ECLT5810/SEEM5750 E-Commerce Data Mining Technique

2024-2025 First Term

Assignment 1 Solution reference

Please note that the provided solution is not the only correct approach for solving the problems.

1) Data preprocessing

3.

Open file Open URL	weka.gui.GenericObjectEditor		×	Edit	Save
ilter	weka.filters.unsupervised.attribute.Dis	cretize			
Choose Discretize -B 6 -M -1.0 -R 1 -pr	recisio About				Apply Stop
Current relation	An instance filter that discretizes a ran	ae of numeric	More		
Relation: bank-additional	attributes in the dataset into nominal a	ttributes.	Capabilities	vet: 67	Type: Numeric
Instances: 4119				KC: 67	Unique: 5 (0%)
in money	attributeIndices	; 1		10	Value
All None	binBangePrecision	6		88	
No.	Nar			40.114	
1 age	bins	6		10.313	
2 🗌 job	debug	False	~		
3 marital	desiredWeightOffpstapsorPedptapse	10		1	
4 default	desired weight Onlistances remitterval	-1.0			
6 housing	doNotCheckCapabilities	False	~		
7 🗌 Ioan	findNumBins	False	~		
8 contact	1	(False			 Visualize
9 month	ignoreClass	raise	~		
10 day_of_week	invertSelection	False	~		
12 Campaign	makeBinan	False	~	1	
13 pdays	makeoniary				
14 previous	spreadAttributeWeight	False	~		
15 poutcome	useBinNumbers	False	~	_	
16 emp.var.rate	useEqualErequence	Ealco			
18 cons.conf.idx	usezquairrequency	rdise	Ť		
40	0	01	Const	-	
atus		18		53	
Weka Explorer					- 0
Weka Explorer Treprocess Classify Cluster Assoc	ciate Select attributes Visualize				- 0
Weka Explorer Preprocess Classify Cluster Assoc Open file_ Open URL	ciate Select attributes Visualize	ate	Undo	Edit	- 🗆
Weka Explorer reprocess Classify Cluster Asson Open file Open URL	ciate Select attributes Visualize Open DB Genere	ate	Undo] Edit	
Weka Explorer reprocess Classify Cluster Asso Open file Iter Choose Discretize -86 - M - 1.0 - R1 - pr	ciate Select attributes Visualize Open DB_ Generi ecision 6	ate	Undo	Edit	- D
Weka Explorer reprocess Classify Cluster Asso Open file Open URL tter Choose Discretize -B 6 -M - 1.0 - R 1 - pr urent relation	clate Select attributes Visualize Open DB Generi ecision 6	ate	Undo	Edit	- D Save
Weka Explorer Treprocess Classify Cluster Asso Open file Open URL Uter Choose Discretize -B.6 -M -1.0 -R.1 -pr urent relation Relation: bank-additional-weka.filters.uns	ciate Select attributes Visualize Open DB_ Gener ecision 6 supervised.att. Attributes: 21	ate Selected attrib Name: age	Undo) Edit	
Weka Explorer Teprocess Classify Cluster Asso Open URL Open URL Choose Discretize - 86 - M - 1.0 - R1 - pr urrent relation Relation: bank-additional-weka.filters.um statances 4119	ciate Select attributes Visualize Open DBGener ecision 6 aupervised.attAttributes: 21 Sum of weights: 4119	selected attrib Name: age Missing: 0 (09	Undo ute 6) Di	Edit_	
Weka Explorer reprocess Classify Cluster Asso Open URL Uter Choose Discrettze-B6-M-1.0-R1-pr urrent relation Relation: bank-additional-weka.filters.une stance: 4119 tributes	clate Select attributes Visualize Open DB Generi ecision 6 supervised.att Attributes: 21 Sum of weights: 4119	ate Selected attrib Name: age Missing: 0 (09 No.	Undo ute 6) Di Label	Edit	- C Save Apply Sto Type: Nominal Unique: 0 (0%) Weight
Weka Explorer Treprocess Classify Cluster Asso Open file Open URL Iter Choose Discretize -86M -1.0 - R 1 - pr arrent relation Relation-bank-additional-weka.filters.ann shances: 4119 Tributes All None	clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern	ate Selected attrib Name: age Missing: 0 (09 No. 1 '(-in 2 '(29	Undo ute 6) Di Label 1-29.6666671' 666667-41 33333	stinct: 6 Count 551 1938	Apply Sto Type: Nominal Unique 0 (%) Weight 551 1938
Weka Explorer reprocess Classify Cluster Asso Open file Open URL tter Choose Discretize -B 6 -M -1.0 - R 1 - pn ruren trelation Relation bank-additional-weka.filters.un stances. 4119 No No.	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29 3 '(41	Undo ute 6) Di Label (-29.66667) .666667-41.3333 33333-53]	stinct: 6 Count 551 	Apply Sto Type: Nominal Unique: 0 (7%) Weight 551 1938 1115
Weka Explorer reprocess Classify Cluster Asso Open file Open URL Iter Cluster Asso Choose Discretize - 86 - M - 1.0 - R1 - pr retration Relation: bin-additional-weka.filters.um nstances: 411 None No. 1 age	ciate Select attributes Visualize Open DB_ Genera ecision 6 supervised.att_ Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Name: age Missing: 0 (0? No. 1 '(-in 2 '(29 3 '(41 4 '(53	Undo ute (b) Di Label (-29.666667)' (-666667-41.3333 -33333-513)	Edit stinct: 6 551 	Apply Stop Type: Nominal Unique: 0 (0%) Weight 551 1938 1115 457
