DATA WAREHOUSING AND DATA MINING LABORATORY

(Semester -VI of B.Tech)

As per the curricullam and syllabus of Bharath Institute of Higher Education & Research

(DWDM Lab Manual)



PREPARED BY

DR. M.K.VIDHYALAKSHMI

NEW EDITION







SCHOOL OF COMPUTING

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

LAB MANUAL

SUBJECT NAME: Data Warehousing and Data Mining Laboratory

SUBJECT CODE: BCS6L1

Regulation R2015 (2015-2016)

BCS6	6L1		DA	TA W	ARE	HOU	SING	AND	DAT V	A MIN	ING	L	T	Р	С
		_	Total	Conta	ct Hoi	$\frac{LAL}{1rs - 3}$	$\frac{0}{0}$		1			0	0	3	2
		-	Prerec	uisite	–Data	a ware	Hous	ing ar	d Dat	a minin	g	Ŭ	Ű		
		_	Lab M	Ianual	Desig	ned b	v - D	ept. of	Com	puter So	cience a	nd Eng	ineerin	<u>y</u> .	
OBJI	ECTI	VES					5	1	1			- C		<u>.</u>	
Data compa indica	Data mining is primarily used by the companies with a strong consumer focus. It enables these companies to determine the factors such as price, product positioning, or staff skills, and economic indicators, competition, and customer demographics														
COU	RSE	OUŤ	COMI	ES (C	Os)			1							
CO1	P	rovide	efficie	ent dis	tribut	ion of	inform	nation	and e	asy acc	ess to d	ata			
CO2	C	reate u	iser fri	endly	report	ting er	viron	ment.							
CO3	F	Find th	ind the unseen pattern in large volume of historical data that helps to manage an												
0.00	0	rganiza	anization efficiently.												
CO4	U	nderst	and th	e conc	epts c	of vari	ous da	ıta miı	ning T	echniqu	ies.				
CO5	U	nderst	and th	e conc	epts c	of Prep	proces	sing.							
CO6	E	xplain	the co	ncept	of Da	ta mir	ning.								
	MA	PPIN	G BE	TWF	EN C	COUI	RSE (OUT	COM	ES &]	PROG	RAM	OUTC	OMES	.
(3/	2/1 I	NDI	CATE	S ST	REN	GTH	OF (CORI	RELA	TION	l) 3- H i	gh, 2-	Mediu	m, 1-L	OW
COs	PO1	PO2	PO3	PO4	PO5	PO6	PO7	PO8	PO9	PO10	PO11	PO12	PSO1	PSO2	PSO3
CO1	2	2	3		3							3		2	
CO2	2		3			2						3	2	2	
CO3			3	2	3				3		3	3		2	
CO4	CO4 2 2 3 3 2					3		2							
CO5					3							2		2	
CO6					2				2		3	3		2	
Catego	ory	Profe	ession	al Co	re (PC	C)									
Appro	val	37th Meeting of Academic Council, May 2015													

LIST OF EXPERIMENTS: 1. Listing applications for mining

- 2. File format for data mining
- 3. conversion of various data files
- 4. Training the given dataset for an application
- 5. Testing the given dataset for an application
- 6. Generating accurate models
- 7. Data pre-processing data filters
- 8. Feature selection
- 9. Web mining
- 10. Text mining
- 11. Design of fact & dimension tables
- 12. Generating graphs for star schema.

DATA WAREHOUSING AND DATA MINING LABORATORY- BCS6L1

LIST OF EXPERIMENTS

	NAME OF THE EXPERIMENT						
1	Listing applications for mining						
2	File format for data mining						
3	Conversion of various data files						
4	Training the given dataset for an application						
5	Testing the given dataset for an application						
6	Generating accurate models						
7	Data pre-processing – data filters						
8	Feature selection						
9	Web mining						
10	Text mining						
11	Design of fact & dimension tables						
12	Generating graphs for star schema.						

CONTENT

	NAME OF THE EXPERIMENT	Page No.
1	Listing applications for mining	6
2	File format for data mining	8
3	conversion of various data files	11
4	Training the given dataset for an application	14
5	Testing the given dataset for an application	16
6	Generating accurate models	19
7	Data pre-processing – data filters	27
8	Feature selection	30
9	Web mining	39
10	Text mining	43
11	Design of fact & dimension tables	46
12	Generating graphs for star schema.	47

EX.NO:1

LISTING APPLICATIONS FOR MINING

AIM:

To list all the categorical (or nominal) attributes and the real-valued attributes separately.

RESOURCES: Weka mining tool1.

PROCEDURE:

Open the Weka GUI Chooser.
 Select EXPLORER present in Applications.
 Select Preprocess Tab.
 Go to OPEN file and browse the file that is already stored in the system "bank.csv".
 Clicking on any attribute in the left panel will show the basic statistics on that selected attribute.1.4

OUTPUT:

🕼 Weka Explorer							
Preprocess Classify (Cluster Associate Select	attributes Visualize					
Open file	Open URL	Open DB.		nerate	Undo	Edt	Save
Filter							
Choose None							Apply
Current relation				Selected attribute			
Relation: bank-data Instances: 600	1	Attributes: 12		Name: age Missing: 0 (0%)	Distinct:	S0 Uni	ype: Numeric que: 0 (0%)
Attributes				Statistic		Value	
				Minimum		18	
All	None	Invert	Pattern	Maximum		67	
				ChilDay		92.395	
No, Name				Stabley		19.925	
1_id							
2 000							
4 1000							
5 Jincome							
6 married							
7 children	6						
B car							
9 save_a	ct			Class: pep (Nom)			Visualize All
10 current	_act			1			
11 mortga	ge -						
12 pep	N			87		87	90
				.00	69 69	57	03
					_		
	Ren	10%8		1			
				18		42.5	67
Status							
ок							Log CXU

Result:

Thus the listing applications for the data mining was studied.

EX.NO:2

FILE FORMAT FOR DATA MINING

Aim: To study the file formats for the data mining.

Introduction:

WEKA supports a large number of file formats for the data. The complete list of file formats are given here:

- 1. arff
- 2. arff.gz
- 3. bsi
- 4. csv
- 5. dat
- 6. data
- 7. json
- 8. json.gz
- 9. libsvm
- 10. m
- 11. names
- 12. xrff
- 13. xrff.gz

The types of files that it supports are listed in the drop-down list box at the bottom of the screen.

This is shown in the screenshot given below.

00		Open	
ook In: 🗎	drsarang	Y	
pycachi anaconda AndroidSi AndroidSi AndroidSi Applicatio	e ia bar i3 ia blobo tudioProjects ia bowe tudioProjects copy ia Calibo ons ia Deski	ity r_components e Library op	 Invoke options dialog Note: Some file formats offer additional options which can be customized when invoking the options dialog
iles of <u>T</u> ype:	Arff data files (*.arff) Arff data files (*.arff) Arff data files (*.arff.gz) C4.5 data files (*.names) C4.5 data files (*.data) CSV data files (*.csv) JSON Instances files (*.jsor JSON Instances files (*.jsor libsvm data files (*.libsvm)) .gz)	

As you would notice it supports several formats including CSV and JSON.

The default file type is Arff.

Arff Format

An Arff file contains two sections - header and data.

The header describes the attribute types.

The data section contains a comma separated list of data.

As an example for Arff format, the Weather data file loaded from the WEKA sample databases is shown below:



From the screenshot, you can infer the following points -

The @relation tag defines the name of the database.

The @attribute tag defines the attributes.

The @data tag starts the list of data rows each containing the comma separated fields.

The attributes can take nominal values as in the case of outlook shown here -

@attribute outlook (sunny, overcast, rainy)

The attributes can take real values as in this case -

@attribute temperature real

You can also set a Target or a Class variable called play as shown here -

@attribute play (yes, no)

The Target assumes two nominal values yes or no.

Result:

Thus the different file formats for the data mining was studied.

EX.NO:3a CONVERSION OF TEXT FILE INTO ARFF FILE

Aim:

To convert a text file to ARFF(Attribute-Relation File Format) using Weka3.8.2 tool.

Objectives:

Most of the data that we have collected from public forum is in the text format that cannot be read by Weka tool. Since Weka (Data Mining tool) recognizes the data in ARFF format only we have to convert the text file into ARFF file.

Algorithm:

- 1. Download any data set from UCI data repository.
- 2. Open the same data file from excel. It will ask for delimiter (which produce column) in excel.
- 3. Add one row at the top of the data.
- 4. Enter header for each column.
- 5. Save file as .CSV (Comma Separated Values) format.
- 6. Open Weka tool and open the CSV file.
- 7. Save it as ARFF format.

Output:

Data Text File:

7	-146	dand 1	144.000	d Sidde Dec D	-	10 M	6											
S	5.2			1. 1. 2.	1.00	-	r, h. ha	1	1.44	1.1	- 24	1.00		151.5				
6	1.4		10.0	- B.		10.0	1960	4.12	1.	-	- 14	1		14.3	1.0			
1	-	141 184		and a brighten	1.00		9140	- 2	-		-			1.1				
		11-10-1	-	- Alleria			1			- Marc	-			1.54				
	18.1			N		-												
	4		A	Rendered	. 6	1.0	141				- 41		1.0			- 4	1.1	
		10	- 14	A2 Investme														
	185	- 00	3.4	Alterative														
	4.7	- 33	- MC	A.2 my entropy														
	48	84	15	8.0 Ho emile														
	· . X	4.4	14	KEMLERINA.														
	14	1.1	17	24/01/0108														
		2.4	18	A 2191 entities														
	1	3.0	13	T.ITRI-Selice			1											
	14	- 24	1.0	R.D.We infinite														
	4.7	110	1.1.	611044694														
	14	- 370	3.8	ADMINING.														
	- 64	- 14	18	111 investme														
	44	- 1	14	ALT RECEIPTER														
	14	- X.	11	0.0100 41004														
	14	- 14	- M.	KEING BARRY														
	58	44	1.5	SAINLASSNA														
	- 54	4.17	1.8	0.4 Househout														
	- 58	- 64	14	8.8103.800.8														
	10.	1.0	- 02	El Manthia														
	- 11	1.0	13	12101-0208														
	54	1.8	17	12thronize														
	11	8.2	131	12 Invidend														
	- 84	10.		KETTO OFFICE														
	- 64	8.8	12	AS HOLEDNA														
	144	14	18	Tildenting														

Data ARFF File:



Result:

Thus, conversion of a text file to ARFF(Attribute-Relation File Format) using Weka3.8.2 tool is implemented.

EX.NO:3b. CONVERSION OF ARFF TO TEXT FILE

Aim:

To convert ARFF (Attribute-Relation File Format) into text file.

Objectives:

Since the data in the Weka tool is in ARFF file format we have to convert the ARFF file to text format for further processing.

Algorithm:

- 1. Open any ARFF file in Weka tool.
- 2. Save the file as CSV format.
- 3. Open the CSV file in MS-EXCEL.
- 4. Remove some rows and add coreseponding header to the data.
- 5. Save it as text file with the desire delimiter.

