Effective Customer Patterns Analysis Using Open Source Weka Data Mining Tool

¹N.Suresh Kumar, ²Dr. M. Thangamani

¹Research Scholar, Anna University, Chennai ²Assistant Professor, Kongu Engineering College, India

Abstract: Data mining playing vital information in extracting useful information from large amount of data set. Apriori algorithm generate useful rule by finding frequent itemset from huge data set. In this paper can apply the Apriori Algorithm to generate rules for the given data set (bank) using Waikato Environment for Knowledge Analysis tool. Bank dataset is taken from UCI machine learning repository. These articles explore and visualize the apriori technique in data mining concept and analysis the customer good patterns to take the decision to give loan for customers.

Keywords: Data mining, Apriori algorithm, Waikato Environment

1. INTRODUCTION

The data mining represents mining the knowledge from large data. Topics such as knowledge discovery, query language, decision tree induction, classification and prediction, cluster analysis, and how to mine the Web are functions of data mining. Manual analyses are time consuming in the real world. In this situation, Waikato Environment for Knowledge Analysis tool (WEKA) can use for automating the task.

Weka is a collection of machine learning algorithms for data mining tasks. Classification was performed using WEKA in data mining research. WEKA is a data mining workbench that allows comparison between many different machine learning algorithms. In addition, it also has functionality for feature selection, data pre-processing and data visualization [1]. The algorithms can either be applied directly to a dataset or called from Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules and visualization. Well-suited for developing new machine learning schemes. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

2. RELATED WORK

The more associations between accident factors and accident severity were illustrated when applying Apriori algorithm [2]. The predictive Apriori algorithm could derive more number of rules that could be useful when studying the effect of each individual factor to accident severity. These results can help the decision makers in the traffic accident department to take actions based on various hidden patterns from the data. The swarm based techniques to extract association rules for student performance prediction as a multi-objective classification problem is analysis by [3]. In this algorithm takes a low convergence time and it used a few number of parameters. Honeybee Colony Optimization and Particle Swarm Optimization are the

two used metaheuristics to extract association rules. These are used in this investigation and WEKA, Rapidminer and KEEL tools are used for comparing the technique. Various type of analysis is carried out using association rules [4-6] in data mining through WEKA environments. The Weka tool is used for disease prediction. Data mining is a well known technique used by health organizations for classification of diseases such as dengue, diabetes and cancer in bioinformatics research [7].

3. EXPERIMENTS DESIGN

Implementation of Association Rule Mining is carried out in bank datasets using Weka tool.

3.1 Dataset description

Association rule works only with nominal type and the data values are discrete in nature. Number of Instances: 600 Number of Attributes: 12

3.2 Attributes description

Table.1 shows the list of attributes in bridge dataset. It also represents the data type for each attributes. Fig.1 and Fig.2 show the front panel of the Weka and location of the bank-data.csv file respectively. Bank datasets attributes are viewed by viewer in the WEKA explorer panel. It is illustrated in Fig. 3.

Attribute	Description	Data type
id	a unique identification number	Nominal
age	age of customer in years	Numeric
sex	MALE / FEMALE	Nominal
region	inner_city/rural/suburban/town	Nominal
income	income of customer	Numeric
married	is the customer married (YES/NO)	Nominal
children	number of children	Numeric
car	does the customer own a car (YES/NO)	Nominal
save_acct	does the customer have a saving account (YES/NO)	Nominal
current_acct	does the customer have a current account (YES/NO)	Nominal
mortgage	does the customer have a mortgage (YES/NO)	Nominal
рер	did the customer buy a PEP (Personal Equity Plan) after the last mailing (YES/NO)	Nominal