Weka Explorer reprocess Classify Cluster Asso Open IIIa Open URL Iter	ciate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 1 (-1) 2 '(29 3 '(41 4 '(53 5 '(64	Undo ute 6) Di 1.29.66667] 6.66667-11.3333 33333-53] -6.4.666677 .666667-76.33333	stinct: 6 Solution: Still:	
Weka Explorer reprocess Classify Cluster Asso Open file Open URL Iter Choose Discretize -86 -M -1.0 - R1 - pr arrent relation Relation-bank-additional-weka.filters.uns shances. 4119 None All None 1 age 2 _ lob 3 _ marital 4 _ does view	clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 1 '-(-in 2 '2-23 3 '(41 4 '(53) 5 '(64) 6 '(76)	Undo ute (a) Di Label 1-29 666671 666667-11.3333 .33333-531 -64.6666677 .666667-7.633333 .333333-531	Edit stinct: 6 Count 551 1938 1115 457 40 18	
Weka Explorer reprocess Classify Cluster Asso Open file. Open velocities Open velocities	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29 3 '(41) 4 '(53 5 '(64) 6 '(76	Undo ute 6) Di Label 4-29.666667] .666667-41.3333 33333-5 .64.666667] .666667-76.3333 .33333-inf)	Edit 551 1938 1115 457 	
Weka Explorer reprocess Classify Cluster Asson Open file Open URL tter Choose Discretze - B 6 - M - 1.0 - R 1 - pr urrent relation Instances 4119 I	clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name Name	ate Selected attrib Name: age Missing: 0(3) No. 1 1 2 3 4 5 6 7(76)	Undo ute (b) Di Label (-29.666667)* (.666667-41.3333 .33333-53 .33333-53 .645.66667]* .6666667-76.33333 .333333-inf)*	Edit stinct: 6 Count 551 1938 1115 457 457 40 18	
Weka Explorer reprocess Classify Cluster Asso Open III Open URL Iter	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Invert Pattern Name Name	ate Name: age Missing: 005 No. 1 '(-in 2 '(20 3 '(41 4 '(53 5 '(76 6 '(76	Undo ute (5) Di Label 1-29.666671 666667-76.33333 .33333-531 -64.6666677-66.33333 .333333-inf)	stinct: 6 Count 551 1115 457 - 40 18	
Weka Explorer reprocess Classify Cluster Asson Open URL ter Choose Discretize B 6 -M -1.0 -R1 -pn rent relation B and additional weka.filters.uns stances All None All None 1 age 2 job 3 marital 4 clustation 5 deaut 6 housing 7 Gon 8 contact	clate Select attributes Visualize Open DB Genere ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29) 3 '(41 4 '(53) 5 '(54) 6 '(76) Class: y (Nom)	Undo ute Label 1.29 666671 .666667-41.3333 33333-531 -64.6666677 .666667-76.33333 .33333-inf)	Edit	
Weka Explorer reprocess Classify Cluster Assor Upen file Open URL Urrent relation Relation: bank-additional-weka filters and nstances: 4119 Non All None I age 2 lob 3 marital 4 education 5 default 6 lousing 7 loan 8 contact 9 month 1 due of weak	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29) 3 '(41) 4 '(53) 5 '(64) 6 '(76) Class: y (Nom)	Undo Undo Label (-29.666667) -666667-41.3333 33333-3 -64.6666677 .666667-76.33333 .33333-inf)	Edit	- C Save Apply Stop Type: Nominal Unique: 0 (0%) Weight 551 1938 1938 1938 1938 1938 1938 1938 193
Weka Explorer reprocess Classify Cluster Asso Open Ifile Open URL Iter Image Cluster Asso Choose Discretize - B6 - M - 1.0 - R1 - pr Image Cluster All None Relation: bank-additional-weka.filters.uns Image Im	clate Select attributes Visualize Open DB_ Genera ecision 6 aupervised.att_ Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Name: age Missing: 0(0) No. 1 '(-in 2 '(2) 3 '(41 4 '(53) 5 '(64 6 '(76 Class: y (Nom)	Undo ute (c) Di Label (c-29.66667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.6667) (c-29.667) (c-29.667) (c-29.667) (c-29.67) (c-2	Edit	
Weka Explorer Open file Open URL Iter Choose Discretize -B 6 -M - 1.0 - R 1 - pr rent relation Relation: bank-additional-weka.filters.uns Relation: bank-additional-weka.filters.uns All None 1 age 2 ob 3 marital 4 education 5 default 6 housing 7 Gan 8 contact 9 month 10 day.of week 11 duration	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Invert Pattern Name Name	ate Selected attrib Name: age Mismieg: 00° 1 'Cin 2 '220 3 '(41 4 '(53 5 '(64 6 '(76 6 '(76	Undo Label 1-29.666671 666667-11.3333 .33333-531 -64.6666677 .666667-76.33333 .333333-inf)	Edit	