Data ARFF File:



Data Text File:

6	14.2			-			and the second second	100		1000	Contraction of the	an ord page.
2	-	-	1	1	100	ren.	31			122		
-		19.50				9.44	-	-	-			
-	-						_					
-	-		-		1.0.00							
÷	17			A Distance in the second se				-				
	- 22	- 12	- 50	1.1 10 10 10 10								
	- 22		- 12	ALC: NO DECK								
	- 22	- 0	- 12	ALC: NO. OF THE OWNER.								
	- 22	- 52	32	121001000								
	12	- 22		A CONTRACTOR OF THE OWNER								
	- 22			and the second								
÷		- 22	- 12	TTOD INCOME.	_	-						
	1.0	- 22	- 12-	R.L. Marketing								
	- 22	-22	1.0	A 1 per antiset								
	- 10	- 52		A Data data a								
0-	- 22		- 10	TTime and and								
2	1.1	- 72	14	A Transmitted								
2	100	- 2	14	ALC: NOT COMPANY								
2	100	- 22	- 10	of Color and store								
	100		10	Adventures.								
2	14	- 14	- 52	a desta sectors.								
Ŀ	- 52	- 12	14	a kine setting								
2	- 10	14	12	of Finite and State								
2	54	- 10	1.5	Exception and the								
	14	1.6	17	Titlestow								
1	14	1.2	15	hit meaning								
i.	- 14	10	- 72	RETURN ATTING								
í.	64	11	1.7	D.S. HOLESTING								
2	14	14	10	III in allow								
24	81 mm	185						_	1.0		-	in the second se

Result: Thus conversion of ARFF (Attribute-Relation File Format) into text file is implemented.

EX. No: 4 TRAINING THE GIVEN DATASET FOR AN APPLICATION

Aim:

To apply the concept of Linear Regression for training the given dataset.

Algorithm:

- 1. Open the weka tool.
- 2. Download a dataset by using UCI.
- 3. Apply replace missing values.
- 4. Apply normalize filter.
- 5. Click the Classify Tab.
- 6. Choose the Simple Linear Regression option.
- 7. Select the training set of data.
- 8. Start the validation process.
- 9. Note the output.

LINEAR REGRESSION:

In statistics, Linear Regression is an approach for modeling a relationship between a scalar dependent variable Y and one or more explanatory variables denoted X.the case of explanatory variable is called Simple Linear Regression.

Coefficient of Linear Regression is given by: Y=ax+b

PROBLEM:

Consider the dataset below where x is the number of working experince of a college graduate and y is the corresponding salary of the graduate. Build a regression equation and predict the salary of college graduate whose experience is 10 years.

Input:

		≓ - I=	lir	iear - M	licrosoft E	xcel				×
G	ile Hoi	me Insert	Page Layo	ut For	mulas D	ata Re	view	View		df 23
Pas	ste	Calibri B I U - E - Sort Font	- 11	≡ ≡ ≣ ≡ in th Alignm	■ □ · ■ □ · ≫ ·	% Number	A Styles	Cells	∑ - 2√- 	
	F5	- 1	6.	fx						~
4	A	В	С	D	E	F		G	н	
1	×	Y								
2	3	30								
3	8	57								
4	9	64					_			
5	13	72								
6	3	36								
7	6	43								
8	11	. 59								
9	21	. 90								
10	1	20								
11	16	83								
12		C1	(et use	1000				1011	line in the second seco	~
Rea	ady	ear / Sheet2	Sheet3				100%	$\overline{\bigcirc}$		 ■ ● ● ●

Output:

9				Weka Explorer – 🗖	×			
Preprocess Classify Classifier	Cluster	Associate	Select attributes	Visualize				
Test options	cenical	C	lassifier output					
Use training set Supplied test set Set Cross-validation Folds 10 Percentage split % 66 More options			Scheme: weka.classifiers.functions.SimpleLinearRegres Relation: linear Instances: 10 Attributes: 2 X Y					
(Num) Y Start	Stop	¥	Test mode: === Classifie	evaluate on training data r model (full training set) ===				
Result list (right-dick f 23:12:32 - functions.	for option SimpleLine	s) arRegr	Linear regres 3.54 * X + 23	sion on X .21	j			
e			Time taken to	build model: 0 seconds				

a					Weka Explorer		×
Preprocess Cl	assify	Cluster	Associate	Select attributes	Visualize		
Classifier							
Choose	Simpl	eLinear	Regressio	n			
Test options			c	Classifier output			
Use trainin Supplied t Cross-vali Percentag Mor (Num) Y	ng set est set dation ge split re optio	Folds %	t 10 66	Linear regres 3.54 * X + 23 Time taken to === Evaluatio === Summary =	.21 build model: 0 second n on training set ==	onds	Ŷ
Start		Stop					
Result list (righ	t-dick f	for option	s) 1	Correlation c Mean absolute Root mean squ Relative abso Root relative Total Number	efficient error ared error lute error squared error of Instances	0.9721 4.5238 5.111 24.4264 % 23.4449 % 10	
							~

Result: Thus the concept of Linear Regression for training the given dataset is applied and implemented.

EX. No: 5 TESTING THE GIVEN DATASET FOR AN APPLICATION

Aim:

To apply the Navie Bayes Classification for testing the given dataset.

Algorithm:

- 1. Open the weka tool.
- 2. Download a dataset by using UCI.
- 3. Apply replace missing values.
- 4. Apply normalize filter.
- 5. Click the Classification Tab.
- 6. Apply Navie Bayes Classification.
- 7. Find the Classified Value.
- 8. Note the output.

Bayes' Theorem In the Classification Context:

X is a data tuple. In Bayesian term it is considered "evidence". H is some hypothesis that X belongs to a specified class C .P(H|X) is the posterior probability of H conditioned on X.

Example: predict whether a costumer will buy a computer or not " Costumers are described by two attributes: age and income " X is a 35 years-old costumer with an income of 40k " H is the hypothesis that the costumer will buy a computer " P(H|X) reflects the probability that costumer X will buy a computer given that we know the costumers' age and income.

Input Data:

C		- (~ -) :	-						_	nb.csv - Micros
E	Home	Insert	Page La	yout Fo	rmulas Data	Review	View	Team		
ľ	Cut	:	Calibri	~ 11	- A A =	= =	» =	Wrap Text		General
Pa	aste 🍼 For	py mat Painter	BI	U - 🖽 -	<u>≫ - A -</u> (≡			Merge & Cer	nter 👻	\$ - % •
	Clipboar	d 🖻		Font	G		Alignment		6	Number
	G12 \bullet f_x									
	A	В	С	D	E	F	G	Н	- I	J
1	age	income	student	credit	buys computer					
2	youth	high	no	fair	no					
з	youth	high	no	excellent	no					
4	middle	high	no	fair	yes					
5	senior	medium	no	fair	yes					
6	senior	low	yes	fair	yes					
7	senior	low	yes	excellent	no					
8	middle	low	yes	excellent	yes					
9	youth	medium	no	fair	no					
10	youth	low	yes	fair	yes					
11	senior	medium	yes	fair	yes					
12	middle	medium	no	excellent	yes					
13	middle	high	yes	fair	yes					
14	senior	medium	no	excellent	no					
15	youth	medium	yes	excellent	yes					
16										

Output data:

o l	Weka Explorer 💦 🗕 🗖 🗙
Preprocess Classify Cluster Associate S Classifier	elect attributes Visualize
Choose NaiveBayes	
Test options	Classifier output
Use training set Suppled test set Cross-validation Folds Percentage split More options	Run information Scheme: weka.classifiers.misc.InputMappedClassifier -I -trim -W weka.classifiers.bayes.l Relation: nb Instances: 14 Attributes: 5 age income
(Nom) buys computer Start Stop Result ist (right-cick for options) 17:44:02 - misc.inputNeppedClassifier	Student oredit buys computer Test mode: user supplied test set: 1 instances Classifier model (full training set)
	InputMappedClassifier: Naive Bayes Classifier
	Class Attribute no yes (0.38) (0.63)
	age youth 4.0 3.0 middle 1.0 5.0 senior 3.0 4.0 [total] 8.0 12.0 income

0		Weka	Explorer	- 🗆 🗙
Preprocess Classify Cluster Associate	Select attributes Visualize			
Classifier				
Choose NaiveBayes				
Test options	Classifier output			
O Use training set	youth 4.	0 3.0		~
	middle 1.	0 5.0		
Supplied test set Set	senior 3.	0 4.0		
○ Cross-validation Folds 10	[total] 8.	0 12.0		
O Percentage split % 66	100000			
More options	high 3.	0 3.0		
	medium 3.	0 5.0		
(New) huns computer	1ow 2.	0 4.0		
(inclusion of the second	[tota1] 8.	0 12.0		
Start Stop				
Result list (right-click for options)	student			
Resolution (influence in options)	no 5.	.0 4.0		
1734402 - miscur putwapped Lass ter	(total) 7	0 11 0		
	[cocar] /.			
	oredit			
	fair 3.	0 7.0		
	excellent 4.	0 4.0		
	[total] 7.	0 11.0		
	Attribute meppings:			
	Nodel attributes		Incoming attributes	
	(nominal) age		> 1 (nominal) are	
	(nominal) income		> 2 (nominal) income	
	(nominal) student		> 3 (nominal) student	
	(nominal) credit		> 4 (nominal) credit	~
	<			>

Q	Weka Exp	lorer			×						
reprocess Classify Cluster Associate Select attributes Visualize											
Classifier											
Choose NaiveBayes											
Test options	t options Gassifier output										
O Use training set	Time taken to build model: 0 secon	ds			^						
Supplied test set Set											
O Cross-validation Epids 10	Evaluation on test set										
O Perceptane mit % 65											
Percentage spin 78 00	Correctly Classified Instances	1	100	8							
More options	Incorrectly Classified Instances	0	0	8							
	Kappa statistic	1									
(Nom) buys computer 🗸 🗸	Root mean amared error	0.1404									
Start Size	Relative absolute error	37.4302 %									
Start Stop	Root relative squared error	37.4302 %									
Result list (right-click for options)	Coverage of cases (0.95 level)	100 %									
17:44:02 - misc.InputMappedClassifier	Mean rel. region size (0.95 level)	100 %									
	Total Number of Instances 1										
	Detailed Accuracy By Class										
	IF Rate FF Rate	Precision Recall	r-measure	ROC Area	CIASS						
	1 0	1 1	1	2	VES						
	Weighted Avg. 1 0	1 1	1	0							
	Confusion Matrix										
	a b c - algorified as										
	0.0 a = po										
	01 b = yes										

Result:

Thus the Navie Bayes Classification for testing the given dataset is implemented.

EX. No: 6 GENERATE ACCURATE MODEL

Aim:

To find the good result (by improving the performance) using the training set and testing data set for numerical values.

Objectives:

To develop training and testing data using numerical data set in order to get accurate model for classification.

ALGORITHM:

- 1. Download any data set.
- 2. Save the file with .ARFF format.
- 3. Apply 'Replace Missing Values' filter.
- 4. Normalize the values by applying normalize flter.
- 5. Go to unsupervised instance remove percentage
- 6. Right click on that (show properties) option then select 70% true and save it as training.arff
- 7. Select the original data set then right click on show properties then give 70% false and save it as testing.arff
- 8. Select classification and apply various algorithms.

