lift")[1:5])</pre>	ares by fift				
36 #finding subsets of rules containing any berry items 37 berryrules <- subset(groceryrules, items Xin% "berries") 39 #mriting the rules to a CSV file 40 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar, 446 (top Level) : msde Image: Comparison of the comparison of	35 1					
37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 39 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 446 (Top Lewel) : Rsorpt : Rsorpt : nsde Image: Comparison of the subset of	36 #finding subsets of rules containing any berry	items	1			
<pre>38 inspect(berryrules) 39 4 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ", ", quote = TRUE, row.nar, 446 (Top Level) : R Script : assle 6 7 m C A La La</pre>	37 berryrules <- subset(groceryrules, items %in%	"berries")				
39 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ", ", quote = TRUE, row.nar, 446 (top level) : msde Image: Comparison of the top level	38 inspect(berryrules)					
umite(groceryrules, file = "groceryrules.csv", sep = ", ", quote = TRUE, row.na; 446 (top Level) : R Script :	40 Empiting the pulse to a CSV file					
A46 (top Level) : R Sorie : Sorie : Sorie : R	41 write(grocervrules, file = "grocervrules, csv".	sen = " " quote = TRUE, row na				
446 (Rop Lewi) 2 R Script 2	a minergrocerynanes, inte = grocerynanes.csv ,	rep = 1 1 spore = 1koc, Tourne				
	4:46 (Top Level) \$	R Script	0			
	onsole	60	3			

Now, I want to find a subset of rules containing any various items and how for that I have to use the subset function these is the it contains the rules and item is varies. Now I just execute this command after executing this command, I have to inspect this command.

(Refer Slide Time: 30:50)