Weka Explorer reprocess Classify Cluster Asso Open file. Open URL. Iter Open URL. Choose Discretize - 86 - M - 1.0 - R1 - pr Iter Iter Iters - 1.0 - R1 - pr Relation: base - 4ditional-weka filters unstances. All None None All None None 1 age 2 lob 3 marital 4 4 education 5 5 deaut 6 6 housing 7 7 Gan 8 9 month 10 10 day of week 11 11 duation 12 12 campaign 13 doks	clate Select attributes Visualize Open DB Genere ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29) 3 '(41 4 '(53) 5 '(66 6 '(76 Class: y (Nom)	Undo ute (b) Di Label 1-29 666677 -646666677 -646666677 -646666677 -646666677 -64697 -64697 -64697 -64697 -64697 -64697 -6497	stinct: 6 551 1938 1115 457 40 18	- C Save Type: Nominal Unique: 0 (0%) Weight 551 1938 1115 457 40 18
Weka Explorer reprocess Classify Cluster Asso Open Ifile Open URL Iter Image Cluster Asso Relation: base definional-weka.filters.unsinstances: 4119 All None Image 2 lob 3 marital 4 education 5 default 5 default Image Image 7 Image Image Image 9 month Image Image 1 age Image Image 2 lob Image Image 1 age Image Image 2 lob Image Image 3 marital Image Image 1 dousing Image Image <tr< td=""><td>clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name Name</td><td>ate Name: age Missing: 0(0) No. 1 '(-in 2 '20') 3 '(41 4 '(53 5 '(64 6 '(76 6 '(76 Class: y (Nom))</td><td>Undo ute (6) Di Label (-29.6666671' .666667-1' .666667-76.33333 .333333-537' .666667-76.33333 .333333-inf)'</td><td>Edit</td><td></td></tr<>	clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name Name	ate Name: age Missing: 0(0) No. 1 '(-in 2 '20') 3 '(41 4 '(53 5 '(64 6 '(76 6 '(76 Class: y (Nom))	Undo ute (6) Di Label (-29.6666671' .666667-1' .666667-76.33333 .333333-537' .666667-76.33333 .333333-inf)'	Edit	
Weka Explorer reprocess Classify Cluster Asso Open IIIa Open URL Iter Iter Choose Discretize -86 - M - 1.0 - R1 - pr Instance: 4110 None Itersum Relation: base additional-weka.filters.uninstance: 411 None No. 1 age 2 job 3 marital 4 education 5 default 6 housing 7 Ioan 8 contact 9 month 10 day of week 11 day of week 11 jodays 12 campaign 13 jodays 13 previous 15 poutcome	clate Select attributes Visualize Open DB Generation ecision 6 aupervised.att Attributes: 21 Sum of weights: 4119 Invert Pattern Name Name	selected attrib Name: age Missing: 000 No. 1 '(-in 2 '(20 3 '(41 4 '(53 5 '(76 6 '(76 Class: y (Nom)	Undo ute 6) Di 1-29.666671 666667-16.33333 .333333-531 -64.6666677-66.33333 .333333-51	stinct: 6 Count 1938 1115 457 - 40 18	
Weka Explorer reprocess Classify Cluster Asso Open file. Open URL Iter Choose Discretize - B 6 - M - 1.0 - R 1 - pr urrent relation Relation: bank additional - weka.filters.un stata.ces. 4119 All None All None 1 age 2 ob 3 marital 4 education 5 default 6 housing 7 loan 8 contact 9 month 10 day.of.week 11 duration 12 contact 9 month 13 opdays 14 previous 15 opottcome 16 emp.var.ate 17 cons.price/dx	clate Select attributes Visualize Open DB Genera eccision 6 aupervised.att Attributes: 21 Invert Pattern Name Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(20) 3 '(41 4 '(53) 5 '(54 6 '(76 6 '(76 Class: y (Nom)	Undo ute (c) Di (-29 66667) (-666667-41.3333 .33333-53) -64.666667.7 .666667-7 .666667-7 .666667-7 .666667-7 .666667-7 .53333 .33333-inf) 1115 1115 1115 1115 1115	Edit	
Weka Explorer reprocess Classify Cluster Asso Open file Open URL Image: Cluster Asso Choose Discretize - 86 - M - 1.0 - R1 - pr Image: Cluster Asso Relation: Image: Cluster Asso Image: Cluster Image: Cluster All None Image: Cluster	clate Select attributes Visualize Open DB Genera ecision 6 supervised.att Attributes: 21 Invert Pattern Name Name	ate Selected attrib Name: age Missing: 0 (05 No. 1 '(-in 2 '(29 3 '(41 4 '(53) 5 '(54) 6 '(76) Class: y (Nom)	Undo ute (a) Di Label (-29.666667) -64.666677 -64.666677 -64.6666677 .666667-76.33333 .33333-inf)	stinct: 6 551 1938 1115 457 40 18	
Weka Explorer reprocess Classify Cluster Asso Open IIIa Open URL Iter Image Cluster Asso Choose Discretize - 86 - M - 1.0 - R1 - pr Image Cluster All None Relation: hark additional-weka.filters.undistances: 4119 None Image	clate Select attributes Visualize Open DB Genera ecision 6 aupervised.att Attributes: 21 Invert Pattern Name Name	ate Selected attrib Name: age Missing: 0(0) No. 1 '(-in 2 '2-9' 3 '41 4 '53 5 '64 6 '(76 Class: y (Nom)) Class: y (Nom)	Undo ute (-29.66667] -6.666677 -6.4666667] .666667-76.33333 .333333-51] -1115	stinct: 6 551 - 1938 1115 457 - 40 18	- C Save Apply Stor Type: Nominal Unique: 0 (0%) Weight 551 1938 1115 457 40 18