TRAINING DATA:

Vener															011-14	P
ation: trail	ving aet weeks.	fitereunspe	evied.at	tsibuta Aa	iplaceHite	ing'iakas vieks. fi	ters who	pervised.at	ttikuta N	annalize G	1.8-70.0					- 7
D	c Nonnai	First Neme Norsital	City Noninal	State Nominal	Gender	Student Stetus Noninal	Major Nominal	Epartry Nonical	Age Numeric	SAT Numeric	Average score (grade) Numeric	up in in-	and seen	Z AutoSan - Ar a		
1.00	000801	DAN601	Los An.	Califor	Pemale	Graduate.	Politica	US.	1.571	8.951	87.0		1.0	a = 2 1		
0.030	posta	DWNEDD2	LOB ATL.	Arizona	Pethale:	Undergraduate	Math	15	3.047	8.607	63.0	ALL DOT N. THE CO.	and the second second	2 N		
0.060	D0E01	20E01	Elwa	Nevi Y	Mate	Graduate	Math	US .	1.381	1.905	38.0			W		
0.090		20802	Letha	Flant Y	Penale	Freduete	Econ	1.5	8.714	1.385	0.87				<u> </u>	_
0.121		20803	Defiance	Ohio	Fenale	Graduata	ECD1	LG .	1.904	1.271	0.23					
0.331	00804	30804	THE AVE	29/36	Mae	oraduate	BC04	19/86	1.552	1.01	89.0	1000 000 000				
0.351	poexs	DOBUS .	CIT4K	North	Make	Graduata	Fallisca	15	1.0	1.24	95.0	CONF.	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	LER	- AC	
0.212		229603	LIDECK	KANKAK	2-60230	Puge Bag mie	PORISO	LN	8,190	8.119	87.0	and the second second				
0.242	D0804	DANEO4	Marined	Cariote	reside	Undergraduate	Math	C3/008	8.0	1.463	91.0	0				
0.272	DOBUS	DANEDS	New Y	Dary Y	Penale	Graduate	Matt	15	1.714	1.721	71.0	and the second se	Summer Street Street	1 A A		
0.303	DEBOR	23bOb	HOEC.+	MOCES -	. 9064	Undergraduate	BC2H	LN .	4.0	0.463	82.0	and the second s				
0.333	1.00006	DAMEDE	Java	sadare	Penae	Graduate	Math	LS	1-701	1.291	79.0					
0.353	DOE07	pueur	VIITIA	pupers	Pate	Graduate	FORIDE	CUD#14	8.571	1.301	78.0	are (grade)	100	201 CO 19991		
0.393	DOE08	10608	Plister	PL658	puse	braduate	PORTO	HUSBB	1.571.0	4.100	6.0	81-	10 A			
0.424	DOB02	DANED1	Drume	PADA C.	remaie	Undergraduate	Math	US .	8,192	1/0	82.0	67,3	1 A 1	L L L		
0.454	DODOB	DANEDE	MERC	Litah	CRUMP .	Undergraduels	ECDI	15	1.0	1.497	93.0	63.0		Insert Delate Format		
0.464	D0609	1498-09	APSTE	Holand	resae	frue-baoreie	Masi	Holiane	1.09/10	1.361	75-0	78.0				
0.515	DOPID	SANCIE	Mexico	Nexco	rense	Graduate	FORTES	MENCO	1.615	8.931	\$5.0	79.0		2308		
0.540	DUCIL	DWELL	Lifecte	perer	PEGAN.	Fuge Buschere	PATI	WEINEL.	.8.0	1.3941	84.0	65.0				
0.5/5	-DOE09	10609	San Juan	Puerto.	Mae	RANDORS	P00103	18	1./1*	1.602	55.0	6.93				
0,606		DANE L2	Formate.	Chalbau	Penak	Undergreduate	CCD4	15	1.047	1.400	17.0	96.3		K L		
0,6.6	Dtheip	206.30	New T	Paper P	P104	Cridergraduate	acos	<u>10</u>	8.040	1.1-91	84.0	87/3	54	AT Average scor		
0.000	DOELS	Depet 13	11e 5	P\$0558	resac.	oraquete	PORISS .	100	8,33211	- Brettler	- 17.0	019	80	2263 57		
0.696	DOEL4	DWNE14	Delang	Crima	787548	Undergraduate	Math	Crime	1.0	1.314	79.0	71.0	30	2008 87		
0.727		33631	Macon.	. swieden	puse	ruge/graz.use	PORTICE	neveden	1.041	1.199	88.0	82.0	18	2000 45		
0.757	DOCLE	20232	EMD/or	Parent	Plac	Greduptz	ECO1	105	1.476	8,008	70.0	79.0	p P	2221 78		
0.707	DOCUS	DUE 13	LPBST	Parrie	Page 1	Undergressiene	Patri	10	3.090	1.004	20.0	79.0	10	1716 78		
0.908	DOEDS	JANE IN	1009	CREAT-	resae	Dute Dag tele	6004	08	1.099	1.0	64.0	70.0	37	1701 65		
0.010	DOBUS	1AC 21	DECTR.	- gen	200	Creductor	100103	A deutoro	8.571	1.707	. #3.0	82.0		1700 60		
0.070	DOD12	SHIPS	PATER .	College Street	Casala	Conding and and	Pelline.	10	3.041	8.200	19.0	80.0				
0.909	n poets	(APR:0)	105 Ah.	5.84104	reside	Signal and	PORIS	1.0	\$10/100	8-256-1	. 67.0	75.0	100	1577 96		
0.939	home.	promote .	Second .	Prizona -	rense.	Criste preduente	PORT'S	10	3.546	8,007	13.0	95.0	1	1842 87		
0.909	ADDELD	JUDUL NORTH	Low-	Primys F	Phese	Grad alte	Pari	1.2	3.205	1,89	35.0	92.0	18	1813 51		
-	VIDOE19	hoens	190411	PICH F	pac	for addience	E704	108	0.000	8128711	38.0	98.0	1	2049 74		
												\$7.0	110	20712 72 *		
													100 N (2			
			_	_			_		_			IN THE	-	100	-	
-	_	_	_	_	_	_	_	_	_	_	Carrie (Carrie) Carrier			(IIII) 2000 (3- 1	
	A	- Caller	100	1991.	and the second second	All and a second					A DESCRIPTION OF TAXABLE PARTY.	The second second			1000	60 B

ZeroR:

Weks Explorer	the second se	and the part lies		
Prepraces Chisity Cluster Assoc	ats falset attributes Visualize			
Clean fer				
Choose ConjunctiveRule N	1-112.0-2-1-41			
Test options	Orssifer output			
 Use training set 	Instances: JA	*		
C Suppled bettert	ACTIDUCET 12			
	Last Sene			
Criss-vieuseuri roka 10	First Same			
O Reicontago quit 10 100	City	1		
Mare options	State			
	Otodre Grudare Granus			
(Ners) City	Nancz			
	Country			
Star1 Step	las l			
Reault list (right-click for options)	257			
14-46,52-Fulles Zerolt	Average acore (graae)			
14:49:29 - rules. Decelori Table	Jest model evaluace on training oata			
14 SEID4 - rules PART	Classifier model (full training set)			
14:55:09 - rules. Rider				
14.55:32 - rules. 3Rip	ZEROR presions class value: 79.32882941176471	1		
14-55-175 - miles Date anti-article		Contract of the second s		
artistical frame car parts reveal	Time takes to build model: 0 seconds	A Alert unter		
	and Production on registers and and	WARNING		
	and Distance of Contrary and the			
	Convelation deefficient d	Advanced System Protector has detected		
	Nean absolute error 0.1263	Kin bisto or represented to chest them		
	Foot mean squared error 12.0285	in mediately.		
	Felative absolute error 100 %			
	FOR RELATIVE EQUARES EFFOR 101 4	Contract of the local division of the local		
	10:07 Newser is inscended 34	Clean Now		
Shikui		Do noishow this window Microsoft =		
OK .		- again Partier		
		252.04		
		· ••• 12 10 4/23/2013		
		11 Oct Doct Doct Doct Doct Doct Doct Doct Do		

Ridor:

Weka Saplover	and the second	-	and the second day of	and the second
Respracess Classify Cluster Associate (Select attributes visualize			
Classifier				
Chaose ConjunctiveRule N 3 -N 2	1.0 -P -1 -S t			
Test options Use torming set Septied teartast Criss-visication Ancorresga split Mani optiona Plant) City *	Classifier output 2003/01/2012 Starger COULLEY Ape SAT Average store (grade) Test mode: === Classifier model (fill temining	data deCl ====		
days in the	Ripple Down Rale Learner(Ridor) rule			
Stat				
14-48 (25 - rules 2010) 14-58 (25 - rules Devisor Table 14-58 (25 - rules Devisor Table 14-58 (25 - rules Devisor 14-58 (25 - rules Devisor 14-58 (25 - rules Devisor 14-58 (25 - rules Devisor 14-57 (25 - rules Devisor Devisor 14-57 (25 - rules Devisor Devisor)	City = Lot Angeles (34.000.0) Total number of rules (incl. the des Time takes to build model: 3 second Fuelsation on treining bet 	ferit rulej: 1 s		
	Correctly Classified Instances Incorrectly Classified Instances Mappe statistic Near mean squared error Relative should error Relative should error Total Ramer of Instances Detailed Accuracy by Class	3 01 0.0628 0.2508 84.7139 8 137.7271 8 34	6.1235 % 21.1763 %	
2				
Statul CK				lig 🛷 '
🛞 🙂 💌 💽	1 🔄 🐼 🔜	1000		• 40 10 10 10 250 PM

PART:

· Weks Explorer			manage through the little	and the second sec
Respracess Chesty Chister Associate S	Gelect attributes Visualize			
Classifier				
Chocee ConjunctiveRule N 3 -H 2	Ø ₽-1-€1			
Test options Use training set	Ossifer output Taket same = DOEDET MEMBER (2.0/1.0)	C		(a)
C Suppled test set	Last Make = DOED9: Amaterdam (2-1/1	·05		
© Percentage split % 16	Last Name - DdEld: New York (2.0/1.	01		
Mark options	Last Name = DOE11: Caracas (2.0/1.0	1		
(Nem) Oty 🔹	Last Neme = DGE13) The I (2.0/1.1)			
Start Stat	Last Name = DOE14: Deljing (2.0/1.0	1		78
Result list (right-click for optional)	: Benote (8,0/2.0)			
14-Hit 52 - rules ZeroR 14-Hit 22 - rules Decetor/Table 14-54:29 - rules OneR	Number of Rules (16			2 1
14:53:39 - rules. Rider 14:53:33 - rules. Jäp 14:53:40 - rules. J7146 14:53:533 - rules. CotjunctiveRule	Time takes to build model: 0.00 sec Evaluation on tanining set	onta		
	Correctly Classified Instances Incorrectly Classified Instances Papps statistic Near Assolute error Foot beam agained error Foot melative agained error Noot relative agained error Noot stative against	13 13 0.5093 0.0011 0.1247 45.5477 t 65.4923 t 34	55.8224 9 44,1276 9	
	Decalled accuracy by class			-
Statul OK				
🔞 🔮 💌 😭) 🖾 🧭 🔤 🔛	A STREET	ALC: NOT	* 41 13 11 259 PM

OneR:

n n ne	
a xa	*
a xz	
n xa	*
peconda	
S	
# 23 97.8585 % 0.9693 0.9693 0.003 0.045 1.0552 % 28.7565 % 38	
-	s 33 pt. 1988 % stat 1 2.9412 % 0.9693 0.003 0.043 3.0033 % 34.7943 % 34 34

JRip:



DTNB:



TEST DATA:

Tennen Construction internet i	Weka Esplorer	Property in the second s			the second division in which the	100 million (1990)	
Same Same	Prepracess Clipsify Children Associati	Select attributes Visualize	_				
Comme Compactive/Mark N 14120 Pt 141 Comme Compactive/Mark N 14120 Pt 141 Comme Compactive/Mark N 14120 Pt 141 State Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 Compactive/Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 New Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 New Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 New Mark N 14120 Pt 141 New Mark New Mark N 14120 Pt 141 New Mark N 14120 Pt 141 New Mark New Mark New Mark New	Classifier	of the state of th					
Addition Addit	Choose ConjunctiveRule N 3	M 2,0 P -1 -5 1					
<pre>9 Up transmission control of the second of the second</pre>	Test options	Ossiller output					
Chyperbolisetist Chyperbolisetist Description Neurolage setting	Use traning set	First Hans					-
Construction Construction New contrast, edit Construction New conterming Constendee <	C Surded test set 1 Set	CLTY					
Conservations of the second se		Sender					
<pre>Neuronage aff * * * * * ** ** ** ** ** ** ** ** **</pre>	C Cress-validation Moks 20	Student Status					
Here options Apr. Apr. Spil Average score (grade) Here for a list of training set (status on training set) Here for a list on training set (status on training set) Average score (grade) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Average score (grade) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Average score (grade) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Average score (grade) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Here for a list on training set (status on training set) Average score (grade) Here for a list on training set (status on training set) Here for a list on training set (status on training	🗇 Percentago quit 🛝 💷	Major					18
Aug Aug Bat Zevrege score (grado) Aug Zevrege score (grado) Medit City Bat Evrege score (grado) Aug Evrege score (grado) Medit City Bat Evrege score (grado) Aug Evrege score (grado) Medit City Bat Evrege score (grado) Aug Evrege score (grado) Medit City Bat Evrege score (grado) Aug Evrege score (grado) Medit City Bat Evrege score (grado) Aug Evrege score (grado) Evrege score (grado) Medit City Bat Evrege score (grado) Evrege score (grado) Evrege score (grado) Evrege score (grado) Medit City Bat Derde score (grado) Evrege score (grado) Evrege score (grado) Evrege score (grado) Medit City Derde score (grado) Derde score (grado) Evrege score (grado) Evrege score (grado) Evrege score (grado) Evrege score (grado) Medit City Derde score (grado) Derde score (grado) Evrege score (grad) Evrege score (grado) Evrege score (gr	Mare options	Country					
NT Lowspe score (grade) Set Set Lowspe score (grade) Set Set Lowspe score (grade) Tet Edds: wollace on training dete Tet Edds: wollace on training dete Tet Edds: wollace on training set Arcd predicts class value: fel kuiv Herd Drivelors Alge Herd Drivelors Alge Herd Drivelors Alge Herd Drivelors Class value: fel kuiv The takes to build model: d seconds UtakeDrive Herd Drivelors Alge Herd Drivelo	0.000.000	Apt					
Subject Descripte accure (gradies)	and an	SAT					
Samt Num The mode: would do not training date Image mode: would do not training date Samt in Spholad for option]	land out	Zuwrage score (grade)					
Head Star Juph and the galaxy == Classifier model (fell teaming set) == Head Star Juph and Star The	Start Start	Test mode: evaluate on training date					z.
<pre>444152 rules 2006 444152 rules Chemonitalies 445150 rules Chemonitalies 445150 rules Chemonitalies 455150 rules Ender 445150 rules Ender 445150 rules Ender 445150 rules Ender 445150 rules Ender 455150 rules Ender 505157 rules Chemonitalies 505157 rules Chemon</pre>	Data it last (sight click for options)	www Classifier wodel (fall training act) -					
<pre>NameSourceTable Add20-what.Demonstrate A</pre>	14-48-52 - niles 3eroR						
14-54-20 - vaka Const 14-5500 - vaka Fabr 14-5500 - vaka Fabr 15-5000 - vaka Fabr 15-50000 - vaka Fabr 15-5000 - vaka Fabr 15-500	14:45:29 - rules, Decelor/Table	Zerok presiots class value: Tel kviv					
<pre>M4503PMcB.FakeT M4503PMcB.FakeT M4503PMcB.FakeT M4503PMcB.FakeT M4503PMcB.FakeT M4503PMcB.FakeT M4503PMcB.FakeT M5043P.</pre>	14:54:29 - rules, OneR.						
<pre>HetDop.rules.Expr HetDop.</pre>	14:S5:04 - rules, PART	Time takes to build model: 0 seconds					
Haddel Jones Jope Haddel Jones Jope Haddel Jones Jones Jope Haddel Jones Jones Jones Haddel Johns Jones Jones Jones Haddel Johns Jones Jones Haddel Jones Jones Jones Jones Haddel Jones Jones Jon	14:55:579 - rules. Fister						
<pre>statisty - decoupled tables - decoupled active statisty classified instances 2 12.5 % Derrectly Classified instances 1 4 57.5 % Defres - decoupled instances 1 4 57.5 % Defres - decoupled instances 2 12.5 % Defres - decoupled instances 1 4 57.5 % Defres - decoupled instances 2 12.5 % Defres - decoupled instances 1 4 57.5 % Defres - decoupled instances 2 12.5 % Defres - decoupled instances 2 10 % Defres - decoupled instances 1 6 </pre>	1415123 - rules Jap	Evaluation on training set					
Line and a set of the	14-51-33 - rules Dorburch a Date	ama SUMMARY and					
300-03 -rules Node Information Constraints of Internet V Classified Internets 1 1.1.5 1 500-03 -rules Node Information Constraints of Internets 1 0.1.5 1 500-03 -rules Node Note The Solution 0 1.1.5 1 500-03 -rules Node Note The Solution 0 1.1.5 1 500-03 -rules Node Note The Solution 0 1.1.15 1 500-03 -rules Node Note The Solution 0 0.1.15 1 500-03 -rules Constrained Note The Solution 0 0.1.15 1 500-03 -rules Constraint Note The Solution 0 0 0 1 500-03 -rules Constraint Note The Solution of T	15 DH 17 - Marc Swink	Connection Classes Black Instances and	101	2262.000	1.1411		
1504:37 - nuke RAT Paper Finishing 0 1504:37 - nuke NAT Paper Finishing 0 1504:37 - nuke Nath Paper Finishing 0 1505:37 - nuke Company Paper Finishing 0 1505:37 - nuke Company Paper Finishing 0 1505:37 - nuke Company Pater TP Finishing 16 1505:37 - nuke Company Pater TP Finishing 16 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing 1505:37 - nuke Company Pater TP Finishing Pater TP Finishing	13 01:25 - rules fider	Incorrectly Classified Instances 14	9	87.5	1.0		
15 Orte - Jules OneR Needi shoolutte error 0,116 15 Orte - Jules DPb Foot make egreed strong 0,2111 15 Orte - Jules DPb Foot make egreed strong 0,2111 15 Orte - Jules DPb Foot make egreed strong 0,2111 Foot make DPb Foot make egreed strong 0,2111 Foot make DPb Foot make egreed strong 100 S Orte - Jules Despirations/se Foot make strong 100 Foot make Despirations/se Total Demker of Instantser 16	15:04:34 - rules, FART	Keens stabilition	8		1098.17		
13 0102 / UNES JPD SIGOID - UNES JPD SIGOID - UNES DECONVIDUE SIGOID - UNES DECONVIDUE SIGOI	15.04148 - rules. OneR	Mean abablute error	1.1161				
Autority wheeles becompile Storigs - whee Decompile Storigs - whee Decompile Storigs - whee Decompile Storigs - wheele Decompile Storigg	15 (01/02 - roles Jhp	Root mean equated source	.2118				
15 001/24 -view Derganchedbile Torical Hender of Instances 16 Peralled Accursor By Class T7 Bate TP Rate TP Rate Precision Recall F-Measure ROC Area Class Rates R 201 10 10 10 10 10 10 10 10 10 10 10 10 1	15 (K) 19 - niles DerisionTable	Falative absolute arror 100	10 1	1			
Total Banker of Instances 16 Decalled Acoursov Bt Class T7 Bata 75 Bata Frecision Recall F-Measure ROC Area Class . Table * * * * * * * * * * * * *	15-05:24 - rules. ContunctiveRule	Root relative squared error 100	6 1	8			
Decalled Accursor Bt Class T7 Rate TP Rate Precision Recall 7-Measure ROT Area Class . R R R R R R R R R R R R R		Total Number of Instances 10	6				
TP Jata 72 Fata 72 Fat		Parent last Assesses Br. Olans					
T7 Bate TF Bate Frequencies ROC Area Class . taka * (a) (b) (b) (b) (b) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c) (c)		and teraries woodady of crass and					
nau * * * * * * * * * * * * * * * * * * *		TP Bate TP Bate Precisi	on D	scall T-Measu	re ROC Area Class		-
* من الله الله من الله الله من الله الله من ال * 10 الله من ال	Steka						
🛞 🥹 🔛 📋 🚳 💶 🖉 🖓 🔤 👘 🖓 🔤	OK						Leg. A A
						1.46	A #1 27 10 347 PM
				1.00			4/25/2013

ZeroR:

	ion: test act-risks fits	NUMBER OF	et d'attrib	de Carlo	or Meterine	(aluan)		() () () () () () () () () ()	Y	N	12			
5	D Last Nerv	Frat Name	City	State	Gender	Student Status	Major	Country	Age	SAT	Average score (grade)		2.02	1 100
	35.000820	30803	Tel Avia	Ohio	Maie	Graduate	Ecor	1.5	37.0	1701.0	10.00		599011	- Arten
	36.0 00021	00004	Tel Avir	Nes Y	Mala	Graduate	Ecos	larant	25.0	1706.0	0.69			
1	37/0D0E22	J0605	Onex	N011	Mate	Graduate	Folia	LS .	38.0	1577.0	95.0			
	38.000823	DANED3	Liberal	Kanasa	Mela	Undergrad.ate	Politica	1.5	21.0	3842.0	87.0			11 1000
	20.0 D0624	DANED+	Mantral	Cenada I	Fensale	Undergraduate	Math	Carada	18.0	3913.0	11.0			and an inclusion of
	40.0 D0825	DANE03	NOV Y	New You	reisie	to aduate	Math	U8 .	35-0	2041.0	71.0			per tumerc
	41.0 D0E26	20206	Hot C	Masia	Male	Undergraduate	Ecos	US .	18.0	1787.0	82.0	18		and the factor of
	42.0 DOE27	DWEDE	18/2	Virgitia	Feaulte	Graduate	Math.	1.6	28.0	1513.0	19.0		Volue	
	43.0 D0E38	DOE07	Vania	Bulgaria	Mae	eraduate.	PORTICI	Bulgaria	39.0	3637.0	19.0		35	
	94,000229	20208	Mascon	Risse	Make	Graduata	Pairsa	Putte	38.0	1513.0	10.0		50	
	45.0 DOE 20	MNE07.	LELIK	PARKY		undergraduate	NAD	LN .	21.0	1228.0	82.0		42.5	
	-16.0 DOES1	DANEOR	PROC	Utah	- cwac	Ander grødvale	CC04	10	18.0	3521.0	60.0		4.761	
è	47.0D0E52	DWICLS	Anax	roland	renae	Undergraduate	MADI	roland	18.0	3494.0	15.0			
Ļ	48.0 D0E33	JANE 16	M6.000	Mexico	resae	0/201202	PORTOS	MENCO	31.0	2248.0	95.0			
-	49.0 D0854	00801	Emora	New You	PNHO	Graduate	Math	US .	25.0	2221/0	18.0			
		_	_	_	_		_	_	_	_	_	Unit] . Canal .	1	
					Ran	208								
											18		425	

Ridor:

Welca Explorer			Street State of State	and the second s
Prepracese Classify Cluster Associate	Select attributes Visualze			
Classifier				
Choose ConjunctiveRale N 2 -	12.0.0-1.61			
Test options	Classifier output			
Use tranvig set	NACOT.			
C Second has not	country			
	SAT			
C 0.69 Allegerari 1963 30	Rverage acone (grade)			
C Percentage split 16 16	Test mode: evaluate on training data			
Mare aptons	Reconstruction of the second second second second			11
	Clearing: model (rell transing are)			
(Nent) City	Ripple Diun Rule Learner (Ridor) rules			
Start Res				
Date of last Sixthe dark for and you'll	City a Tal June (15.0/1.0)			4
14-48/52 -n.ies 2woR	The second			
14-49:29 - rules. Decelor/Table	Total number of rules cincl, the default role; i	12		
14:54:29 - rules, GreR				
14/56/04 - rules, PART	Time taken to build model: 0 seconds			
1455:12 - n.lee, Fip	man FUELDEDAR ON TRAINING AFC man			
14:00140 - rules.0118	mm Schildry mm			
14:57:03 - rules, ConjunctiveRule				
The End of a result of the second sec	Correctly Classified Instances 2	12.5		
15:D4:39 -rules.FART	Incorrectly Classified Instances 14	57.5	1.1	
15:04:48 - rules OneR	Pappa Statistic 0			
15.00/02 roles 30p	Front mean equared ervor 0.2007			
15:05:19 - nies DevisionTable	Falative absolute error \$3.7235 8	6		
15:05:24 - rules, ConjunctiveRule	Fnot relative squared error 137.055 8	E.		
Contraction and a second second	Total Number of Instances 16			
	Decailed knowrany By Class			
	TF Rate FF Pate Precision Re	icall F-Nearurs	ROC Ares Class	
Status OK				10.000
		_		
	(A) (A) (A)			• 41 17 No 347 PM
				A DISTRICT OF A DISTRICT

PART:

Weks Explorer	Construction of the second	
nepraceos 019957/ Chister Associa	te Select attributes Viscolize	
Lanefor		
Chocee ConjunctiveRule N 3	#20#-1-61	
Test colons \$ Use tommy set \$ Spepled setset C tress-velideton 1985 Percentage pilt % in: Here options	Obsofter output Wenner - Yessiy and 10 c+ 0.400007 (Bostreal (2.071.3)) Student Status = Greduate AND 10 c+ 0.510031; Cinax (1.0/1.0) Gender = Sale AND 10 c+ 0.40007; Liberal (2.071.0)	
Nen) Cty	TD de 0.7322881 Moenow (2.0/1.0)	
Start Sigi	: Amstantian (2.0/1.0)	-
Result list (right-click for options)		
14:97:25 - rules. Decement Table 14:57:25 - rules. Caracter 14:55:26 - rules. FART 14:55:26 - rules. FART 14:55:26 - rules. FART 14:55:26 - rules. Detain the Salar 15:04:12 - rules. Caracter Marka 15:04:12 - rules. Caracter 15:04:13 - rules. Table 15:04:13 - rules. Decembral 15:04:13 - rules.	Time takes to build model: 0 metonds 	
tatui X	TF Bate IF Fate Frecision Recall F-Measure ROT Area Class	
		346 PM
		· • • 17 18 1/3/2013

OneR:

 Welca Explorer 				-	Strength Strength Training	(acre (C) and So
Propraceox Classify Cluster Associate	Gelect attributes Visualize					
Classifier						
Chases ConjunctiveRule N.3 -H	2.0.4-1-61					
Test options	Ossafler output					
W Use tarwig ret.	DUE23 -> LIDEFEI					-
C Rended Statust	DOE24 -> Montreal					
C etter	DEELS -> Box Foffe					
Cress-validation Polos ID	DEE27 -> Java					
Percentage oplit % 66	DOE28 -> Varna					
Mare options	DOE28 -> Moscow					
the statement of the st	D0E30 -> Drunkard Creek					
Nam) City	DGEN1 -> Mexican Hat					
	DOELD -> LANGERGAN					
Start Stop	DOESS -> Fintre					
Result list \$ ight-cick for options)	DOE35 -> Lackswana					
14-48-52 - rules ZeroR	(14/16 instances correct)					
14:45:29 -rules DecisionTable						2
14:54:29 - rules. OneR						
14/55/04 - rules Fakit	Time takes to build model: 0 seco	nda				
14.55122 - rules, Rio						
11-31-10 -rules DTNB	SUMPATU and	2.11				
14:57:03 - rules, ConjunctiveRuler	a subject of the second s					
15 DH: 17 - rules, JavaR	Correctly Classified Instances	16		100	3	
15 Of 125 TUES FROM	Incorrectly Classified Instances	a		0	3	
15 DHHE - rules OneR-	Mappe statistic	r				
15-05:02 -rules 20p	Mean absolute error	0				
15:05:09 - rules, DTNB	Fold mean squared error	4				
15.05(19 - rules DecisionTable	Fort relative smared error	a				
12 22 Contract Contract Annual	Total Number of Instances	14				
	Decailed Acoursoy By Class	#21				
		24 OR	13 23	663	32243e 2235	
	IF RADA IF RADA	Frecision	Recall	T-Nearur	e ROT Area Class	
Status						and the second se
OK						Log
		-	-			
	A20 100					* 41 TT 18 230711

JRip:

fect attributes Visualize	
£-141	
Ossifer output	
Country	
2.0%	
Source entre (prode)	
Test mode: evaluate on training data	
Classifier model (full training set)	
JRIP rules:	
and the second se	
-0 CICP-001 MULY (10.0/10.0)	÷
Number of Rules : 1	
Time takes to Build model: 0.01 seconds	
and fealuation on training ast and	
and Pumpory and	
Correctly Classified Instances 2 12.5 M	
Incorrectly Classified Instances 14 67.5 8	
Pagpa statistic 0	
POOL Beat Solated strop 0.2411	
Felative absolute error 99,5816 %	
Root relative squared eroor 99.0951 %	
Total Number of Instances 10	
num Detailled Argurany No Class and	
75 Aans 75 Fars Precision Recall F-Measure ROC Lies Class	*
	<pre>#-1-61 Desile requet Country Age Sat Average source (grade) Test mode: evolution of the initial date == Classifier model (fill training set) === AllP rules: == Country -> furg-Sel knir (16.0/14.0) Number of Rules : 1 The takes to huld model: 0.01 seconds == Fealustics on training set === == Temmary === Correctly Classified Instances == Temmary === Correctly Classified Instances == Temmary === Correctly Classified Instances == 0.1102 News absolute error 0.1102 News absolute error 94.0011 % Dest relative spaced error 94.0011 % Dest relative spaced error 10 Accorrect These TF have TF have Precision Recall F-Measure ROI laws Class </pre>

DTNB:

Weka Esplorer				and the second second	
Prepracess Classify Cluster Associate S	elect attributes Violatite				
Cleantflat					
Choose ConjunctiveRule -N 3 -H 2	0-0-1-51				
Test options	Ossiller output Country				4
C Suppled testant Set	SAT				
C CONTRACTOR NO. 1	Zverage acore (grade)				
Aecontage akt % 00	Test mode: evaluate on training d	ată			
Mare optore	Classifier model (full training :	set! ene			
(filent) City +	Decision Tables				
Start Star	Number of training instances: 16 Number of Rules : 15				
Result list (right-click for options)	Non matching orvered by Sajority clar	a :			6
14-48152 - rules ZeroR	Evaluation (for feature selection)) (CV (leave one of	(73		
14:46:29 - rules, Decision/Table	Feature set: 1.4				
14:54:29 - rules. OneR	and the second se	-			
14-00-04 - FUES PART	Time takes to build model: u.ue seco	nda			
14:55:30 - rules. Rip	FUNTURCION ON SENIAIDO deC				
14:55:40 - rules, DTHB 14:57:513 - rules, ConjunctiveRule	\$008619				
15 D4: 17 - rules Zerob	Correctly Classified Instances	15	95.75	* ·	
15.04:25 rules fider 18.04:12 rules fider	Incorrectly Cleasified Instances	1	6.25	а.	
15.0+i48 - rules OneR	Negge statistic	0.9925			
15 00 0Z -roles Mip	Hean absolute error	0.0977			
1500509 (Hales 2016)	Relation shallors around	43,2552 8			
15.05/19 - rules Decision table	Root relative squared error	54,1356 8			
15 (01:24 - rules) conjunctivenue	Jotal Number of Instances	14			
	Detailed Acouracy Br Class				
	TF Sate fF Fate S	recision Rece	11 T-Measury	ROC Ares Class	
Status CR					100.00
	- 49 - 12 E				· 41 13 14 1 13 14 10

Result :

Thus, the good result (by improving the performance) using the training set and testing data set for numerical values is found out.

EX. No: 7 DATA PRE-PROCESSING – DATA FILTERS

Aim:

To perform the data pre-processing by applying filter.

Objectives:

The data collected from public fourms have plenty of noise or missing data. Weka provides filter to replace the missing values and to remove the noisy data. So that the result will be more accurate.

Algorithm:

- 1. Download a complete data set (numeric) from UCI.
- 2. Open the data set in Weka tool.
- 3. Save the data set with missing values.
- 4. Apply replace missing value filter.
- 5. Calculate the accuracy using the formula

Accuracy= $\sqrt{\sum (\text{old-new})^2}$

Percentage of accuracy= Accuracy $\times 100$

 \sum old value

OUTPUT:

Student Details Table: Missing values

o v	lewer				
Relat	tion: weather	r -			
No.	1: outlook 2 Nominal	temperature	3: humidity Numeric	4: windy Nominal	5: play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy		96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0		TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny			FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast			TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy		91.0	TRUE	no

Student Details Table: Replace Missing values:

	📿 V	liewer					
F	Relat	tion: weathe	er-weka.filters.u	insupervise	d.attribut	e.Replac	e
	No.	1: outlook 2	2: temperature	3: humidity	4: windy	5: play	
		Nominal	Numeric	Numeric	Nominal	Nominal	
	1	sunny	85.0	85.0	FALSE	no	
	2	sunny	80.0	90.0	TRUE	no	
	3	overcast	83.0	86.0	FALSE	yes	
	4	rainy	74.8	96.0	FALSE	yes	
	5	rainy	68.0	80.0	FALSE	yes	
	6	rainy	65.0	83.0	TRUE	no	
	7	overcast	64.0	65.0	TRUE	yes	
	8	sunny	72.0	95.0	FALSE	no	
	9	sunny	74.8	83.0	FALSE	yes	
	10	rainy	75.0	80.0	FALSE	yes	
	11	sunny	75.0	70.0	TRUE	yes	
	12	overcast	74.8	83.0	TRUE	yes	
	13	overcast	81.0	75.0	FALSE	yes	
	14	rainy	74.8	91.0	TRUE	no	
		•					

CALCULATION:

Data	Old Data	Predicted data	Errors	(Error)2
Location				
J2				
J4				
J6				

Result:

Thus the data pre-processing by applying filter is performed.

EX. No: 8

FEATURE SELECTION

AIM:

To find the good results by feature selection.

OBJECTIVES:

Any classifier/model has internal feature, those feature gives more accurate and optimal result.