Table.1 List of attributes



Fig.1 Weka GUI Chooser



Fig.2 bank-data.csv file

🗘 Weka Explorer 🗕 🗖													ð X		
Preprocess Classify Cluster Associate Select at	ttributes Visualize														
Open file	Open URL Open DB Generate Undo Edit									5	Save				
Filter															
Choose None															Apply
Current relation				4						Viewer					X
Relation: bank-data	a11.1.	12		Relat	ion: hank-r	lata				Herei					
Instances: 600	Attribu	tes: 12		No.	id	age	sex	region	income married	children car	save act	current act	mortgage	pep	
Attributes					Nominal	Numeric	Nominal	Nominal	Numeric Nominal	Numeric Nominal	Nominal	Nominal	Nominal	Nominal	
All No	ne	Invert	Pattern	1	ID12101	48.0	FEMALE	INNER	17546.0 NO	1.0 NO	NO	NO	NO	YES	^
				2	ID12102	40.0	MALE	TOWN	30085.1YES	3.0 YES	NO	YES	YES	NO	- 11
No. Name				4	ID12103	23.0	FEMALE FEMALE	TOWN	20375.4 YES	3.0 NO	NO	YES	NO	NO	
1 id				5	ID12105	57.0	FEMALE	RURAL	50576.3 YES	0.0 NO	YES	NO	NO	NO	
2 age				6	ID12106	57.0	FEMALE	TOWN	37869.6 YES	2.0 NO	YES	YES	NO	YES	
3 sex				7	ID12107	22.0	MALE	RURAL	8877.07 NO	0.0 NO	NO	YES	NO	YES	
4 region				8	ID12108	58.0	MALE	TOWN	24946.6 YES	0.0 YES	YES	YES	NO	NO	
5 lincome				9	ID 12 109	37.0	FEMALE	SUBU	25304.3 YES	2.0 YES	NO	NO	NO	NO	
6 married				10	ID12110	54.0	MALE	TOWN	24212.1 YES	2.0 YES	YES	YES	NO	NO	
				- 11	ID12111	66.0 52.0		TOWN	59803.9 YES	0.0 NO	YES	YES	NO	NO	
9 Isave act				12	ID12112 ID12113	44.0	FEMALE FEMALE	TOWN	15735.8 YES	1.0 NO	VES	VES	VES	VES	
10 current_act				14	ID12114	66.0	FEMALE	TOWN	55204.7 YES	1.0 YES	YES	YES	YES	YES	
11 mortgage				15	ID12115	36.0	MALE	RURAL	19474.6 YES	0.0 NO	YES	YES	YES	NO	
12 рер				16	ID12116	38.0	FEMALE	INNER	22342.1 YES	0.0 YES	YES	YES	YES	NO	ľ
				17	ID12117	37.0	FEMALE	TOWN	17729.8 YES	2.0 NO	NO	NO	YES	NO	
				18	ID12118	46.0	FEMALE	SUBU	41016.0 YES	0.0 NO	YES	NO	YES	NO	- h
				19	ID12119	62.0	FEMALE	INNER	26909.2 YES	0.0 NO	YES	NO	NO	YES	
				20	ID12120	31.0	MALE	TOWN	22522.8 YES	0.0 YES	YES	YES	NO	NO	
				21	ID12121	61.U		TOWN	5/880.7 YES	2.0 NO	VEC	NU	NO	YES NO	
				22	ID12122 ID12123	54.0	MALE MALE	INNER	38446 6 YES	2.0 NO	VES	VES	NO	NO	
				24	ID12124	27.0	FEMALE	TOWN	15538.8 NO	0.0 YES	YES	YES	YES	NO	
				25	ID12125	22.0	MALE	INNER	12640.3 NO	2.0 YES	YES	YES	NO	NO	
				26	ID12126	56.0	MALE	INNER	41034.0 YES	0.0 YES	YES	YES	YES	NO	
				27	ID12127	45.0	MALE	INNER	20809.7 YES	0.0 NO	YES	YES	YES	NO	
				28	ID12128	39.0	FEMALE	TOWN	20114.0 YES	1.0 NO	NO	YES	NO	YES	
				29	ID12129	39.0	FEMALE	INNER	29359.1NO	3.0 YES	NO	YES	YES	NO	
	Remove			30	ID12130	61.0	MALE	RURAL	24270.1 YES	1.0 NO	NO	YES	NO NO	YES	v
	/ Series 75					£1.0		- I I I A	- CRAFTIN E			U	ndo (Ж	Cancel
Status OK					_	-	-	-	_	_	-			Log	×0

Fig.3 Weka Database Viewer and front panel

4. IMPLEMENTATION STEPS

Since Apriori algorithm works with only nominal data, the data set is preprocessed. Save the intermediate files after each step. The preprocessing WEKA is shown in Fig.4 and Fig.5. The Fig.6 represents the pure data after preprocessing.

The following preprocessing methods are applied:

• Removing the attribute:

- Remove the attribute id, since it uniquely identifies the tuples. It is done by selecting the remove attribute filter.
- Remove the attribute location, since it does not play a vital role in generating the rules.

0			Weka E	Explorer			- 0 ×
Preprocess Classify Cluste	er Associate Select attributes Visualize						
Open file	Open URL	Open DB	Gener	rate	Undo	Edit	Save
Filter							
Choose Remove -R 1	[Apply
Current relation Relation: bank-data-weka	, 🗘 weka.gui.Gener	icObjectEditor	X	Selected attribute Name: id		Туре:	Nominal
Instances: 600	weka.filters.unsupervised.attribute.Remove			Missing: 0 (0%)	Distinct: 6	500 Unique:	600 (100%)
Attributes	About			No. Label		Count	
All	A filter that removes a range of attribute	s from the dataset. More		1 ID121	01	1	<u>^</u>
		Capabilities		2 ID121 3 ID121	02	1	
No. Name				4 ID121	04	1	
1 🗸 id	attributeIndices 1			5 ID121	05	1	
2_age				6 ID121	06	1	
4 region	invertSelection False	V		7 ID121	07	1	
5 income				9 ID121	09	1	
6 married	Open Save	OK Cancel		10 ID121	10	1	
7 children				11 ID121	11	1	
8 car				12 ID121	12	1	
10 current act				13 ID121	13	1	V
11 mortgage				Class: pep (Nom)			✓ Visualize All
12 pep							
	Damaya						
	ĸemove						
Status							
OK							Log x0
-							



0			Weka	Explorer			- Ö X
Preprocess Classify Cluster As	sociate Select attributes Visualize						
Open file	Open URL	Open DB	Gen	erate	Undo	Edit	Save
Filter							
Choose Remove -R 1							Apply
Current relation Relation: bank-data-weka.filter Instances: 600	s.unsupervised.attribute.Remove-weka Att	.filters.unsupervised.attribute.Rem tributes: 11	ove-weka.filters.unsup	Selected attribute Name: age Missing: 0 (0%)	Distin	Ty ct: 50 Uniq	pe: Numeric ue: 0 (0%)
Attributes				Statistic		Value	
Al	None	Invert	Pattern	Minimum		18	
				Maximum		67	
No. Name				StdDev		42.395	
8 save_act 9 current_act 10 mortgage 11 pep				Class: pep (Nom)			√ Visualize Al
				87 89	78 69	87 57	90 83
	Remove						
0h-h-s				18		42.5	67
Status OK							Log 🛷 x0