Undetdd1*x Invicement Histery Connection 28 Flook at the first three rules Import Dataset • 2 29 inspect(groceryrules[1:3]) Import Dataset • 2 31 Improving model performance sorting grocery rules by lift Import Dataset • 2 32 #Improving model performance sorting grocery rules by lift Import Dataset • 2 33 #Improving model performance sorting grocery rules by lift Import Context (groceryrules, by = "lift")[1:5]) 36 #finding subsets of rules containing any berry items Import Performance sorting grocery rules. 37 berryrules <- subset(groceryrules, items %in% "berries") Import Performance sorting grocery rules. 39 #minting the rules to a CSV file Import Performance sorting groceryrules. 401 (top tervet): Import Performance sorting groceryrules. 411 (berries) => (mbiped/sour cream) 0.009049314 0.2721713 3.796886 69 Import Performance sorting sortidence lift 31 (berries) => (whole milk) 0.01074479 0.3180428 2.279848 104 Import Performance sortidence lift		ction Addins •			Project: (None)
Improving model performance sorting grocery rules by lift inspect(groceryrules[1:3]) Improving model performance sorting grocery rules by lift inspect(sort(groceryrules, by = "lift")[1:5]) Improving model performance sorting grocery rules by lift inspect(sort(groceryrules, by = "lift")[1:5]) Improving model performance sorting grocery rules by lift inspect(sort(groceryrules, items %in% "berries") inspect(berryrules) inspect(berryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nat. 401 (berries] => (whipped/sour cream) 0.009049314 0.2721713 3.796886 89 11 (berries] => (yogurt} 0.010574479 0.3180428 2.279848 104 31 (berries] => (whole milk)	Untitled1* ×		-0 6	invironment His	tory Connections
<pre>28 #look at the first three rules inspect(groceryrules[1:3]) 31 32 32 33 34 35 35 35 36 #finding subsets of rules containing any berry items 35 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules, file = "groceryrules.csv",sep = ",", quote = TRUE, row.nat. 40 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv",sep = ",", quote = TRUE, row.nat. 40 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv",sep = ",", quote = TRUE, row.nat. 40 (optewet): 52 berries] >> (whipped/sour cream) 0.009049314 0.2721713 3.796886 89 31 (berries) >> (other vegetables) 0.01076479 0.3180428 2.279848 104 31 (berries) >> (other vegetables) 0.01076479 0.3180428 2.279848 104 31 (berries) >> (other vegetables) 0.010794611 0.3547401 1.388328 116</pre>	💷 🗊 🛛 🔚 🗋 Source on Save 🛛 🔍 🎢 🗤	Run 😁 Sou	rce • 🖹 🧯	🚰 🔒 🛛 🖃 Impor	t Dataset 🔹 🔏 🗏 List
<pre>29 inspect(groceryrules[1:3]) 30 31 32 31 #Improving model performance sorting grocery rules by lift 32 32 33 #Improving model performance sorting grocery rules by lift 33 34 35 35 35 36 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 39 40 #inting the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 41 the rules to a CSV file 45 minute to the total solution of total solution of the total solution</pre>	28 #look at the first three rules		1.0	Global Environme	nt • Q,
30 31 32 32 33 34 #Improving model performance sorting grocery rules by lift 33 35 #finding subsets of rules containing any berry items 36 #finding subsets of rules containing any berry items 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 39 40 #ariting the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 421 (top tevel) : 43 marting the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 421 (top tevel) : 43 marting the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 421 (top tevel) : 42 marting the rules to a CSV file 43 marting the rules to a CSV file 44 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 421 (top tevel) : 43 marting the rules to a CSV file 44 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 42 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 42 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 42 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 42 martic(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 43 (therries) => (this performance = groceryrules.csv", sep = ",", quote = TRUE, row.nar. 43 (therries) => (therries) = 0.000269446 0.308685 1.596280 101 43 (berries) => (ther vegetables) 0.010269446 0.308685 1.596280 101 43 (berries) => (therries) => (ther weight = groceryrules) 0.011794611 0.3547401 1.388278 116	<pre>29 inspect(groceryrules[1:3])</pre>		D	ata	
<pre>#mproving model performance sorting grocery rules by lift #inspect(sort(groceryrules, by = "lift")[1:5]) #mproving model performance sorting grocery rules by lift grocer_Formal class tr_ grocer</pre>	30		0	berryr Form	nal class ru… Q
<pre>33 #Improving model performance sorting grocery rules by lift 34 inspect(sort(groceryrules, by = "lift")[1:5]) 35 36 #Finding subsets of rules containing any berry items 37 berryrules - subset(groceryrules, items %in% "berries") 38 39 40 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv",sep = ",", quote = TRUE, row.nar, 401 (top tewn) : 40 (top tewn) : 41 (top tewn) : 42 (top tewn) : 43 (top tewn) : 44 (top tewn) : 45 (top tewn) : 46 (top tewn) : 47 (top tewn) : 47 (top tewn) : 48 (top tewn) : 48 (top tewn) : 49 (top tewn) : 40 (top tewn) : 41 (top tewn) : 42 (top tewn) : 43 (top tewn) : 43 (top tewn) : 44 (top tewn) : 44 (top tewn) : 45 (top tewn) : 45 (top tewn) : 45 (top tewn) : 46 (top tewn) : 47 (top tewn) : 47 (top tewn) : 48 (top tewn)</pre>	32		0	grocer Form	nal class tr Q
34 inspect(sort(groceryrules, by = "lift")[1:5]) 35 #finding subsets of rules containing any berry items 37 berryrules <- subset(groceryrules, items %in% "berries")	33 #Improving model performance so	orting grocery rules by lift	0	grocer Form	nal class ru Q
<pre>35 #finding subsets of rules containing any berry items 37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 39 40 writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 421 (Optive(g) 2</pre>	<pre>34 inspect(sort(groceryrules, by = 25</pre>	"lift")[1:5])	1		
<pre>37 berryrules <- subset(groceryrules, items %in% "berries") 38 inspect(berryrules) 40 #writing the rules to a CSV file 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 401 (Optevel): 40 #writingData/ 40 Terminal × 401 (Optevel): 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 401 (Optevel): 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 402 Terminal × 403 (Optevel): 40</pre>	36 #finding subsets of rules conta	ining any berry items		riles Plots Pac	Rages Help Vinne
38 inspect(berryrules) 39 inspect(berryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 41 write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 401 (top level) : inspect(berryrules) 5mode Terminal × inspect(berryrules) 1 top level) : inspect(berryrules) 5mode Terminal × inspect(berryrules) 1 (top level) : inspect(berryrules) 1 top level) : inspect(berryrules) 1 (top revel) : inspect(berryrules)	37 berryrules <- subset(groceryrul	es, items %in% "berries")			- Export • 🔍
39 41 #writing the rules to a CSV file write(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nar. 40.1 (top tevel) : @ R Script : oncole Terminal <	<pre>38 inspect(berryrules)</pre>				
Imrite(groceryrules, file = "groceryrules.csv", sep = ",", quote = TRUE, row.nal, N01 (top tevel) : Image: Script : Inset Image: Script : Image: Script : Inset rhs support confidence lift count [] {berries} => (whipped/sour cream) 0.009049314 0.2721713 3.796886 89 Image: Script confidence lift count [] {berries} => (other vegetables) 0.010259446 0.308685 1.596280 101 Image: Script confidence lift count [] {berries} => (other vegetables) 0.011794611 0.3547401 1.38828 116 Image: Script confidence lift count	40 Ewriting the rules to a CSV fil	I	1		
401 (Pop Level): Image: Construction of the second se	41 write(groceryrules, file = "gro	ceryrules.csv", sep = ",", quote = TRUE	row.nar.		
Iterminal × Image //MiningData/ >> support confidence lift count 1) bs rhs support confidence lift count 1) berries} >> (whipped/sour cream) 0.009049314 0.2721713 3.796886 89 1) (berries} >> (yogurt) 0.010574479 0.3180428 2.279848 104 1) (berries} >> (other vegetables) 0.010269446 0.3086855 1.596280 101 1) (berries} >> (whole milk) 0.011794611 0.3547401 1.388328 116	40:1 (Top Level) \$	6	R Script ©		
<pre>/MiningData/ ⁽²⁾ Ths rhs support confidence lift count [) {berries} → {whipped/sour cream} 0.009049314 0.7721713 3.796886 89 [) {berries} → {yogurt} 0.010574479 0.3180428 2.279848 104 [] {berries} → {other vegetables} 0.010269446 0.3088685 1.596280 101 [] {berries} → {whole milk} 0.011794611 0.3547401 1.388328 116</pre>					
Ths support confidence lift count] berries} => {whippedyour cream} 0.009049314 0.2721713 3.796886 89] (berries] => {yogurt} 0.010574470 0.3180428 2.279848 104] (berries] => {other vegetables} 0.010259446 0.308685 1.596280 101] (berries] => {whole milk} 0.011794611 0.3847401 1.388328 116	onsole Terminal ×				
[] (berries] ⇒> (mhipped/sour cream) 0.009049314 0.7/21/13 3.790886 89 [] (berries] ⇒> (yogurt) 0.01057447 0.3180428 2.729848 104 [] (berries] ⇒> (other vegetables) 0.010269446 0.3088685 1.596280 101 [] (berries] ⇒> (mhole milk) 0.011794611 0.3547401 1.388328 116	2/MiningData/ A				
(berries) => (orbit vegetables) 0.010269446 0.3088685 1.598680 101 (berries) => (whole milk) 0.011794611 0.3547401 1.388328 116	nnsole Terminal × ://MiningData/ ↔ Ths rhs	support confidence lift count	1		
]{berries} ⇒ {whole milk} 0.011794611 0.3547401 1.388328 116	nnsole Terminal × //MiningData/ ∞ hs rhs] {berries} => {whipped/sour cream} {harries} => {wount}	support confidence lift count 0.009049314 0.2721713 3.796886 89 0.010574479 0.3180428 2.278848 104			
1	<pre>/MiningDuta/ % /MiningDuta/ % /MiningDuta/ % // Serries} => {whipped/sour cream} ? { berries} => {yogurt} } { berries} => {other vegetables}</pre>	support confidence lift count 0.009049314 0.2721713 3.79684 104 0.010574479 0.3180428 2.279848 104 0.010269446 0.3088685 1.596280 101			
	<pre>onsele Terminal >> //MiningData/ +> lhs rhs l {berries} => {whipped/sour cream} } {berries} => {other vegetables}] {berries} => {whipped/sour cream} } {berries} => {whipped/sour cream}</pre>	support confidence lift count 0.009049314 0.2721713 3.796886 89 0.010574479 0.3180428 2.279848 104 0.010269446 0.3088685 1.59280 101 0.011794611 0.3547401 1.388328 116			