ALGORITHM:

- 1. Download any dataset with nominal values.
- 2. Save it as text.arff.
- 3. Split it into training and testing data set.
- 4. Go to unsupervised instance remove percentage.
- 5. Right click on that show properties then select 70% true and save it as training.arff
- 6. Right click on that show properties then select 70% false and save it as testing.arff using original data set.
- 7. Open the parameter for classifying .
- 8. Fix the set of changing values.
- 9. Look at the performance.
- 10. Go to step 3 until the expected values of maximum value is reached.

theme City State Gende mital Noninal Noninal Nonina DI Sedora Califor Penale DJ Sedora Ven 7., Penale DI Eleva Nev 7., Penale D Lacka, Nev 7., Penale	r Student Statue Rominal Graduats Undergraduate Graduatz	Major Nominal Politics Marth	Country Noninai US	Age Numeric	SAT Numeric	Average scare (grade)	he		10 4-4
 Sedona Califor. Pemale. Sedona Nevi Y., Pemale. Ellera Nevi Y., Fendle. Lacka., Nevi Y., Male. 	Graduate Undergraduate Graduate	Politica Math	US	1.171		Thurse and the second sec		500	55/0
 Sedora New Y., Penale Ellera New Y., Penale Lacka, New Y., Male 	Undergraduate Graduate	Math		8127.4111	8.952	87.0			
1 Elwa New Y., Fenale 12 Lacka., New Y., Main	Graduate		1.5	8.047	8.607	63.0			
12 Lecka. New Y., Main		Math	US	1.381	1.908	78.0			10
	Graduate	Econ	LS	1.714	1.389	78.0			17.85
Defiance Ohio Male	Graduate	Math	LS	1.904	8.373	6.23			
H Tel Avia basel Male	foraduate	B004	19 aci	\$.533	1,461	19.0			Type: Numeric
5 Drus North Male	Graduate	Palitics	LS .	1.0	1.248	96.0	100		Cinques 24(100/W
21 Liberal Kancad Fesuale	Lindergraduate	Politics	1.6	1.140	8.514	87.0		Volue	
p+ Manifed Canada Penale	undergraduate.	Mat)	cariada	-1.0	0.465	91.0		10	
 New Y., New Y., Pamaie 	Graduate	Math	1.5	1.714	1.723	71.0		12	
6 Hot C. Mosies Male	Lindergraduate.	Ecol .	1.6	3.0	3.461	82.0		15	
06 Jove Wrythia Penale	Greduetz	Math	1.5	9.952	1.381	13.0		0.302	
7 Nema Bugeta Mele	Graduate	Folitica	Bulgeria.	1.571	8.307	79.0		10.346	
8 Missow Russia Male	Graduate	Folitics	Aussia -	8.571	1.178	70.0			
07 Drunk New You Pemale	Undergraduate	Math	US	3.142	9.0	#2.0			
08 Maxic., Litalt Penale	Undergraduate	Econ	1.5	1.0	1.497	0.04			
09 Anste Holand Festale	Undergraduate	Math	Hollant	1.04T	1.361	75.0			
18 Mestro Mexico Pernale	Greduets	Politica.	Nexico	1.619	1.937	95.0			
11 Ceraces Versez Persale	Undergraduete	Math	Veneza	8.0	1.941	\$2.0			
9 San Juan Puerto Male	Graduate	Folia .	US	3.714m	1.600	95.0			
12 Remote Oregon Penale	Undergraduate	Econ	15	8.047	1.401	67.0			
D New Y., New Y., Male	Lindergraduate	Ecos	LG .	1.045	1.541	#2.0			
13 The S Mosta Penale	traduate	PORTICS	1.5	8.533	10441.4	19.0			- 50
14 Detjing China Pemale	Undergraduate	Math	Crime	5.0	8.314	79.0			
1 Stoch Sveden Mpie	Lindwgrad.ute	Folitics	Sueden	1.047	3.596	6.68			
Z Enlor Mirre Mdc	Oreducitz	Eco1	US .	1.420	1.058	95.0			
3 Interc., Penne., Male	Undergrøduete	Math	LS	1.095	1.504	85.0			
19 Loos Okah., Fenale	Indepratiate	Econ	1.5	1.095	1.0	64.0			
H Buorow Argon Mdc	Greduetz	Politica	Argentino	1.571	8.965	85.0			
5 Acres Louisters Male	Undergraduate	Econ.	1.5	8.047	1.500	79.0			
01 Los An., Califor., Fessale	Graduate	Folitics	1.6	1.571	1.95	Right click (or left+alt) for context menu			
00 Sedona Arizona Pemale	Undergraduate	Math	1.5	\$.04T.	1.687.	63.0			
	And the second s	Step.	LS:	1.386	1.905	18.0			
1 Ehma New Y., Male	Greduete	PRESS							
	1 10 MW server, proc. 2 Dimas Merth., Mais 21 Lakeral, Kancak Server, Server	10 HWD B200 HWT Create to nearest Draw B200 HWT Create the Createst Draw B401 Events Eventse Eventse Eventse Harry Carlos Eventse Eventse Harry Carlos	Norman Selection (Normal Constants) Dirace Arthr., Mais Cindualts Palitics Marry, Namy, Pamaia Cindualts Math. Dirace Arthr., Pamaia Cindualts Math. Dirace Arthr., Pamaia Cindualts Math. Dirace Arthr., Pamaia Cindualts Marry, Namy, Pamaia Cindualts Mathematics Mathematics	10 1000 2000 Vencoust Low arrors 10 Laberal Karock Peakaits Polics L.G. 11 Laberal Karock Peakaits Polics L.G. 11 Married Carobic Peakaits Polics L.G. Polics L.G. 12 Married Carobic Peakaits Polics L.G. Polics L.G. 13 Married Carobic Peakaits Polics L.G. Polics L.G. 14 Married Carobics View Policy Resolution Mart L.S. 16 Jove Mrguino Preset Resduate Mart L.S. 17 Varenta Durburborito Polics Resduate Polics Addition 18 Meticos Noted Polics Resduate Polics Addition 10 Durburborito Noted Resduate Polics Mart Noted 11 Durburborito Noted Resduate	North State Park Park States and a room room room room room room room ro	1 10 </td <td>1 10 HW 10</td> <td>Dime Dime <thdim< th=""> <thdim< th=""> Dime Dim</thdim<></thdim<></td> <td>1 10 mm 2000 2</td>	1 10 HW 10	Dime Dime <thdim< th=""> <thdim< th=""> Dime Dim</thdim<></thdim<>	1 10 mm 2000 2

Training Data:

JRip(seed=1):

Woka Explorer								A DOLLAR STOLEN
vepraceos 02001/ Chuster Associate	Select attributes Visualize							
Jamefer								
Chases 384p # 3 11 2.0 02 61								
Test options	Classifier output							
se taiwig set								
O Supplied text and Sel	JRIP pulses							
Cress-validation tokis 37								
🕐 Percentage epit 🐁 🔟	(First Same = JOE02) =>	City=Leck	emena 12.0/0	. 25				
Mana optiona	(First Same = JOEO1) =>	CITY-ELMI	26 (2.0/0.0)					
	=> £3£3=960088 (3110/2 =>	1-01						
(Nam) City +	Sumber of Rules : 1							
dark (inc.								
Dam & East Study and Bar antonnal	Time takes to build mod	el: 0.04 #	econde					
-164 12-104s 340								
	Evaluation on train	ing set as	7.5					
	Correctly Classified In	stences	2		20.5682	8		
	Incorpectly Classified	Instances	27	29	79.4118	1		
	Nean absolute error		0.05	83				
	Foot mean squared error		0.17	07				
	Relative absolute error		87, 18	88 9 07 8				
	Jotal Sunkey of Instano	1201	24	71 ÷				
	- Istailed Accuracy 8	Class						
	TF Sace	FF Fate	Precision	Recall	F-Measure	ROC Area	Clapp	
	0	٥	٥	0	0	19-2.0	los Angeles	
	1	0.571	0.1	1	0.182	0.565	Sedona	
			1	1			Limira	
	đ	0	õ	0	0	0.561	Defiance	
	à	à	ŵ	0	4	0.641	Tel Ariv	
	5.0	- A	Ő.	A	Ó.	0.541	off male	

JRip(seed=2):

Woka Taplatar			And and the second s	and a second sec
repracedze Closefv Cluster Associate I	Gelect attributes Visualme			
Interfor				
Chapes Ridler # 3 -5 1-N.2.0				
Test options Use training set Supplied test set Criss-validation risks D	Obsilier output Poudznio Statius Major Country Age			2
Nercontage split % 00	Int Intropy score (grafs) Test moderevaluate on training data			
Non) City 🔹	- Classifier model (full training	44E)		
Start] Sun]	JRIP rules:			
Salashi ya Kata Sala Salashi ya Kata Sala Salashi ya Kata Salashi 20 Shi Si - Naka Raja 20 Shi Si - Naka Raja 21 Shi Si - Naka Raja	(First Name - JOBO1) -5 City-Lacisson -5 City-Sedone (32.0/28.0) Rumber of Roler + 2 Time takes to build model: 0.00 seco Evaluation on training set Funnaly Correctly Classified Trainings Neepe restirtle	na (3.6/0.8) nda 23 3.0963	14.7059 % 85.2041 %	
1886	Heam shoolyde troot Foot mean squared stroot Felarite shoolrte stroot Foot relative squared stroot Total Hamber of Instances Decalled Accultacy By Class	0.0022 0.1764 98.7591 k 96.5997 k 34		

JRip(seed=3):



Ridor(seed=1):

Weka Taplatar		
Represented Construction Associate	Select attributes / Visualize	
Classifier		
Chapes Rider # 1 -5 1-N 2.0		
Test options	Ossiller output	
Use training set	Student Status	
C Suppled test and Sel.	Nagar	
C research the star In	Country	
	Los	
D vercentage det	Shi Zwytane Alone (IIIAde)	
Mare options	Test mode evaluate on training data	
(Ham) City ·	Classifier model (full training set)	
Stert Stop	RUNNE DOWN MUTE TARGET RUNNEL FULES	
Data & last births and for anti-and		
38.54(52 -n arc Rin		
30:59:15 - rules. Rip	City = Setons (34.0/0.1)	
33:58:34 - rules. JRip		10
20.55153 - rules Rider	Tougt underer of lates Vinor, one denote intelle 1	
2101125 - rules Fictor	Time takes to holls model: 0 seconds	
210131-rules Mp		
23 03:22 -rules Paler	Evaluation on training set	
01(03:28 -rules Fider	see Sumary.eve	
	Connective Classified Toutagnes 8 & 2005 6	
	Incorrectly Classified Instances 31 91.1761 V	
	Kapps statistic 0	
	Near sharlute error 0.0428	
	Rot men equared error 0.2008	
	Helative should be error 94-7139 b	
	Total Manher of Instances 34	
	Denalled Accuracy By Class	
	TF RACE TF FAILS Predictor Recall F-Measure Not Lives Class	-
DE	r r	Log and x0
20.P		

Ridor(seed=2):