19

										_						
Open file	Open URL	Open DE	3	Generate				Undo				Edit			Save	
se Remove -R 1				_												
relation				4						View	/er					Х
tion: bank-data-weka.filt	ers.unsupervised.attribute.Remove	e-weka.filters.unsupervised.attribu	ute.Remove-weka.filters.u	sup Rel	ation: bank-	lata-wek	a.filters.ur	nsupervised	l.attribute	e.Remove	-weka.filt	ers.unsuper	vised.attribut	e.Remove-v	veka.filter	s
ces: 600		Attributes: 11		N), age	sex	region	income	married	children	car	save_act	current_act	mortgage	рер	
es					Numeric	Nominal	Nominal	Numeric	Nominal	Numeric	Nominal	Nominal	Nominal	Nominal	Nominal	
		· .		1	48.0	FEMALE	INNER	17546.0	NO	1.0	NO	NO	NO	NO	YES	۸
All	None	Invert	Pattern	2	40.0	MALE EEMALE		30085.1	YES	3.0	YES	NU	YES	YES NO	NO	
News				4	23.0	FEMALE	TOWN	20375.4	YES	3.0	NO	NO	YES	NO	NO	
Ivame				5	57.0	FEMALE	RURAL	50576.3	YES	0.0	NO	YES	NO	NO	NO	
1 age				6	57.0	FEMALE	TOWN	37869.6	YES	2.0	NO	YES	YES	NO	YES	
3 region				7	22.0	MALE	RURAL	8877.07	NO	0.0	NO	NO	YES	NO	YES	
4 income				8	58.0	MALE	TOWN	24946.6	YES	0.0	YES	YES	YES	NO	NO	
5 married				9	37.0	FEMALE	SUBU	25304.3	YES	2.0	YES	NO	NO	NO	NO	
6 children				10	54.0	MALE	TOWN	24212.1	YES	2.0	YES	YES	YES	NO	NO	
7 car				11	52.0			59803.9	YES NO	0.0	NO	YES	YES	NU	NO	
8 save_act				12	52.0 44.0	FEMALE FEMALE	TOWN	20030.0	VES	1.0	NO NO	VES	VES	VES	VES	
9 current_act					66.0	FEMALE	TOWN	55204.7	YES	1.0	YES	YES	YES	YES	YES	
10 mortgage					36.0	MALE	RURAL	19474.6	YES	0.0	NO	YES	YES	YES	NO	1
ш_рер				16	38.0	FEMALE	INNER	22342.1	YES	0.0	YES	YES	YES	YES	NO	
				17	37.0	Female	TOWN	17729.8	YES	2.0	NO	NO	NO	YES	NO	
				18	46.0	Female	SUBU	41016.0	YES	0.0	NO	YES	NO	YES	NO	
				19	62.0	FEMALE	INNER	26909.2	YES	0.0	NO	YES	NO	NO	YES	
				20	31.0	MALE	TOWN	22522.8	YES	0.0	YES	YES	YES	NO	NO	
				21	50.0	MALE MALE	TOWN	16407.3	VES	2.0	NO	VES	VES	NO	ICS NO	
				22	54.0	MALE	INNER.	38446.6	YES	2.0	NO	YES	YES	NO	NO	
				24	27.0	FEMALE	TOWN	15538.8	NO	0.0	YES	YES	YES	YES	NO	v
													Undo	OK	Cance	
	Ren	move														

Fig.6 After preprocessing



Discretization: Association rule mining can be applied on categorical data, so the three numeric attributes erected, length and lanes in the data set are discretized and it shown in Fig.7. The Fig.8 represents the how to modify the normalized value for discretization.

0		Weka	Explorer			_ 🗇 🗙
Preprocess Classify Cluster Associate Select attributes Visualize						
Open file Open URL	Open DB	Gene	erate	Undo	Edit	Save
Filter Choose Remove -R 1						Apply
Current relation Relation: bank-data-weka.fiters.unsupervised.attribute.Remove-weka.fi Instances: 600 Attrib	ters.unsupervised.attribute.Remove-wek utes: 11	ka.filters.unsup	Selected attribute Name: children Missing: 0 (0%)	Distinct	דן ג 4 Unic	ipe: Numeric jue: 0 (0%)
Attributes			Statistic		Value	
All None	Invert Pa	attern	Minimum Maximum		0	
No. Name			Mean		1.012	
2 sex 3 region 4 income 5 maried 6 ✓ dridren 7 car 8 save_act 9 current_act 10 mortgage 11 pep			Class: pep (Nom) 280 0	135		V Visualize All
Status OK			U		1.0	Log "x0

21



Fig.7 Discretization in customer datasets

The attribute children has only the values 0, 1, 2 and 3, so it is discretized by just removing the keyword numeric from the input file and replacing it with set of discrete values.