Now it is showing rules with berries.

(Refer Slide Time: 30:53)

Untitled1* ×			-0	Environment History Connection
💷 🖉 🔚 🗋 Source on Save 🛛 🔍 🎢 📲		👄 Run 🛛 😁 🕞 Sou	rce • ≥	💣 🔒 🖙 Import Dataset 🔹 🥑 🗉
28 #look at the first three rules				Global Environment • Q,
<pre>29 inspect(groceryrules[1:3])</pre>				Data
30				Oberryr Formal class ru.
32				Ogrocer Formal class tr
33 #Improving model performance so inspect(sert(arccarvulas, by -	rting grocery rules by	lift		Ogrocer Formal class ru.
35 anspect(sort(groceryrules, by =	((ft;)[1;))			Files Plots Packages Help V
36 ·			D.Curlet A	a a P Zoon Export - C
oct (top cever) +			K Script +	
nsole Terminal ×				
/MiningData/ 🐵			1	
tropical fruit} => {pip fruit}	0.009456024	0.2634561 3.4826	549 9	
	The second se			
<pre>#finding subsets of rules containin berryrules <= subset(procervrules.</pre>	ig any berry items			
<pre>#finding subsets of rules containin berryrules <- subset(groceryrules, inspect(berryrules)</pre>	ng any berry items items %in% "berries")			
<pre>#finding subsets of rules containin berryrules <- subset(groceryrules, inspect(berryrules) lhs rhs</pre>	ng any berry items items %in% "berries") support confidenc	e lift count		
<pre>#finding subsets of rules containin berryrules <- subset(groceryrules, inspect(berryrules) lhs rhs] (berries) => (whipped/sour cream)] (berries) => (unut)</pre>	ng any berry items items %in% "berries") support confidenc 0.009049314 0.2721713 0.010574479 0.3180438	e lift count 3.796886 89		
<pre>#finding subsets of rules containin berryrules <- subset(groceryrules, inspect(berryrules) lhs rhs l] {berries} => {whipped/sour cream} l] {berries} => {orgourt}</pre>	ng any berry items items %in% "berries") support confidenc 0.009049314 0.2721713 0.010574479 0.3180428 0.010269446 0.3088685	e lift count 3.796886 89 2.279848 104 1.596280 101		

Berries whipped and sour cream berries and yogurt with confidence level these berries and other vegetables and berries and whole milk, these are the most frequently bought items with berries. Now I have to write the all the rules into a particular file for that I have to use right command, right the grocery rules into the file grocery rules csv. Now I am now trying to execute this. Now the file has been created. Now I have to check the file this is the rule files here inter rule files has been saved. Now, this is the procedure of how to generate association rules and save into the csv file the library is arules.