Woka Taplater	-	_					-		-		
Anepracess Classify	Chieter Associate	Select attributes Visual	ize.								
Casefier											
Chocen Ridor	£1.51-N2D										
Test options		Classifier output									
🔹 Use training set		5	AT								
C Suppled test set	Set.	3	werege at	ore igned	e0						
Cress-validation	Polos 20	Test moderevals	mie on to	eining da	ba						
🗇 Recontage split	16. 06	Classifler	model (fu	11 traini)	ng art inte						
Mane opt	tonu	Ripple DOwn Ral	e Learner	(Ridor) r	ular						1
(Hare) Oby	*	Pitte - Salora	WEAR A	5							
Start	Skap			0							-
Result list (right-click fi	or options)	Total number of	t rules (d	ncl. the	default rule	1: 1					
30:54:52 - rules, JRID	10.00	THE DATE OF D	other layers	NO ROBERT	DBACK						
20:58:15 - rules, Rip 20:58:34 - rules, Rip											
30 (59) 53 - rules. Rider		- Ivaluation	Dvaluation on training set								
21:00:03 - rules Rolar		Sumary									
23101131 - Miles 200		Correctly Class	ufied Ins	baisons	2		0.1235	N			
21:03:21 - rules, Richr		Incorrectly Cla	estfied 1	nétances	31,		91,1765	1			
1101138 - niles Rider	100	Rappa statistic			4						
		Near absolute a	eroc		0.06	29					
		Enlative shorts	te arror		9.40	75.8					
		Root relative s	uuared er	ror	137.72	71.4					
		Total Runker of	Instance	Ð	34						
		- Decailed Ac	cursey By	Class ***							
			IF RACE	IT Fale	Frecision	Recall	r-measure	NOC Ales	Clapp		
			0	0	0	0	0	0.6	Los Angeles		
			1	1	0.026	1	0.162	0.5	Sediona		
			0		0			0.5	Lationana		
			a	0	0	D	0	0.0	Defiance		-
STARK.											11201002 000
DK											Log Alla x0

Test Data:

	BUT BUT YOUNA TH	eutronenbeuro	and a tribe	za Repis	ceMating	Alues	a numerabou	veed.atre	subb hitems	alle St. D	-T0.8				
1.16	D Last Nerv	 Prot Name Reminal 	City	State	Gender	Student Status	Major	Country	Age	SAT	Average score (grade)		60	10	law.
	0.000020	30POS	Tel Auto	Ohin	Male	Crackate	Free.	1.6	1.904	1.508	15.0		600011		
0.0	66D0021	20204	Tel Avia	Nex Y.	Hele	Graduate	Ecot	Intel	1.333	8,492	0.63				
14.5	133D0E22	20605	Cinex	North	Fenale	Graduate	Folitics .	US	1.0	1.261.	95.0				- Acco
	0.200673	DAMESO -	Liberal	Karmet	Penale	Graduate	Politics	1.5	1.141	1.553	17.0				
D.2	6600624	DAMED+	Mantheal	Canada	Fenale	Undergraduate	Econ .	Cartada	8,0	1.521	81.0			400 March 19	
D.3	133 DOE29	DANEOS	NOV You	New Y	renaic	braduate	Math	U8	1.7146	9.772.	31.0			Typel Muneric	
1.2	0.4D0826	00506	Hot C	Maxin.	.Maie	Undergraduate	Econ	1.5	1.0	1.493	82.0			Caldher to Fronte	£3
D.4	66D0527	DWHEDE	Lava	Virginia	Feixale	Graduate	Nati	LS	1.952	8.992	79.0				
0.5	133 DOE26	30807	Nama	Bulgaria	Male	biaduate	Politics	Bulgaria	8.371	1.321	79.0				
	0.6 D0E29	DOBOB	Mascow	Runnier	Male	Graduate	Palitica	Rumia	1.571	1.391	. 10.0				
0.4	eepasso	DONEO7	Dnak	Nevi Y	Festale	Lindergraduate	Math	LG.	1.140	1.0	#2.0				
0.7	53. poest.	DANE08	MORELLE	Utah	Penale	Undergraduole	EL01	2.6	1.0	8.538.0	0.08		17		
-	0.8D0E32	DANEOS	Ares	holand	renae	Undergraduate	Rati	rolard	1.047	1.171	. 15.0				
0.8	66D0E33	DANE 18	346,000	Mexico	resae	Graduate	F081CI	MENCO:	1,619	- 3.0	95.0				
- 9,9	33. D0E34	20801	Elmiro	Peper Y	Met	Graduate	Math	05	3,386,	1.976	18.0				
												Units DE Decot	F		+ Rousia

JRip(seed=1):



JRip(seed=2):



JRip(seed=3):

Wolca Taplosar		_								يكانيكا الكارك المت	
Prepracess Classify	Cluster Associate	Select attributes Visio	doe								
Classfer											
Chaose Ridor	€3-\$1-N20										
Test options		Classifier output									
 Use taring set 			Age .								*
Supplied test set	Set		SAT								
Criss-validation	HING DIE	Test mode upser	Average d	test set	etas paknow	m Dead	or increases	with:			
Percentage galit	14 GC	1000 1000 10000	and the second		VARY GRADUE	in Income	ing annu sites	(daagt			
- American		Classifier	model (f	all traini	ng set)						
1	0.0718	1									100
Rivel Oto		Thir rules									
for the state											
Start	(im)	-> dicydal k	wiy (15.0	/14.03							
Repuit fat (right-click f	br optonaj	Traduce of Sola	20.								
21:06:55 - rules, JRio		Manper or Mule									F
J 1075JS - PUMA JAp	1	1000 (1000 SV) (1000 SV)									
2107120 - rules Rider	8	Tips takes to	poild mod	al: 0.01 a	econda						
21:07:27 - rules. Rider		- Family atting									
		Instary	-								
		Correctly Clas	aified In	stances	್ಷಕ್ಕೆ		6.35	8 S			
		Incorrectly CL	sputtied	Instances	15		95.75				
		Neen absolute	ertor		0.11	172					
		Foot mean squa	ned etror		0.24	191					
		Relative absol	ste error		100						
		Total Monhat a	f Instant	ar .	100.01	1.9 8					
		Decalled A	coursey a	5 Class ==	-						
			TI Jaca	TF Fate	Frecision	Recall	T-Measure	ROC Area	Class		
			a	a	D	p	0	0.5	Defiance		
			1	1	D.063	1	0.118	0.5	Tel Aviv		
-			. a	. 9.	D	0	0	g.8	CLIEROR		-
status											Lim

Ridor(seed=1):

Woka Taplatat	-					-		and the second se
eprecision Classify Cluster Associate	e Gelect attributes Visualze							
auther								
Chapes Ridker # 3 -5 1-H 2.0								
est options	Classifier output							
🗇 use training set	Average at	ore igrad	c).					
Suppled bet set] lest modernser supplied	cest set:	size unknow	n treatin	og inorement	47774		
Cress-validation rokis (0	Classifier model (fr	11 vraini	ng sec) week					
Percentage epit 🐀 🕫								
Have options	Ripple Down Rale Learner	(Rider) r	ules					
iani) City	 City - Tal Aviv (16.0/) 	1.01						
Quel 1	Total number of rules (s	incl. the	default rule	1: 1				
station definition								
106:55-rules. Rio	line takes to build mode	ili 0 secs	nde					
107503 -rules. Pip	- Evaluation on test a							
107; 10 - rules. Rip	Jumary							
1:07:27 - rules Rider	Constants Constants on the	11221	- C2		1.40	25		
	Incorrectly Classified 1	Datances	15		42.15			
	Kapps statistic	any paracest.	- C		1000	3.0		
	Nean absolute error		0.11	72				
	Foot mean squared error		0.34	23				
	Relative absolute error		102					
	Poot relative squared er	TOT	141.27	43.9				
	spins winger at instance		1.0					
	Detailed Attrivety Sy	Class	•					
	IF Ance	FF Fate	Precision	Recall	F-Measure	ROC Area	Class	
	â	4	0	0	ů.	0.6	Cefiance	
	L	4	0.065	- L	0.115	0.5	Tel Aviv	
	a	a	D	0	0	0.5	Ciman	
	a	4	D	0	0	0.5	Liberal	
	a	2	0	0	0.	0.5	MinCreal	
	. 0	0.0	0	<u>0</u> 5	30	0.6	Bew York	
88.6								
Ł								Lop

Ridor(seed=2):

Ownfer								
Ghapes Ridler # 1 -5 1 -N 2.0								
Test options	Classifier output							
Supplied teatteat Supplied teatteat Crists-validation No. No.	Rverspr s Test modernser supplied Classifier model (f Fipple DVm Fals Learns	oore (grad teat set) all traini r(Rider) :	ng setj	n (peadar	g inorement	elly		
Mare options								
(len) Cty .	City - Tal Aviv (16.0/	6.0)						
Start Itus	Total number of rules (incl. the	default rule	1: 1				
Result list (right-click for options)	Time takes to boild not	-	7.64					
21.06:55 - rules JRip	and once of forms and	eri a peca	1000					
11:07:03 -rules. Rip	Evaluation on test	ant end						1
110720-04-500	Sonmary							
21:07:17 - rulen: Richr	Correctly Classifies In Incorrectly Classified Papps statistic Near Autolute stor Foct mean agained error Pelative Robulte throw Foct relative squared e Total Rankes of Instanc	stances Instances Froz	i 15 0,11 0,34 100 141.27 14	72 23 43 4	6,25 93,75			
	Detailed Accuracy 5	y Class	-					
	IP Bace	78 Fate	Precision 0 0.045	Recall 0	F-Measure 0 0.115	800 Ares 0.5	Class Definice Tel Ariv	
	a	a :	D	0	0	0.5	Climate	
	a	a	D	D	0	0.8	liberal	
	0	0	0	0	0	0.6	Man Week	
	M.	20	0.6	0.00	10,22	4.15	Carrier Society	

Training Data Set Performance:

	TRAINING SET				
CLASSIFIER	PARAMETER SETTING	PERFORMANCE			
JRip	Seed=1	Root Mean Squared Error=0.1707 Mean Absolute Error=0.0583			
JRip	Seed =2	Root Mean Squared Error=0.1764 Mean Absolute Error=0.0622			
JRip	Seed =3	Root Mean Squared Error=0.1764 Mean Absolute Error=0.0622			
Ridor	Seed =1	Root Mean Squared Error=0.2508 Mean Absolute Error=0.0629			
Ridor	Seed=2	Root Mean Squared Error=0.2508 Mean Absolute Error=0.0629			

Testing Data set Performance:

	TEST SET				
CLASSIFIER	PARAMETER SETTING	PERFORMANCE			
JRip	Seed=1	Root Mean Squared Error=0.2431			
		Mean Absolute Error=0.1172			
JRip	Seed =2	Root Mean Squared Error=0.2431			
		Mean Absolute Error=0.1172			
JRip	Seed =3	Root Mean Squared Error=0.2431			
		Mean Absolute Error=0.1172			
Ridor	Seed =1	Root Mean Squared Error=0.3423			
		Mean Absolute Error=0.1172			
Ridor	Seed=2	Root Mean Squared Error=0.3423			
		Mean Absolute Error=0.1172			

Comparison between training and testing data set:

	TRAINING				
JRip	Seed=1	Root Mean Squared Error=0.1707			
		Mean Absolute Error=0.0583			
Ridor	Seed =1	Root Mean Squared Error=0.2508			
		Mean Absolute Error=0.0629			

TEST					
JRip	Seed=1	Root Mean Squared Error=0.2431			
		Mean Absolute Error=0.1172			
Rider	Seed =1	Root Mean Squared Error=0.3423			
		Mean Absolute Error=0.1172			

Result:

Thus the good results by feature selection were found.

EX. No: 9

Web Mining

Aim:

To apply the web mining technique clustering algorithm for the given dataset.

Introduction to Web Mining:

Web mining is an application of data mining techniques to find information patterns from the web data. Web mining helps to improve the power of web search engine by identifying the web pages and classifying the web documents.Web mining is very useful to e-commerce websites and e-services.