The **filters.unsupervised.attribute.Normalize** and **filters.unsupervised.attribute.Standardize** tools cannot separate the normalization and standardization for different variables, so the **MathExpression** tool is recommended for Q4 and Q5. (the **PartitionedMultifilter** is also ok)

Open file Open URL	Open DB	Generate	Undo	Edit	Save
ilter]][
Choose MathExpression -E (A-MIN)/(MAX	-MIN) -V -R 11				Apply Stop
urrent relation		Selected attrit	oute		
Relation: bank-additional-weka.filters.unsupe	rviced att Attribute	e 21 Name dur	ation		Type: Numeric
Instances: 4119	weka.gui.GenericObje	ctEditor	×	ct: 828	Unique: 282 (7%)
Attributes	weka.filters.unsupervised.a	ttribute.MathExpression			Value
All None	About			0	
Ne	Modify numeric attributes	according to a given	More	3643	
	mathematical expression.		Capabilities	256.766	
4 education				Lonior	
5 default	debug	False	~		
7 loan	deNetCheckCapabilities	Falsa			
8 🗌 contact	uoivoteneekeapabiiities	raise	Ť		
9 🗌 month	expression	(A-MIN)/(MAX-MIN)			
10 day_of_week	ignoreClass	False	~		 Visualize /
11 duration					
13 pdays	ignoreRange	11			
14 previous	invertSelection	True	~		
15 🗌 poutcome					
16 emp.var.rate	Open Sa	ve OK	Cancel		
17 cons.price.idx					
19 euribor3m					
20 nr.employed					
21 🗌 y					
Remove					
				1921.6	~