Decision Trees					
 The decision tree induction is label training data The Cars Evaluation data set #V1 buying price #V2 price of the maintenance #V3 doors #V4 persons #V5 luggage boot #V6 safety #V7 class Packeges caret rpart.plot Rpart e1071 	: learning of decision tree from class v-high, high, med, low v-high, high, med, low 2, 3, 4, 5-more 2, 4, more small, med, big low, med, high unacc, acc, good, v-good				
	Data Mining With R				

(Refer Slide Time: 32:53)

Now, I am going to show; how to generate the decision tree for creating decision tree we have to use 4 packeges; one is caret another is Rpart dot plot and for decision tree generation e1071 and Rpart e1071 is the most important packege for creating decision trees in I am using one data set; this name is car evaluation data sets, it contains 5 7 attributes price of the car maintenance value of the car doors of the car number of persons can sit into the car the booted the luggage area of the car the safety matters of the car and class; the class is either unsatisfactory, it is satisfactory good or very good these are the class now you want to make decision tree based on this information.

(Refer Slide Time: 34:05)

decisionine Notepad		100 100 200	Spin
le fåt fumat Vex Hdp			
#5("(04-) #(1127-16()) #tw0("C)/#linteguata")		1	· Side Farmat Size · Background
totall_packager("caret") totall_packager("part_biot") thrany(caret) thrany(part_biot) thrany(part_biot) thrany(part_biot)		9 - 30 - 11 - 12 - 14 - 14	Carbonia 18 - 18 -
hata Import) ar.df read.csv("car.data", sep = ',', header = TRUE) r(car.df)			
to check top 5-6 rows of the dataset, we can use head() ead(car_df)			
Data blicks is a step to split data into stato and test set the site, seed) matched is used to make our work registed. States it splits randomly and same value sill pains both will get identical matched or consultationstanticiton(y = car_offscl, p= 6.7, 15st = FALSE) matched = car_off(intering) matched = car_off(intering)	l results.	class	
(beck dimensions of train & test set im(training); dim(testing);			
summarized details(ummary(car_df)			
train() method() "Boot", "BootEl2", "cv", "repeatedcv" for training our dat number- number of resempling iterations repeats the number of complete sets of folds to compute	ta for various algorithms		
<pre>rtcl - crysic(correl(correl(s) = "repeate(o', number = 10, repeate = 1) trad_file (correl(c), data = revision_ments = repeate(c) rem_file (correl(c), data = revision_ments = repeate(c) repeate(c), data = revision_ments = repeate(c) repeate(c), data = revision_ments = repeate(c) revision_revision_ments = repeate(c) revision_revision_ments = repeate(c) revision_revision_ments = revision_revision_revision_ments = revision_revisio</pre>			
Plot Decision Trae rp(dtree_fitifinalwodel, tweak = 1.2)			
Prediction Now, our model is trained with cp = 0.01122596 estimp(1.) redic(2d)res_fit, newdata = testing(1.3)			
		· · · · · · · · · · · · · · · · · · ·	
And and and and a second	•Rpart		
Till and the second sec	1071		
19- 10a -	-610/1		
phi .			
	Data Mir	ing With R	
The second secon			1
OF I UP DELOGING		A NORS I COMMENT IS II II I	· · · · · · · · · · · · · · · · · · ·

After generating the decision tree; we are we will predict one value.

(Refer Slide Time: 34:36)

irce				50	Environment	History C	onnections
nsole Terminal ×				-0	Global Env	Import Dataset	• 🖌 🗉
/MiningData/ 🕫					Data		
<pre>whole milk} => {root vegetables}</pre>	0.007015760	0.4107143	3.768074	6	o berryr o grocer	Formal cla Formal cla	ass ru (
] {beef,					o grocer	Formal cla	ass ru
<pre>other vegetables} => {root vegetables} l {other vegetables.</pre>	0.007930859	0.4020619	3.688692	7	Files Plots	Packages	Help Vin
<pre>tropical fruit} => {pip fruit}! #finding subsets of rules containing any b</pre>	0.009456024	0.2634561	3.482649	9			
berryrules <- subset(groceryrules, items % inspect(berryrules)	in% "berries")						
<pre>hspc:(orr)(disp) hs rhs suppor lberries} => {whipped/sour cream} 0.0090 lberries} => {yogurt} 0.0105 lberries} => {other vegetables} 0.0102 lberries} => {whole milk} 0.017</pre>	t confidence 49314 0.2721713 74479 0.3180428 69446 0.3088685 94611 0.3547401	e lift 3.796886 2.279848 1.596280 1.388328	count 89 104 101				
Awriting the rules to a CSV file	34011 0.3347401	1.300320	110				

(Refer Slide Time: 34:42)



Now, I want to add the library caret Rpart Rplot and e701 e1071. Now, I have to read the data from card data, then I have to view the structure of the data set.