📇 l 🔜 为 🤭 -	-	bank-r1 - WordPad	- 🗖 🗙
File Home	View		^ ()
Paste	Courier New • B I U abe × ₂ × ³	· A A 译 译 注 - ↓ ↓ · ↓ · ↓ · ↓ · ↓ · ↓ · ↓ · ↓ · ↓ ·	
Clipboard	Font	Paragraph Insert Editing	
		L · 1 · · · · · · · · · · · · · · · · ·	
		<pre>@</pre>	
		57, FEMALE, TOWN, 37869.6, YE9, 2, NO, YE9, YE9, NO, YE9 22, MALE, RURAL, 897.07, NO, 0, NO, NO, YE9, NO, YE9 58, MALE, RURAL, 897.07, NO, 0, NO, NO, YE9, YE9, NO, NO 37, FEMALE, SUBURBAN, 25304.3, YE9, 2, YE9, NO, NO, NO 54, MALE, TOWN, 242121, YE9, 22, YE9, YE9, YE9, NO, NO 66, FEMALE, TOWN, 5903.9, YE9, 0, NO, YE9, YE9, NO, NO 52, FEMALE, INNER_(TIY, 26584, 8, NO, 0, YE9, YE9, NO, NO 54, FEMALE, INNER_(TIY, 26584, 8, NO, 0, YE9, YE9, YE9, YE9, NO 44, FEMALE, TOWN, 15735.8, YE9, 1, YE9, YE9, YE9, YE9, YE9 66, FEMALE, TOWN, 5504.7, YE9, 1, YE9, YE9, YE9, YE9, YE9 36, MALE, RURAL, 19474.6, YE9, O, NO, YE9, YE9, YE9, YE9 36, MALE, RURAL, 19474.6, YE9, O, NO, YE9, YE9, YE9, YE9, NO	
			100% 😑 💷 🕀

Fig.8 Labels assigned for the attributes and the changes in the instances.

The input file with the above changes is shown Fig.9

Fin Yes A A A Image: Decomposition	<u>≅ ⊒ > C = </u>	bank-r2 - WordPad	- 0 ×
Control New (13) N N (14) N N (15) N (14) N	File Home View		^ 😯
Cipboard Fort Paragraph next Edding Protection and a constrained at the second at the	Paste Courier New 13 B I U abs ×₂ ײ	· A A 详 律 语 := ・ := · ○ · · · · · · · · · · · · · · · · ·	
<pre>% ************************************</pre>	Clipboard Font	Paragraph Insert Editing	
<pre>grelation bank-data-weka.filters.unsupervised.attribute.Remove- weka.filters.unsupervised.attribute.Remove-Rl @attribute age numeric @attribute region (INNER_CITY,TON,RUBAL,SUBURBAN) @attribute income numeric @attribute married (No,YE8) @attribute married (No,YE8) @attribute car (No,YE8) @attribute car (No,YE8) @attribute car (No,YE8) @attribute car (No,YE8) @attribute carrent (No,YE8) @attribute carrent (No,YE8) @attribute gare (No,YE8) @attribute (NO,YE8) @attribute (NNER, CITY, 1575.4,YE8,NO,NO,NO,NO S7,FFMALE,TONN, 3765.5,YE8,JO,YE8,NO,NO,NO S5,FFMALE,TONN, 3765.5,YE8,JO,YE8,NO,NO,NO S5,FFMALE,TONN, 5421.1,YE8,JO,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,ALE,RUNAL, 9971.0,YE8,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,ALE,RUNAL, 9971.0,YE8,YE8,YE8,YE8,NO %a,ALE,RUNAL, 9971.0,YE8,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,JO,YE8,YE8,YE8,NO %a,ALE,RUNAL, 9971.0,YE8,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,YE8,YE8,YE8,NO %a,FFMALE,TONN, 5532.1,YE8,YE8,YE8,NO %a</pre>		1	
an FEMALE, INNER CITT.22342.1.TES.0.TES.TES.TES.TES.TO		<pre>Grelation bank-data-weka.filters.unsupervised.attribute.Remove- weka.filters.unsupervised.attribute.Remove- Weka.filters.unsupervised.attribute.Remove-R1 @attribute age numeric @attribute age numeric @attribute sex (FFMALE,MALE) @attribute income numeric @attribute income numeric @attribute children (0,1,2,3) @attribute children (0,1,2,3) @attribute age act (NO,YES) @attribute age act (NO,YES) @attribute age (YES,NO) @data 48,FEMALE,INNER_CITY,1546,NO,1,NO,NO,NO,NO,YES 40,MALE,TONN_30085.1,YES,3,YES,NO,YES,YES,NO 51,FEMALE,NON,30085.1,YES,3,VES,NO,YES,YES,NO 53,FEMALE,FINNER_CITY,15575.4,YES,0,YES,YES,NO 53,FEMALE,FINNER_CITY,15575.4,YES,0,NO,YES,YES,NO 53,FEMALE,FINNER_CITY,15575.4,YES,0,NO,YES,NO,NO 53,FEMALE,FINNER_CITY,15575.4,YES,0,NO,YES,NO,NO 53,FEMALE,FINNER_CITY,15575.4,YES,0,NO,YES,NO,NO 54,FEMALE,FINNER_CITY,15575.4,YES,0,NO,YES,NO,NO 55,FEMALE,FINNER_CITY,0,NO,O,NO,YES,NO,NO 56,FEMALE,FINNER,25076.3,YES,2,NO,YES,NO,NO 57,FEMALE,FINNER,25076.3,YES,2,NO,YES,NO,NO 54,MALE,FONN,24946.6,YES,0,YES,YES,NO,NO 54,MALE,FONN,24946.6,YES,0,NO,YES,YES,NO,NO 54,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 54,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 54,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 55,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 56,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 57,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 58,MALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 59,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 50,FEMALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 52,FEMALE,FONN,5506.3,YES,YES,NO,NO 53,MALE,FONN,5506.3,YES,2,NO,YES,YES,NO,NO 54,FEMALE,FONN,5506.3,YES,YES,YES,YES,NO 55,MALE,FONN,506.3,YES,YES,NO,NO,NO 56,FEMALE,FONN,506.3,YES,YES,YES,NO,NO 56,FEMALE,FONN,506.3,YES,YES,NO,NO,NO 57,FEMALE,FONN,5506.3,YES,YES,NO,NO 56,FEMALE,FONN,5506.3,YES,YES,NO,NO 56,FEMALE,FONN,5506.3,YES,YES,NO,NO 56,FEMALE,FONN,5506.3,YES,YES,NO,NO 56,FEMALE,FONN,5506,YES,YES,NO,NO 56,FEMALE,</pre>	
			100% 😑 💷 🕀