or

0,12-20 Attribut ui.GenericObj unsupervised. meric attribute: tical expression debug	Acc 21 Selected attr Name du lectEditor attribute.MathExpression a according to a given h. False	ibute rration × More Capabilities	ct: 828	0 3643 256.788 254 704	Apply Type: Num Unique: 282 I Value	eric (7%)
),12-20 Attribut jui.GenericObj .unsupervised. imeric attribute tical expression debug	Selected attr lectEditor attribute.MathExpression s according to a given h.	ibute rration X More Capabilities	ct: 828	0 3643 256.788 254.704	Apply Type: Num Unique: 282 (Value	eric (7%)
Attribut jui.GenericObj unsupervised. imeric attribute: tical expression debug	Selected attr lectEditor attribute.MathExpression s according to a given h.	ibute irration × More Capabilities	ct: 828	0 3643 256.788 254.704	Type: Num Unique: 282 (Value	neric (7%)
Attribut gui.GenericObj unsupervised. imeric attribute: tical expression debug	tec 21 Name du jectEditor attribute.MathExpression s according to a given h.	More Capabilities	ct: 828	0 3643 256.788 254.704	Type: Num Unique: 282 (Value	neric (7%)
unsupervised. Imeric attribute: tical expression debug	attribute.MathExpression s according to a given h. False	More Capabilities		0 3643 256.788 254.704	Value	(***)
imeric attribute: tical expression debug	s according to a given	More Capabilities		0 3643 256.788 254 704	value	
imeric attributer tical expression debug	s according to a given n. False	More Capabilities		0 3643 256.788 254.704		
umeric attribute tical expression debug	s according to a given	More Capabilities		256.788 254.704		
tical expression debug	1. False	Capabilities		254 704		
debug	False					
debug	False		_			
LC		~				
debug False doNotCheckCapabilities False						
ckCapabilities	raise	~				
expression	(A-MIN)/(MAX-MIN)					
						Vicualize
ignoreclass	raise	~			-]	VISGUIL
ignoreRange	1-10, 12-20					
	E-L-					
ivertSelection	False	~				
	Gave OK	Cancel				
	ignoreClass ignoreRange nvertSelection	ignoreClass False ignoreRange 1-10, 12-20 wertSelection False Save OK	ignoreClass False v ignoreRange 1-10, 12-20 wertSelection False v Save OK Cancel	ignoreClass False ~ ignoreRange 1-10, 12-20 wertSelection False ~ Save OK Cancel	ignoreClass False v ignoreRange 1-10, 12-20 wertSelection False v	ignoreClass False v ignoreRange 1-10, 12-20 wertSelection False v . Save OK Cancel

4.

open me	Open URL	Ope	n DB Gen	erate Un	do	Edit		Save
ilter								
Choose Mathl	xpression -E (A-MIN)/(M	IAX-MIN) -R 1-10	,12-20				Apply	Stop
urrent relation Relation: bank-a Instances: 4119	dditional-weka.filters.uns	upervised.att	Attributes: 21 Sum of weights: 4119	Selected attribute Name: duration Missing: 0 (0%)	Distinct: 828		Type: Num Unique: 282 (eric 7%)
ttributes				Statist	ic		Value	
All	None	Invert	Pattern	Minimum Maximum		D 1		
No. 1 age 2 job 3 marit 4 educa 5 defau 6 housi 7 loan	al tion lit ng	Name		Mean StdDev		0.07 0.07		
8 _ conta 9 _ mont 10 _ day_c	ct h of_week			Class: y (Nom)			~	Visualize
11 durat 12 camp 13 pday 14 previa 15 pouto 16 emp. 17 cons.	ion aign s sous come var.rate price.idx							

5.