(Refer Slide Time: 35:26)

Minutes n Life (see Yees Fens Senson Built Debug Purtes Tools Help	10.00
🕨 🔹 😪 🖌 🔛 🎒 🥌 🍌 Go to file/function 🛛 🛛 🔂 🔹 Addins 🔹	Project: (None)
Outstedt** ■ Outstedt**	Environment History Connections
<pre>Poata Import foata Import</pre>	Global Environment Q Data Car_df 1728 obs. of 7
17 mead(car_or) 18 19 #Data slicing is a step to split data into train and test set 10 161 [(oplevel) = RScript =	Files Plots Packages Help Vinver
Console Terminal × C/MiningData/ > Tibrary(61071)	
> forary(eu/) > forary(eu/) > car_df <- read.csv("car.data", sep = ',', header = TRUE) > str(car_df) 'data.frame': 1728:obs. of 7 variables:	
<pre>\$ bypr: Factor w/ 4 levels "high","low","med",: 4 4 4 4 4 4 4 4 4 4 4 4 \$ mact: Factor w/ 4 levels "high","low","med",: 4 4 4 4 4 4 4 4 4 4 4 4 4 4 4 \$ dr : Factor w/ 4 levels "2","3", "4","5more": 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 \$ prsn: Factor w/ 3 levels "2","4","more": 1 1 1 1 1 1 1 1 1 1 2 \$ lgbt: Factor w/ 3 levels "big","med","small":</pre>	

It is showing the structure of the data set it has 4 levels bypr means buying price mact means the maintaining cost, it has 4 levels door number of doors peers number of persons here these are the structure of the data set, then I am to display, the first few data set and it is displaying the data set 6 data sets, then I have to divides the data set into 2 parts 30 percent for testing the decision tree and 70 percent for training purpose that is

why I am creating a party one partition based on class column; cl is class based on class column and 0.7 means seventy percent of data will be used for training purpose list is false.

Now, at at first, I have to make a seed on random value, then I have generated a training data set. Now, training data set has been created. Now test data set; these are the test data set. Now I have to check the dimension of the training data set. Now, it is show showing the dimension of the training data set and testing data set the training data set contains 1211 records and test data set has 5 517 records with 7 attributes.

Here is the summary of the car data sets. Now, I have to apply the training with number of 10 folds number of 10 folds number of resampling iterations, these are number and repeats means the number of complete sets of folds to compute I am here using repeated cv method for training, there are lots of method available boot method boot six 32 method cv method repeat cv method for training our data. Now execute training make one random variable, then then I have to create the tree these function training these are training function and it will create store the tree into the tree feet this is a class level data is training data I am using Rpart method is the transaction control tune length is 10; 10 is the number of resampling iterations.

(Refer Slide Time: 39:52)

• M S · Addins •	Project: (None
Untitled1* ×	Environment History Connections
🗈 🗊 🕞 🛛 Source on Save 🔍 🎢 🔹 📄 👘 😁 Source 🔹	😅 🔒 🐨 Import Dataset 🔹 🔏 🗏 List
30 dtree fit - train(c) - data - training method - "reart"	Global Environment • Q
40 parms = list(split = "information").	Data
41 trControl=trctrl,	O car df 1728 obs. of 7
42 tuneLength = 10)	intrain int [1:1211 1]
43 #print dtree_fit variable	O testing 517 obs. of 7 y
44 dtree_rit	Cesting 517 obs. of 7 v
46 #Plot Decision Tree	Files Plots Packages Help Vinne
17	an and B Zoren 2 Ferrert + O
3:1 (Top Level) © R Script ©	Compare and a compare of the
nsole Terminal ×	
/MiningData/ 🕸 🖉	
c]	
cc : 384	
00d : 69	
nacc: 1210	
trctrl <- trainControl(method = "repeatedcy", number = 10, repeats = 3)	
set.seed(3333)	
<pre>dtree_fit <- train(cl ~., data = training, method = "rpart",</pre>	
parms = list(split = "information"),	
tunel ength = 10)	

Now, you have to execute the train data to create the tree. Now, tree is created; now I have to get that tree information; these are the data of a this is the complexity parameters

this is the accuracy and kappa value and it is showing it is using cart algorithm, there are 6 predictors and 4 classes. Now I want to plot the tree plot the decision tree; these are decision tree, but it will be due to graphics problem in this machine, it will not displaying properly. Now we have to test one dataset. Suppose, these one testing data set the testing dataset showing buying price is very high maintenance cost is very high and class is unsatisfactory these my testing data whether these data is correctly predicted or not now.

(Refer Slide Time: 40:41)



I am now I am predicting the same data here new data set predict dtree the data record is this now; now it is predicting correctly because this unsatisfactory and in my dataset it is also unsatisfactory here these both are same.