Web Content Mining :

Web content mining can be used for mining of useful data, information and knowledge from web page content. Web structure mining helps to find useful knowledge or information pattern from the structure of hyperlinks. Due to heterogeneity and absence of structure in web data, automated discovery of new knowledge pattern can be challenging to some extent. Web content mining performs scanning and mining of the text, images and groups of web pages according to the content of the input (query), by displaying the list in search engines. For example: If an user wants to search for a particular book, then search engine provides the list of suggestions.

ALGORITHM:

- 1. Open the weka tool.
- 2. Download a dataset by using UCI.
- 3. Apply replace missing values.
- 4. Apply normalize filter.
- 5. Click the cluster tab.
- 6. Apply all algorithms one by one.
- 7. Find the no of clusters that are formed
- 8. Note the output.

Output:

Cobweb

Justerer		
Choose Cobweb -4 1.0 < 0.00282094791773878	15 4 C	
Guster mode	Clusteer output	
@ Use baning set		
C Supplied test set	Time taken to build model (full training data) : 0.01 seconds	
Percentage split % (66)	Model and evaluation do training set	
Casses to dusters evaluation		
(Nori) contact-lenses -	Canada and and and and and and and and an	
Store clusters for visualization	2 1 440	
	4 1 (48)	
Ignore attributes	b 1 (69) T 1 (69)	
Start Down	E 1 (48)	
and the state of the second	D 1 (48)	
and in the second second	12 1 (40) 13 1 2 40	
	2 2 1 4 17 1 6 4 23 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 25 1 6 4 26 1 6 4 27 1 6 4 28 1 6 4 29 1 6 4 20 1 6 4 21 1 6 4 22 1 6 4 23 2 6 4 24 4 6 4 25 1 6 4	
ons K		in

EM

Cluster mode	Laft-click to edit propertier for this object, right-click/Wt-Shift-left-click for menu	
Use taking set Sachuer test at Sachuer test at Concerninge split Concerninge split Concerninge split Concerninge split Concerninge spliter evaluation Concerninge spliter dualities Conce	<pre>member information ==== Scheme: weiks.clusterers.Bs -7 100 -8 -1 -1 10 -max -1 -11-ov 1.85-6 -11-iter 1.05-6 -H 1.05-6 -man-slote 1 -9 100 Relation: unstanct-innere Instances: 34 Attributes: 5 ap mystocale-prescrip estigmating insta-port-data contort-inners Test mode: evaluate in training set) === Di == Mumber of clusters melected by cross validation: 2 Rumber of clusters melected by cross validation: 2 Rumber of ilerations performed: 10 Cluster Attribute</pre>	
	pre-prestycpin 5,8346 4,1654 prestycpin 6,4085 3,3917	

Farthest First

Clusterer		
Choose ParthestFirst 4/2-51		
Cluster node	Clusterer output	
Lise training set Supplied test set Procentage split Dates to clusters evaluation Table(contact immes Stare clusters for visualization	Memory and Antiparties and Ant	*
Ignore attributes	tes-prod-pss	
Start	Test moder evaluate on training date	
Result list (right-click for aptions)		
00.1354-EM 00.12465-Colveb 00.12122-EM 02.1122-EM 02.1122-EmbedSel	<pre> Clustering model (full training bet) FarthestFirst</pre>	
	time texes to offic model (full fightund date) (n accounts	
Data	Model and evaluation on training set	and and a second
OK .		Log An X

Filtered Cluster:



Hierarchical Cluster

Waka Explorer			and the latter see
Preprocess Classify Custer Associate Select attribute	es. Visualize		and the second
Ousterer			
Choose SimpletCHeans -N.2 -A "weka core.Budde	ear/bitance -R. fret-last" -1 500 -rum-slots 1 -5 30		
Ouster mode	Clusterer sulput		
 Use training set 	Discrete Estisator. Counts = 7 5 (Total = 12)		*
C Supplied text set	Attribute: teas-prod-rate		
a hereiten off	Attribute: contact-larger		
Conner to chatters and alter	Discrete Estimator. Counts = 2 3 8 (Total = 13)		
Over a start branch			
Without distance for all advectors	Claster: 1 Fride probability: 0.3529		
W serve creaters an womantagou	Attribute: spe		
Torroro attributes	-) Discrete Estimator. Counts = 2 1 5 (Total = 8)		
	Attribute: spectacle-prescrip		
Start ling	Discrete Estimator. Counts = 2 5 (Total = 7)		
Result list (right-click for cottons)	Discrete Estimator, Counts = 2 5 (Total = 7)		
00:13:06 -EM	Attribute: tear-prod-rate		
00:14:05 - Cobweb	Discrete Estimator. Counts = 4 8 (Total = 7)		
00-18-52 - Farthestival	Discrete Estimator, Counts = 1.2.5 (Total = 5)		
00:21:29 - ManadQusterer	an company service of the service of		
31:25-20 - HerardskalClusterer	S 3 37 37 37 5 3		
01:25:28 - Singlet/Heats	Time taken to build model (percentage split) t 0 m	e con da	
	Clustered Instances		
	0 6 (678)		
	1 1 (338)		
			1
	log likelihood/ -4.55298		
	CERTER REPORTED AND A REPORT OF A		
	1-12.181	and the second se	
Stetus OK			100
😰 🗀 🖪 🙆 🙆			- 128 AM

Simple KMeans:



Result:

Thus the web mining technique clustering algorithm for the given dataset is implemented.

EX. No: 10

TEXT MINING

Aim:

To find association between data and to find the frequent item set for text mining.

Text Data Mining

Text data mining can be described as the process of extracting essential data from standard language text. All the data that we generate via text messages, documents, emails, files are written in common language text. Text mining is primarily used to draw useful insights or patterns from such data. The purchasing of one product when another product is purchased represents an association rule. Association rules are frequently used by retail store to assist in marketing, advertising, floor placement, and inventory control. Association rules are used to show the relationship between data items.

Keyword-based Association Analysis in text mining:

It collects sets of keywords or terms that often happen together and afterward discover the association relationship among them. First, it preprocesses the text data by parsing, stemming, removing stop words, etc. Once it pre-processed the data, then it induces association mining algorithms. Here, human effort is not required, so the number of unwanted results and the execution time is reduced.

ALGORITHM:

- 1. Open dataset
- 2. Select associate
- 3. Choose different algorithm for association
- 4. Observe the performance
- 5. Select the association rule with the maximum confidence rule.

INPUT: SuperMarket data set

<u>د</u> ک	liewer				×	
Relati	on: supermarket					
No.	1: department1 Nominal	2: department2 Nominal	3: department3 Nominal	4: department4 Nominal	5: departme Nominal	
1						Save
2	t					
3						10
4	t					Apply
5						
6			t			
7	t					ype: N
8						que: 0
9	t		t			Weight
10						1047.0
11						
12	t					
13	t	t				
14						
15						visualize All
16	t				t	
17						
18	t		t			
19	t					
20	t				-	
21		t			t	
22	t	t				
23					-	
< ⊨					•	

OUTPUT: Apriori Algorithm

🕝 Weka Explorer	
Preprocess Classify Clust	er Associate Select attributes Visualize
Associator	
Choose Apriori -N 1	0 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start Stop	ssociator output
Result list (right-click for c 13:43:15 - Apriori	Best rules found:
	<pre>1. biscuits=t frozen foods=t fruit=t total=high 788 2. baking needs=t biscuits=t fruit=t total=high 760 3. baking needs=t frozen foods=t fruit=t total=high 4. biscuits=t fruit=t vegetables=t total=high 815 == 5. party snack foods=t fruit=t total=high 854 ==> br 6. biscuits=t frozen foods=t vegetables=t total=high 7. baking needs=t biscuits=t vegetables=t total=high 8. biscuits=t fruit=t total=high 954 ==> bread and c frozen foods=t fruit=t vegetables=t total=high 83 </pre>
Status OK	Log 💉 0
_	

FP-Growth Algorithm:

🥥 Weka Explorer
Preprocess Classify Cluster Associate Select attributes Visualize Associator
Choose FPGrowth -P 2 -I -1 -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1
Start Stop Associator output
Result list (right-dick for (13:43:15 - Apriori 13:49:12 - FPGrowth 1 [fruit=t, frozen foods=t, biscuits=t, total=high] 2 [fruit=t, baking needs=t, biscuits=t, total=high] 3 [fruit=t, vegetables=t, frozen foods=t, total=high] 4 [fruit=t, party snack foods=t, total=high]: 5 [fruit=t, party snack foods=t, biscuits=t, total= 6 [vegetables=t, frozen foods=t, biscuits=t, total= 7 [vegetables=t, baking needs=t, biscuits=t, total= 8 [fruit=t, vegetables=t, total=high]: 954 ==> [bread 9 [fruit=t, vegetables=t, frozen foods=t, total=hig]: -
Status OK Log x 0

Result:

Thus association between data and to find the frequent item set for text mining was found.

EX. No: 11

DESIGN OF FACT AND DIMENSION TABLES

Aim:

To design fact and dimension tables.

Fact Table :

A fact table is used in the dimensional model in data warehouse design. A fact table is found at the center of a star schema or snowflake schema surrounded by dimension tables. A fact table consists of facts of a particular business process e.g., sales revenue by month by product. Facts are also known as measurements or metrics. A fact table record captures a measurement or a metric.

Designing fact table steps

Here is overview of four steps to designing a fact table:

- 1. **Choosing business process to model** The first step is to decide what business process to model by gathering and understanding business needs and available data
- 2. Declare the grain by declaring a grain means describing exactly what a fact table record represents
- 3. Choose the dimensions once grain of fact table is stated clearly, it is time to determine dimensions for the fact table.
- 4. **Identify facts** identify carefully which facts will appear in the fact table.

Fact table FACT_SALES that has a grain which gives us a number of units sold by date, by store and by product.

All other tables such as DIM_DATE, DIM_STORE and DIM_PRODUCT are dimensions tables. This schema is known as the star schema.



Result: Thus design fact and dimension tables are created.

EX. No: 12

GENERATING GRAPHS FOR STAR SCHEMA

Aim:

To generate graphs for star schema.

Introduction:

Star schema is the fundamental schema among the data mart schema and it is simplest. This schema is widely used to develop or build a data warehouse and dimensional data marts. It includes one or more fact tables indexing any number of dimensional tables. The star schema is a necessary case of the snowflake schema. It is also efficient for handling basic queries. It is said to be star as its physical model resembles to the star shape having a fact table at its center and the dimension tables at its peripheral representing the star's points.



In the above demonstration, SALES is a fact table having attributes i.e. (Product ID, Order ID, Customer ID, Employer ID, Total, Quantity, Discount) which references to the dimension tables. **Employee dimension table** contains the attributes: Emp ID, Emp Name, Title, Department and Region. Product dimension table contains the attributes: Product ID, Product Name, Product Category, Unit Price. Customer dimension table contains the attributes: Customer ID, Customer Name, Address, City, Zip. Time dimension table contains the attributes: Order ID, Order Date, Year, Quarter, Month.

In Star Schema, Business process data, that holds the quantitative data about a business is distributed in fact tables, and dimensions which are descriptive characteristics related to fact data. Sales price, sale quantity, distant, speed, weight, and weight measurements are few examples of fact data in star schema. Often, A Star Schema having multiple dimensions is termed as Centipede Schema. It is easy to handle a star schema which has dimensions of few attributes.

Result: Thus the graphs for star schema are generated.