Ø				Weka	Explorer				- 0 X
Preproces	^S Classify Cluster Assoc	iate Select attributes Visualiz	e						
	Open file	Open URL	Open DB	, Gei	nerate		Undo	Edit	Save
Filter									
Choos	e Remove -R 1								Apply
-Current r Relatio Instance	elation m: bank-data-weka.filters.u es: 600	nsupervised.attribute.Remove-ı	weka.filters.unsupervised.attribute Attributes: 11	.Remove-weka.filters.unsup	Selected Name Missing	attribute : children : 0 (0%)	Distinct	Type : 4 Unique	e: Nominal e: 0 (0%)
Attribute	1				No.	Label		Count	
	Al	None	Invert	Pattern		10		263	
					_	21		135	
No.	Name					4 3		68	
	1 age							I	
	2 sex								
	4 lincome				-				
	5 married				1				
	6 🖌 children								
	7 car								
	8 save_act								
	9 current_act								
	10 mor igage 11 nen				Class: ner	n (Nom)			V Visualize All
		Remo	IVÊ		263		135	134	63
Status OK									Log 🛷 × C



			Weka	Explorer			- Ö
eprocess Classify Cluster Associa	te Select attributes Visualize						
Open file	Open URL	Open DB	Gen	erate	Undo	Edit	Save
ter							
Choose Remove -R 1							Ap
rrent relation				Selected attribute			
Relation: bank-data-weka.filters.un nstances: 600	supervised.attribute.Remove-wel A	eka.filters.unsupervised.attribute.R Attributes: 11	temove-weka.filters.unsup	Name: age Missing: 0 (0%)	Distin	ct: 50 Uni	ype: Numeric que: 0 (0%)
ributes				Statistic		Value	
All	None	Invert	Pattern	Minimum		18	
				Maximum		67	
). Name				Mean StdDev		42.395	
4 nome 5 married 6 dhidren 7 car 8 save_act 9 current_act 10 mortgage 11 pep				Class: pep (Nom) 87 89		87	90 63

Fig.11 Attribute Age details

The other two attributes age and income have different (continuous) values for different instances so it is discretized by applying discretize filter in the WEKA tool The age and income ranges are divided into three categories (arbitrary) and the attributes are discretized. Fig.12 and Fig.13 shows parameter modification in age and income attributes.



				Weka	Explorer					- 0
eprocess Classify Cluster Ass	ociate Select attributes Visualize									
Open file	Open URL		Open DB	Gen	erate	Undo		Edit	Save	
ter										
Choose Discretize -B 3 -M -1	1.0 -R first-last									Appl
rrent relation Relation: bank-data-weka.filters nstances: 600	.unsupervised.attribute.Remove-wek A	(a.filters.unsupervis (ttributes: 11	ed.attribute.Remove-we	ka.filters.unsup	Selected attribute Name: age Missing: 0 (0%)		Distinct: 50		Type: Numeric Unique: 0 (0%)	
ributes					Statistic			Value		
Al	None	Invert	F	Pattern	Minimum Maximum			18 67		
o. Name					Mean StdDev			42.395		
8 save_act 9 current_act 10 mortgage 11 pep			attributes in the d	ataset into nomina attributeIndices fi bins 3	st-last	Capabilities		87	۷	Visualiz 90
			ucarcuvelynoniste	findNumBins F ignoreClass F	alse	v	78	57	63	
				invertSelection F makeBinary F	alse	v				
			use	EqualFrequency F	alse	v				
	Remove		Open	Save	OK	Cancel		Т		
atus					18			42.5		
(Log	10