(Refer Slide Time: 42:31)



Now, I want to show how to apply R program to use k means clustering the k means clustering generates cluster where the centre of each cluster denotes mean value of respective objects the input set of k means algorithm the number of clusters you want to generate and the data set and it will create k k clusters in my data set, I am using 1 wholesale customers data set in that data set, there are lots of attributes; I am I am trying to the attributes are channel regions fresh milk grocery I will generate the clusters on grocery attribute.

(Refer Slide Time: 43:59)



It is very simple program at first, we have to read the data from the wholesale customer and data is stored in x very well now and to generate k means cluster on grocery data with 4 clusters and with number of iterations will be 10. Now, I am executing the k means algorithm.

(Refer Slide Time: 45:06)



It will be the result will be stored in km objects and I want to print the [Noise] data it contains 4 clusters with these are the mean values of 4 clusters, these are the cluster vectors within cluster some squares by clusters are this some of the objects of these clusters these are the available components.

(Refer Slide Time: 45:54)



Now, we want to generate the cluster on grocery. Now it is showing the clusters on grocery; there are 4 clusters this blue, one is one cluster, red one is another cluster and green one is another cluster and very small; these are very small cluster, it scattered basically.

(Refer Slide Time: 46:17)



Suppose, I want to display inter cluster all the clusters of all attributes now it is showing all the cloud clusters of all attributes here these are clusters of channel attributes these are clusters of where is the grocery is the clusters of grocery attributes.

(Refer Slide Time: 47:27)



(Refer Slide Time: 47:31)



(Refer Slide Time: 48:03)

	Naive Bayesian Classification
	The mushrooms data set are used Attribute Information: (classes: edible=e, poisonous=p) 1. cap-shape: belib.conical=.convex.v.flat.ef, knobbed=k.sunken=s 2. cap-surface: fibrous=f.grooves.gr.saly=v.smooth=s 3. cap-color: brown=n.putfb=b_chancanec_grave_greener, pink=p.purple=u.red=e,whitewayellowary 4. bruises?: bruises.t.nowf 5. odor: almond=a.anise=i.crosote=c.fishy=v.foul=f, musty=m.none=n.pungent=p.pip(vs 6. gill=attachment: attached=a.descending=d.free=f.notched=n 7. gill-spacing: close=c.crowded=w.distant=d 8. gill=size: browd=h.putfb=b,chacolat=h.gray=g.green=r,orange=o,pink=p,purple=u.red=e, white=w.yellowary 10. stalk-shape: enlarging=e.tapering=t
000000	Data Mining With R

Now, Naive Bayesian classification Naive Bayesian classification, we classify the data set based on the probability Naive based theorem and I am using one table it contains mushroom data set its very big table it contains one class.

 Image: Direction
 Directio

(Refer Slide Time: 48:36)

Whether; it is eatable or poisonous the class are e and p whether it is eatable and poisonous these are the characteristics of mushrooms these are the characteristics values just like odor; odor means odor odor type of odor.

(Refer Slide Time: 49:17)

Dula	0.00
Transfere Verlaged	
Fir Lift Ford Yes Hep	Project: (None) •
partisonabata partisinabata (nushvuodi) trateringuita varitismata advastinguta tottisguita varitismata takatosisuka	amont History Consultant
eClassification	onment History Connections
Tterwy (wec.)	🔒 📑 Import Dataset 🔹 🍯 List 🔹
schemp	lobal Environment • Q
getConfusionMatrix function(probabilityOffests, actuals = actualMesults, threshold = 0.4)	
predictions ifelse (probabilityd/Tents > threshold, 'p', 'e') confusionmetris(textimpleidclass, predictions)	List of 9 0
plotecture <- function(predictionesults, title, color = "red")	
1 plot (ref (reministrate las, prefit (semenols, directions""), print.motoffer, sil = color, be = 1, each = tille),	440 obs. of 8 v
eclassification using waive mayerian classification	Plots Packages Help Vinner
Titray(e071.)	
nk.model <	🕑 🎤 Zoom 🖓 Export • 🔍 🥑
urin() percentusionautis() (m. Frediction(, 2))) alomboClarve()mb.Frediction(, 2), "Rec curve for Native Regestan Classification")	1
Altribute information: (Manuss attibute, attanuard) b. optimize: attant attant attant attant c. Optimize: attant attant attant for attant attant attant attant attant b. Optimize: attant attant attant attant attant attant b. optimize: attant attant attant attant attant attant attant b. optimize: attant a	
Available components:	
[1] "cluster" "centers" "totss" "withinss" "tot withinss"	
[6] "hatwaanse" "siza" "itar" "ifault"	
all for the second size in the second size is the s	
plot(xstrotery, col = kmcluster)	1.0 0e+00 0 0
> plot(xscluster, col = kmscluster)	
Error in plot.window() : need finite 'xlim' values	
	10.00

(Refer Slide Time: 49:32)