Open fi	le	Open URL		Open DB		Gener	ate	Un	do	Edit		Save
ilter												
Choose M	athExpressio	n -E (A-MEAN)/((SD) -	R 11							Apply	/ Stop
urrent relatio	n						Selected attrib	ute				
Relation: ba	nk-additional	-weka.filters.uns	uperv	iseri att 👩 weka qui Ger	Attribute nericΩbie	e: 21	Name: dura	tion	×	ct: 828	Type: Nur Unique: 282	neric (7%)
ttributes	15			- cu					~	ct. 020	onique. 202	(770)
				weka.tiiters.unsup	pervised.a	ttribute.Ma	athExpression			0	Value	
All		None	_	About						1		
No.			Nar	Modify numeric	attributes	according to	o a given	N	lore	0.07		
4	ducation			mathematical ex	pression.			Capa	abilities	0.07		
4 _ e	lefault											
6 🗌 h	ousing				debug	False			~			
7 🗌 le	ban			doNotCheckCap	abilities	False			~			
8 🗌 c	ontact											
9 🗌 n	nonth			exp	pression	(A-MEAN	I)/(SD)					
11 0	lay_of_week			ign	oreClass	False			~		~	Visualize A
12 0	ampaign			Immo	Denne	11						
13 🗌 p	idays			ignor	enange							
14 🗌 p	revious			invertS	election	False			~			
15 🗌 p	outcome											
16 🗌 e	mp.var.rate			Open	Sa	ive	OK		Cancel]]		
18 🗆 c	ons.conf.idx											
19 🗌 e	uribor3m											
20 🗌 n	r.employed											
21 🗌 y												

Reference link:

https://weka.sourceforge.io/doc.dev/weka/filters/unsupervised/attribute/MathExpression.html

ilter Choose MathExpression - E (A-MEAN)/(SD) - R:11 urrent relation Relation: bank-additional-wekafilters.unsupervised.att Attributes: 21 Sum of weights: 4119 Selected attribute Selected attribute None Invert Pattern No. Contact Conta	Open file	Open URL.	Open	DB Ge	nerate	Undo	Edit.	-	Save
Apply Stop Choose MathExpression - E (A-MEAN)/(SD) - R.11 Apply Stop urrent relation Sum of weights: 4119 None Type: Numeric Unique: 0 (0%) Instances: 4119 Sum of weights: 4119 Name Maising: 0 (0%) Distinct: 7 Unique: 0 (0%) All None Invert Pattern Minimum -0.351 Maximum 10.723 No.	Filter								
urrent relation Relation bank-additional-wekafiltersunsupervised.att Attributes 21 Sum of weights 2119 All None Invert Pattern All None Invert Pattern No. Name Ad education A educatio	Choose MathExp	ression -E (A-MEAN)/(SD) -R 11					Apply	Stop
Relations: Attributes: 21 Name: Distinct: Type: Numeric Instances: Sum of weights: 4119 Sum of weights: 4119 Distinct: 7 Unique: 0 (0%) All None Invert Pattern Main: -0.351 Maximum 10.723 No.	Current relation				Selected attribut	e			
Statistic Value All None Invert Pattern No. Invert Pattern Maximum -0.351 Maximum 10.723 Maximum 10.723 Maximum 0 StdDev 1 S default G StdDev 1 G housing StdDev 1 StdDev 1 G housing StdDev StdDev 1 StdDev 11 duration StdDev StdDev Visualize A 12 genpaign StdDev StdDev StdDev 13 pdays StdDev StdDev StdDev StdDev 14 previous StdDev StdDev StdDev StdDev 15	Relation: bank-add Instances: 4119	tional-weka.filters.ur	nsupervised.att	Attributes: 21 Sum of weights: 4119	Name: previo Missing: 0 (0%)	us Dis	tinct: 7	Type: Num Unique: 0 (0%	eric 6)
All None invert Pattern No. Name 0 4 education 0 5 default 0 6 housing 1 7 loan StdDev 1 8 contact	Attributes					Statistic		Value	
No. Name Mean 0 4 education 5 default 5 5 default 6 housing 1 6 housing 7 loan 2 7 loan 2 loanyaign 1 10 day of week 1 Lass: y (Nom) Visualize A 11 duration 2 loanyaign Visualize A 13 pdays 1 Visualize A 16 emp var.rate 1 Visualize A 17 consprice/dx 1 Visualize A 18 consond/dx 1 V 19 eurlor3m 20 n.employed 21 y V V	All	None	Invert	Pattern	Minimum		-0.351		
1	No.		Name		Mean		0		
11 duration 12 campaign 13 pdays 14 previous 15 poutcome 16 emp var.rate 17 cons.price.idx 18 cons.conf.idx 19 euribor3m 20 n.employed 21 y	6 housing 7 loan 8 contact 9 month 10 day_of_v	veek			Class: y (Nom)			~	Visualize A
15 poutcome 16 empvar.rate 17 cons.price.idx 18 cons.conf.idx 19 euribor3m 20 n.employed 21 v	11 duration 12 campaig 13 pdays 14 previous	n							
10 construction 11 consconfidx 19 euribor3m 20 nr.employed 21 v	15 poutcom	ie rate							
19cuistoniata 19euribor3m 20nr.employed 21v	17 _ cons.pric	e.idx							
20 nremployed	10 Cons.con	m							
	20 _ nr.emplo 21 _ v	yed							