Fig.12 Discretization in Age attributes



0		Weka Explorer				- 0 ×
Preprocess Classify Cluster Associate Select attributes Visualize						
Open file Open URL	Open DB	Generate	Unc	lo	Edit	Save
Filter						
Choose Discretize -B 3 -M -1.0 -R first-last						Apply
Current relation Relation: bank-data-weka, filters.unsupervised.attribute.Remove-weka, filt Instances: 600 Attribu	ers.unsupervised.attribute.Remove-weka. Ites: 11	filters.unsup Selected at Name: ir Missing: 0	tribute ncome (0%)	Distinct: 599	Ty Uniq	pe: Numeric ue: 598 (100%)
Attributes		Statistic			Value	
All None	Invert Patt	tern			5014.21	
		Maximum			63130.1	
No. Name	🥥 weka.gui.	.GenericObjectEditor	× _		12899.468	
4 √ ncome 5 married 6 dhidren 7 car 8 save_act 9 arrent_act 10 mortgage 11 pep	An instance filter that discretize: attributes in the dataset into nor attributes in the dataset into nor desired/WeightOfInstancesPerInterva findhumBin: ignoreClass inverISelector makeBinary useEqualFrequency Open Save	s a range of numeric minal attributes. s first-last s 3 d -1.0 s False False y False y False OK	More Capabilities	08	4	V Visualize Al
Remove						
Ctable		5014.21			34072.16	63130.1
OK .						Log 🛷 x O

26



The following Fig.14 depicts the labels assigned for the instances (one instance highlighted) of



27

the attributes age and income. Fig.15 shows the customer dataset after discretization. Fig.14 Labels assigned for the instances of the attributes age and income.

0											Weka E	Explorer						- 0 >
Preproces	S Classify Cluster As	ssociate Select at	tribute	s Visualia	ze													
	Open file		Open	URL		Open DB					Gener	ate	ie		Undo		Edit	Save
Filter																		
Choos	e Discretize -B 3 -M	-1.0 -R first-last																Apply
- Current r	elation											Selected att	ribute					
Relation: bank-data-weka.filters.unsupervised.attribute.Remove-R1-weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1 Instances: 600 Attributes: 11						Name: a Missing: 0	ge (0%)			Distinct: 3	Ty Uniq	pe: Nominal ue: 0 (0%)						
Attribute	s											No.	Label				Count	
	All	Nor	P			Invert			Pa	ittern		1	0_34				195	
			Hone									2	35_51				214	
No.	Name											3	52_max				191	
	1 age		_											_	_			
	2 sex		4						View	/er					х			
<u> </u>	3 region		Relat	ion: bank-	-data-wek	a.filters.u	nsupervise	d.attribut	e.Remove	-R1-weka	filters.uns	upervised.att	ribute.Discre	tize-B3-M	-1			
	5 married		No.	age	sex	region	income	married	children	car	save_act	current_act	mortgage	pep				
	6 children			Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal	Nominal				
	7 car		1	35_51	FEMALE	INNER	0_24386	NO	1	NO	NO	NO	NO	YES	^			
	8 save_act		2	35_51 52 may	MALE	TOWN	24387	YES	3	VEC	NU	YES	TES NO	NO				
	9 current_act		4	0.34	FEMALE	TOWN	0_24386	YES	3	NO	NO	YES	NO	NO				
<u> </u>	10 mortgage		5	52 max	FEMALE	RURAL	43759	YES	0	NO	YES	NO	NO	NO				Vigualize Al
			6	52_max	FEMALE	TOWN	24387	YES	2	NO	YES	YES	NO	YES				• Visualize A
			7	0_34	MALE	RURAL	0_24386	NO	0	NO	NO	YES	NO	YES				
			8	52_max	MALE	TOWN	24387	YES	0	YES	YES	YES	NO	NO		214		
			9	35_51	FEMALE	SUBU	24387	YES	2	YES	NO	NO	NO	NO				191
			11	52_max	FEMALE	TOWN	43759	YES	2	NO	YES	YES	NO	NO				
			12	52 max	FEMALE	INNER	24387	NO	0	YES	YES	YES	YES	NO				52_max [191]
			13	35_51	FEMALE	TOWN	0_24386	YES	1	NO	YES	YES	YES	YES				
			14	52_max	FEMALE	TOWN	43759	YES	1	YES	YES	YES	YES	YES				
			15	35_51	MALE	RURAL	0_24386	YES	0	NO	YES	YES	YES	NO				
			16	35_51	FEMALE	INNER	0_24386	YES	0	TES	TES	TES	YES	NO				
			18	35_51	FEMALE	SUBL	24387	YES	2 0	NO	YES	NO	YES	NO				
			19	52 max	FEMALE	INNER	24387	YES	0	NO	YES	NO	NO	YES				
L			20	0_34	MALE	TOWN	0_24386	YES	0	YES	YES	YES	NO	NO				
			21	52_max	MALE	INNER	43759	YES	2	NO	YES	NO	NO	YES				
			22	35_51	MALE	TOWN	0_24386	YES	2	NO	YES	YES	NO	NO				
Status			23	52_max	MALE	INNER	24387	YES	0	NO	YES	YES	NO	NO				
OK			24	0_34	MALE	INNER	0_24386	NO	2	VES	VES	VES	NO	NO				Log
			20	0_04	MALE	DWNER	0_24500	110	4	163	163	163						

Fig.15 Customer Dataset after discretization

Apriori Algorithm Implementation in Weka:

The preprocessed data file is used for Association rule mining (Apriori Algorithm) and the following rules are generated by setting the necessary measures such as support and confidence is shown in Fig.16 and Fig.17.