Untit	led1* X	- 5	Environment	History	Connecti	ions_ C
501	🔊 🔒 🖸 Source on Save 🛛 🔍 🎢 🗉	Run 🐤 Source • ≷	@ 8 10	Import Datase	et + 🖌	ELis
92	rhizomorphs=z,rooted=r,missin	g=?	🚯 Global Env	ironment •	Q,	
93	12. stalk-surface-above-rin	g: fibrous=f,scaly=y,silky=k,smooth=s	Data			
94	13. stalk-surface-below-ring:	ribrous=r,scaly=y,silky=k,smooth=s	O km	List of	9	0
96	nink=p.red=e.white=w.vellow=v	brown-n,burr-b,crinianon-c,gray-g,brange-b,	0 x	440 obs.	of 8	V
97	15. stalk-color-below-ring:	brown=n,buff=b,cinnamon=c,gray=g,orange=o,				
98	pink=p,red=e,white=w,yellow=y					
99	16. veil-type:	partial=p,universal=u	Files Plots	Packages	Help	Viny
100	17. veil-color:	brown=n,orange=o,white=w,yellow=y	600	700m -2	mort +	0
101	18. ring-number:	none=n,one=o,two=t		Loon I Car	Apont - 1	
102	19. ring-type:	cobwebby=c,evanescent=e,flaring=f,large=l,				
104	20. spore-print-color:	black-k brown-n buff-b chocolate-b green-r	1.	0 6	0 0	
105	orange=o.purple=u.white=w.vel	low=v	8			9
106	21. population:	abundant=a,clustered=c,numerous=n,				-
107	scattered=s, several=v, solitar	y=y				8
108	22. habitat:	grasses=g,leaves=l,meadows=m,paths=p,				8
109	urban=u,waste=w,woods=d		0.40			
110						. 0
112			0			
113	1		0.1			0
115	<u>k</u>		0			
13:1	(Top Level) 0	R Script 0	1.0	0e+00 0	0	
onsole		60				

It contains lots of attributes cap shape cap surface cap colour, these are the abbreviations of the table, these are the abbreviations of the tables and two classes eatable or poisonous to execute Naive Bayesian algorithm, we need to include two libraries one is Carret and the other one is proc library, then I have to import data same way.

(Refer Slide Time: 50:45)

21	Intitled1*	× mushro	omDf ×					_		Environment History Connections
	12	🖓 Filter					0	λ,		🞯 🔒 🖙 Import Dataset 🔹 🥑 🗏 List
*	class	cap.shape	cap.surface	cap.color	bruises	odor	gillattachment	gill.spacing	gil	Global Environment • Q,
1	р	x	5	n	t	р	f	c	ń	Data
2	e	x	5	у	t	a	f	c	t	O mushro 8124 obs. of 23
3	e	b	\$	w	t	I.	ł	c	Ł	
4	р	x	у	w	t	р	f	c	r	
-	_		<u> </u>							Files Plots Packages Help Vinne
low	ing 1 to 5 o	of 8,124 entries								💷 💷 / 🎢 Zoom 🖓 Export • 🖓
Cons	ole Ter	minal ×						-		
C:/N	AiningDat	a/ 🕸								
1	etwd("C ibrary(ibrary(Data Im ushroom	:/MiningDa (caret) (pROC) uport uDf=read.c: iew	ata") sv("mushroom	ns.csv",he	ader=T)					

Similarly, then want to view the mushroom datasets this is the same data.

(Refer Slide Time: 50:59)

(Internet)* X		- Enviro	teamor	History	Connecti	
Source on Save Q. 2.	- Dun - Source +	2 00 0		Import Datase		E Liet
	De Halt i De Joarde - I	-	Chebal Environment + 0			
8 #Data Import		- G	Global Environment • Q,			
9 mushroomDT=read.csv("mushrooms.csv",header=T)		Data				
1 View(mushroomOf)		0 mu	shro	8124 obs	. of 2	3 🔃
2 head(mushroomDf)						
3		1				
4 #Class View						
5 table(mushroomDfSclass)		Files	Plots	Packages	Help	Vina
6 str(mushroomDf)					inned a	0
7			110	2008 1 201	sport •	Ψ.
8 #Data Partitioning						
9 set.seed(4)						
0						
1 partitionData <- function(data, fractionOfDataFor	rTrainingData = 0.6)					
2 - {						
<pre>3 numberOfRows <- nrow(data)</pre>						
<pre>4 randomRows <- runif(numberOfRows)</pre>						
5 flag <- randomRows <= fractionOfDataFor	FrainingData					
6 trainingData <- data[flag,]						
7 testingData <- data[[flag,]						
<pre>8 dataSetSplit <- list(trainingData = trainingDat </pre>	ta, testingData = testingD	at:				
9 datasetsolit						
1 (Top Level) 1	R Scri	pt 0				
a (opported +						

These are the first few first 6 data sets, it is very big table that is why it is displaying this way first six in second second set of attributes these are third set of attributes these are the fourth set of attributes this because there are lots of attributes that is why it is displaying this way.

Now, I want to view the class table; there are 2 classes eatable and poisonous; there are 4000; around 4000 classes mushrooms are edible around 3000 classes are poisonous mushrooms.