6.

Weka Explorer Preprocess Classify Cluster Assoc	iate Select attributes V	/isualize			– – ×
Open file Open URL	Open DB	Generate	Undo	Edit	Save
Filter					
Choose NominalToBinary -R 2,4					Apply Stop
Current relation		Selected attrib	ute	_	
Relation: bank-additional-weka.filters.uns Instances: 4119	uperv 🥥 weka.gui.GenericO	bjectEditor	×	nct: 6	Type: Nominal Unique: 0 (0%)
Attributes	wereannerstansupervise	d.attribute.rominarrobinary		Count	Weight
All None	About			551	551
	Converts all nominal a	ttributes into binary numeric	More	1938	1938
No.	Nar attributes.		Capabilities	1115	1115
1 age 2 job=blue-collar	attributeIndic	ces 2,4		457	457 40
3 job=services	hinan/AttributorNomi	Falco	~	18	18
4 job=admin. 5 job=entrepreneur 6 job=self-employed	deb	ug False	~]	
7 job=technician	doNotCheckCapabilit	ies False	~]	Visualize All
9 iob=student	invertSelecti	on False	~		
10 job=retired 11 job=housemaid	spreadAttributeWeig	pht False	~]	
12 iob=unemployed 13 iob=unknown	transformAllValu	False	~]	
14 🗌 marital	Open	Save OK	Cancel		
15 ducation=basic.9y 16 ducation=high.school			cuncer		
17 ducation=university.degree					
18 education=professional.course		551		457	
Remo	ove				40
Status OK					Log 💉 x 0

Weka Explorer	– D X
Open file Open URL Open DB Gen	erate Undo Edit Save
Filter	
Choose NominalToBinary -R 2,4	Apply Stop
Current relation	Selected attribute
Relation: bank-additional-weka.filters.unsupervised.att Attributes: 39 Instances: 4119 Sum of weights: 4119	Name: job=blue-collar Type: Numeric Missing: 0 (0%) Distinct: 2 Unique: 0 (0%)
Attributes	Statistic Value
All None Invert Pattern	Minimum 0
	Maximum 1
No. Name	Mean 0.215
$2 \square$ iob-blue-collar	500EV 0.411
3 job=services	
4 job=admin.	
5 🔄 job=entrepreneur	
6 job=self-employed	
$7 \square Job=technician$ 8 \square job=management	Class: y (Nom) Visualize All
9 job=student	
10 job=retired	3235
11 🗌 job=housemaid	
12 job=unemployed	
13 job=unknown 14 - marital	
15 education=basic.9v	
16 ducation=high.school	
17 ducation=university.degree	884
18 education=professional.course	
Remove	
	n ns 1
Status OK	Log x0

7.