0	Weka Explorer	- 0 ×
Preprocess Classify Cl	luster Associate Select attributes Visualize	
Associator		
Choose Apriori -	N 10-TO-C 0.9-D 0.05-U 1.0-M 0.1-S-1.0-c-1	
	Associator output	
Start Stop		٨
Result list (right-click	Apriori	
01:31:10 - Apriori		
	Minimum summart: 0 1 (60 instances)	
	Minimum metric <confidence>; 0.9</confidence>	
	Number of cycles performed: 18	
	Generated sets of large itemsets:	
	Size of set of large itemsets L/11: 28	
	Size of set of large itemsets L(2): 232	
	Size of set of large itemsets L(3): 524	
	Size of set of large itemsets L(4): 277	
	Size of set of large itemsets L(5): 33	
	Dest unles frust.	
	best fules lound:	
	1. income=43759 max 80 ==> save act=YES 80	
	2. age=52_max income=43759_max 76 ==> save_act=YES 76 conf:(1)	
	3. income=43759_max current_act=YES 63 ==> save_act=YES 63 conf:(1)	
	4. age=52_max income=43759_max current_act=YES 61 ==> save_act=YES 61 conf:(1)	
	5. children=0 save_act=YES mortgage=N0 pep=N0 74 ==> married=YES 73 conf: (0.99)	
	6. sex=YEMALE children=0 mortgage=N0 pep=N0 64 ==> married=YES 63 conf: (0.98)	
	/, cnlidten=U current_act=1cs mortgage=UU pep=NU 82 ==> married=1cs 80	
	6. CHILIELEO MOTOJAGENNO PEPENNO IOI ==> MALTIEU-ILS IO4 CONI:(0.97) 0. income_40350 may curvent act-VPS 63 =-> ace-52 may 61 conf:(0.03)	
	10. income=43759 max save act=YES current act=YES 63 ==> ace=52 max 61 conf:(0.97)	
		¥
	(>
Status		
ОК		Log 📣 x0

Support and Confidence threshold:

The following Fig.17 shows the parameters set



0	Weka Explorer	- 0 ×
Preprocess Classify Cluster Associate Select attributes Vis	ualize	
Associator		
Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -5	-1.0 -c -1	
Start Stop	or output	
Result list (right-click for options)	٢	weka.gui.GenericObjectEditor
	weka.associations.Apri	iori
	About	
	Class implement	ing an Apriori-type algorithm. More
		Capabilities
	ca	ar False 🗸
	dassInde	-1
	delt	ta 0.05
	lowerBoundMinSuppor	vrt 0.1
	metricTyp	Confidence V
	minMetr	ric 0.9
	numRule	es 10
	outputItemSet	ts False v
	removeAllMissingCo	False V
	significanceLev	el -1.0
	upperBoundMinSuppor	rt 1.0
	verbos	False v
	Open	Save OK Cancel
Status OK		Log 🗸 x K

Fig. 17. Minimum Support and Confidence threshold

Output-Rules Generated:

The screen shot shows the rules generated by applying Apriori Algorithm for association rule mining is shown in Fig.18.

```
== Run information ===
```

```
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation: bank-data-weka.filters.unsupervised.attribute.Remove-R1-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R1-
weka.filters.unsupervised.attribute.Discretize-B3-M-1.0-R4
Instances: 600
Attributes: 11
age
sex
region
income
```

married children car save_act current_act mortgage

pep



0	Weka Explorer	- 0 ×
Preprocess Classify Clu	uster Associate Select attributes Visualize	
Associator		
Choose Apriori -	N 10-T0-C 0.9-D 0.05-U 1.0-M 0.1-S-1.0-c-1	
Start Stop	Associator output	
Result list (right-dick 01:31:10 - Apriori	Apriori Minimum support: 0.1 (60 instances)	^
	Minimum metric <confidence>: 0.9 Number of cycles performed: 18</confidence>	
	Generated sets of large itemsets: Size of set of large itemsets L(1): 28	
	Size of set of large itemsets L(2): 232	
	Size of set of large itemsets L(3): 524	
	Size of set of large itemsets L(5): 33	
	Best rules found:	
	<pre>1. income=43759_max 80 ==> save_act=YES 80 conf:(1) 2. age=52_max income=43759_max 76 ==> save_act=YES 76 conf:(1) 3. income=43759_max current_act=YES 63 ==> save_act=YES 63 conf:(1) 4. age=52_max income=43759_max current_act=YES 61 ==> save_act=YES 61 conf:(1) 5. children=0 save_act=YES mortgage=NO pep=NO 74 ==> married=YES 73 conf:(0.99) 6. sex=FEMALE children=0 mortgage=NO pep=NO 74 ==> married=YES 63 conf:(0.98) 7. children=0 current_act=YES mortgage=NO pep=NO 82 ==> married=YES 80 conf:(0.98) 8. children=0 mortgage=NO pep=NO 107 ==> married=YES 104 conf:(0.97) 9. income=43759_max current_act=YES 63 ==> age=52_max 61 conf:(0.97) 10. income=43759_max save_act=YES current_act=YES 63 ==> age=52_max 61 conf:(0.97) </pre>	× >
Status OK	Log	×0



=== Associator model (full training set) === Apriori

Minimum support: 0.1 (60 instances) Minimum metric <confidence>: 0.9 Number of cycles performed: 18 Generated sets of large itemsets: Size of set of large itemsets L(1): 28 Size of set of large itemsets L(2): 232 Size of set of large itemsets L(2): 524 Size of set of large itemsets L(4): 277 Size of set of large itemsets L(5): 33