(Refer Slide Time: 52:05)

• 🤏 💣 • 🔒 🔂 🚔	Go to file/function	Idins • 🔋 Project: (None)
Untitled1* ×		Environment History Connections
🗉 🔄 🔚 🗌 Source on S	ave 🔍 🎢 - 🗌	😁 Run 🐤 🕞 Source 🔹 😅 🔂 📅 Import Dataset 🔹 🍯 List
10 #Data view		* 📑 Global Environment * 🔍
12 head(mushroomDf)		Data
12 field(mushroomor)		Data
14 #Class View		🛛 🕐 mushro… 8124 obs. of 23… 🔄
15 table(mushroomDfs	class)	
16 str(mushroomDf)	crussy	
17		
18 WData Partitionin	a	Files Plots Packages Help Vinns
19 set.seed(4)		Com D R Zoom Tomat . 0
20		Contraction of the second s
21 partitionData <-	function(data, fractionOfDa	taForTrainingData = 0.6)
77 - 1		
18:1 (Top Level) =		R Script ©
onsole Terminal ×		
:/MiningData/ 🔅		
ring.type	: Factor w/ 5 levels "e"	,"f","l","n",: 5 5 5 5 1 5 5 5 .
5		
spore.print.color	: Factor w/ 9 levels "b"	, "h", "k", "n",: 3 4 4 3 4 3 3 4
3		
population	: Factor w/ 6 levels "a"	, c , n , s ,: 4 3 3 4 1 5 3 4
4	· Franker - (7 Jaural - "d"	"-" """
hadron to the state	: Factor w/ / levels "d"	, g , i , m ,: 0 Z 4 0 Z Z 4 4
habitat		
habitat		

what is the structure of the; these are these is the structure of the data sets there are twenty three variables in the data sets and around eight thousand observations, then for again have to partition the data for training set and test set.

(Refer Slide Time: 52:33)

Untit	led1*×	- D Enviro	onment	History	Connect	ions_ [
	🖉 📄 Source on Save 🔍 🎢 🔹 📄	n 🍽 🖻 Source 🔹 🥔 😭	a i reci	mport Datase	e - 🖌	ELis	
24	randomkows <- runtr(numberotkows)	1 B G	inhal Emin	• tremos	0		
25	flag <- randomRows <= fractionOfDataForTraining	ata Data	Data O mushro 8124 obs. of 23 Functions				
20	testinoData <- data[flag]	Data					
28	dataSetSplit <- list(trainingData = trainingData, test	ngData = testingData					
29	dataSetSplit	Funct					
30	}	pa	partit function (data,				
31							
32	paritionedData <- partitionData(mushroomDf)	Files	Plots	Packages	Help	Viny	
33	trainingData <- paritionedDataStrainingData		P 1	toom 🔁	xport .	0	
34	testingData <- paritionedDataStestingData						
35							
32:1	(Top Level) #	R Script ©					
onsole	Terminal ×						
/Minii	ngData/ 🔅						
nu	mberOfRows <- nrow(data)						
ra	ndomRows <- runif(numberOfRows)						
- fl	ag <- randomRows <= fractionOfDataForTrainingData						
tr	anningData <- data[flag,]						
da	taSatSalit <- list(trainingData - trainingData testingD	ta - testinoData)					
da	tasetsplit	ica - cescinguaca /					

Now training set is being generated now test set is generated now to make a schema class and odor class versus odor schema. Now testing the data set on class equal to p actual results now have to develop the confusion matrix with the probability test with the function probability test. Now confusion matrix has been generated. Now I have to make prediction curve. Now, execute the naive Bayesian algorithm on training data and it will be stored in nb dot model and this will be the predicted data.

Now, prediction on any model testing data type is raw now and to print the confusion matrix the prediction two prediction data two see for prediction data 2, this is the confusion matrix references is e to e is and with accuracy 0.9851; the class is edible here.

(Refer Slide Time: 55:33)

U Rode The first Cale Yan, Parts Series Juilt Dates Partie Task Miles		0.0
• • 🕸 💣 • 🔒 🔒 📥 🗚 Go to file/function 🔤 🔀 • Addins •	Project: ()	None) •
Utented1'* Source on Save S	myionment History Connections myonment History Connections model information model inf	
67.1 (Top Level) ÷	R Script =	
Console Terminal × C/MiningData/ =0 POSILIVE Class : e > print(getConfusionMatrix(nb.Prediction[, 3]))	-0	
<pre>Error in nb.Prediction[, 3] : subscript out of bounds > > </pre>	Show Inceback Renun with Debug	

Suppose, I want to check another data this is the first data no there is no; now I want to plot the ROC curve or Bayesian matrix prediction matrix. Now this is the curve these ROC curve for Bayesian classification is a specificity and sensitivity curve.

(Refer Slide Time: 56:32)



Therefore, for executing the Naive Bayesian algorithm at first, we have to partition the data into training set and test set then make a schema in this class with an particular attribute then the actual results based on poisonous here then generate the confusion matrix after that execute the Naive Bayesian algorithm these are procedure ok.

Thank you.