Best rules found:

- 1. income=43759_max 80 ==> save_act=YES 80 conf:(1)
- 2. age=52_max income=43759_max 76 ==> save_act=YES 76 conf:(1)
- 3. income=43759 max current act=YES 63 ==> save act=YES 63 conf:(1)
- 4. age=52 max income=43759 max current act=YES 61 ==> save act=YES 61 conf:(1)
- 5. children=0 save_act=YES mortgage=NO pep=NO 74 ==> married=YES 73 conf:(0.99)
- 6. sex=FEMALE children=0 mortgage=NO pep=NO 64 ==> married=YES 63 conf:(0.98)
- 7. children=0 current act=YES mortgage=NO pep=NO 82 ==> married=YES 80 conf:(0.98)
- 8. children=0 mortgage=NO pep=NO 107 ==> married=YES 104 conf:(0.97)
- 9. income=43759_max current_act=YES 63 ==> age=52_max 61 conf:(0.97)
- 10. income=43759 max save_act=YES current_act=YES 63 ==> age=52 max 61 conf:(0.97)

CONCLUSION:

The above rules infer that most of the customers whose age is above 52 and income greater than 43,000 have a saving account and current account. Most of the customers who has no children, no mortgage and no personal equity plan is married.

REFERENCES

- Donn Morrison, Ruili Wang, Liyanage C. De Silva, Ensemble methods for spoken emotion recognition in call-centres, Speech Communication, Elsevier, Vol. 49, pp.98-112, 2007
- 2. Amira A. El Tayeb, Vikas Pareek, Abdelaziz Araar, Applying Association Rules Mining Algorithms for Traffic Accidents in Dubai, International Journal of Soft Computing and Engineering, Vol. 5, No.4, pp. 1-12, 2015
- Roghayeh Saneifar and Mohammad Saniee Abadeh, Jan Zizka et al., Association Rule Discovery for Student Performance Prediction Using Metaheuristic Algorithms, pp. 115–123, 2015, DOI: 10.5121/csit.2015.51510
- 4. Amit Dipchandji Kasliwal, Dr. Girish S. Katkar, "Web Usage mining for Predicting User Access Behaviour", International Journal of Computer Science and Information Technologies, Vol.6, No.1, pp-201-204, 2015.
- Rahul Neve 1, K.P Adhiya ", Comparative Study of Web Mining Algorithms for Web Page Prediction in Recommendation System", international Journal of Advanced Research in Computer and Communication Engineering, Vol. 2, No.1, pp.969-976, 2013
- 6. Dhruva Mistry, Kirti Sharma, Samip A.Patel, Recommend Websites through Weblog Files using Association Rule, International Journal of Computer Applications, Vol.126, No.2, pp.16-19, 2015.
- 7. Kashish Ara Shakil, Shadma Anis and Mansaf Alam, Dengue Disease Prediction Using Weka Data Mining Tool, Pp.1-26, 2015.



Author(s) Biography



Thangamani completed her B.E., from Government College of Technology, Coimbatore, India. She completed her M.E in Computer Science and Engineering from Anna University and PhD in Information and Communication Engineering from the renowned Anna University, Chennai, India in the year 2013.

Dr. M. Thangamani possesses nearly 23 years of experience in research, teaching, consulting and practical application development

to solve real-world business problems using analytics. Her research expertise covers Medical data mining, machine learning, cloud computing, big data, fuzzy, soft computing, ontology development, web services and open source software. She has published nearly 70 articles in refereed and indexed journals, books and book chapters and presented over 67 papers in national and international conferences in above field. She has delivered more than 60 Guest Lectures in reputed engineering colleges and reputed industries on various topics. She has got best paper awards from various education related social activities in India and Abroad. She has organized many self-supporting and government sponsored national conference and Workshop in the field of data mining, big data and cloud computing. She continues to actively serve the academic and research communities and presently guiding Ph.D Scholars under Anna University. She is also seasonal reviewer in IEEE Transaction on Fuzzy System, international journal of advances in Fuzzy System and Applied mathematics and information journals. She has organizing chair and keynote speaker in international conferences in India and countries like California, Dubai, Malaysia, Singapore, Thailand and China. She has received many awards for academic activities. She is on the editorial board and reviewing committee of leading research journals, which includes her nomination as the Editor in chief to International Scientific Global journal for Engineering, Science and Applied Research (ISGJESAR) & International Research Journal in Global Engineering and Sciences (IRJGES) and on the program committee of top international data mining and soft computing conferences in various countries. She has Life Membership in ISTE, Member in CSI, International Association of Engineers and Computer Scientists in China, IAENG, IRES, Athens Institute for Education and Research and Life member in Analytical Society of India. She is currently working as Assistant Professor at Kongu Engineering College at Perundurai, Erode District.



Mr. N.Suresh Kumar is currently working as Assistant Professor in the Department of Information Technology at Sri Ramakrishna Engineering College, Coimbatore. He is a part time research scholar under Anna University, Chennai. He has a Master's degree in Software Engineering (2010) at Anna University of Technology, Coimbatore. He has nearly 13 years of academic experience and 2 Years of Industry experience. He is proficient in Data Mining, Software Engineering, Database Management System, Web services and

ontology development. He has published the papers in International and National Journals and presented the papers in National conferences. He is a life member of ISTE, IAENG and IACSIT.