Correlation:



Top 10: duration, nr.employed, pdays, euribor3m, emp.var.rate, previous, poutcome, contact, cons.price.idx, job=retired

InfoGain:

epiecess classify claster resource	Serect attended			
tribute Evaluator				
Choose InfoGainAttributeEval				
arch Method				
Choose Ranker -1 -1.79/6931348623157E	308 -N -1			
tribute Selection Mode	Attribute selecti	on output		
Use full training set				
Cross-validation Folds 10	Search Metho	d:		
Seed 1	Atti	ibute ranking.		
	Attribute Ev	aluator (supervised, Class (nominal): 39 y):		
o class v	Info	rmation Gain Ranking Filter		
	Deeled seeud			
Start Stop	0.1096656	29 duration		
sult list (right-click for options)	0.0951712	36 cons.conf.idx		
5:37:50 - Ranker + CorrelationAttributeEval	0.0874463	37 euribor3m		
5:39:47 - Ranker + InfoGainAttributeEval	0.0873909	38 nr.employed		
	0.0860484	35 cons.price.idx		
	0.0736447	34 emp.var.rate		
	0.0469643	33 poutcome		
	0.0355763	22 month		
	0.031243	32 previous		
	0.0151493	26 contact		
	0.0096277	1 age		
	0.0058546	30 campaign		
	0.0048049	23 default		
	0.003644	2 job=blue-collar		
	0.0017762	10 job=retired		
	0.0017391	14 marital		
	0.001566	15 education=basic.9y		
	0.0014107	17 education=university.degree		
	0.001118	4 job-admin.		
	0.0010139	5 job=entrepreneur		
	0.0002066	25 loan		
	0.00001175	24 housing		
	0	3 job=services		
	0	7 job=technician		
	0	8 job=management		
	0	6 job=self-employed		
	0	13 job=unknown		
	0	11 job=housemaid		
	0	12 job=unemployed		
	0	21 education-unknown		
	0	20 education=basic.4y		
	0	18 education-professional.course		
	0	16 education=high.school		
	0	19 education=basic.6y		
	Selected att	ributes: 29,36,37,38,35,34,33,31,27,32,26,1,30,23,2,10,9,14,15,17,4,5,25,24,28,3,7,8,6,13,11,	,12,22	2

Top 10: duration, cons.conf.idx, euribor3m, nr.employed, cons.price.idx, emp.var.rate, poutcome, pdays, month, previous

Union: duration, nr.employed, pdays, euribor3m, emp.var.rate, previous, poutcome, contact, cons.price.idx, job=retired, cons.conf.idx, month, (age)

8-9:

Bank-test-set.csv



Bank-train-set.csv:

Set "invertSelection" to True

Open file	Open URL	Open DB	Gene	erate	Undo		Edit_	Save
ilter	380 P 200 V							Annha
Choose Kenovereiten	age -r 00.0 - v						l	Apply
Relation: bank-additional-	weka.filters.unsupervised.att	ribute.DiscrAttr	ributes: 14	Name: age	te		Type: N	lominal
Instances: 3295		Sum of w	reights: 3295	Missing: 0 (0%		sunct: 6	Unique: U	(0%)
ittioutes				No.	Label	0	ount	Weight
All	None	Invert Pa	attern	2 '(29.0	566667-41.3333333]	1561	455	1
No.	Nam	e		3 '(41.	333333-53]"	884	884	
1 age				4 '(53-	64.666667]"	355	355	
2 job=retired				5 '(64.6	5666667-76.3333333]"	28	28	
3 contact				0 (/0	555555-Int)	14	14	
5 duration			011.15.01					
6 🗌 pdays		Weka.gui.Generi	cobjectEditor		~			
7 previous		weka.filters.unsuperv	ised.instance.Re	emovePercentage				
8 poutcome		About						
10 cons.price.idx		A filter that removes	a given percent	age of a dataset.	More			
11 _ cons.conf.idx					Capabilities			
12 euribor3m								
13 nr.employed		de	bug False		~			
14 🛄 Y		doNotCheckCapabil	lities False		~			
		invertSeler	tion True		~			Visualize
		interescie	The state					
		percent	tage 80.0					
		0.00	6.00	01	Const.			
		Open	Save	OK	Cancel			
					204			
				453				
	Remove							
	Remove						28	14
tatus	Remove					_	28	

2) Modelling and deployment

10-11:

The classification result on training set:

Preprocess Classify Cluster	Associate Select attrib	outes Vis	ualize								
lassifier											
Choose J48 -C 0.25 -M 2											
est options	Classifier output										
Use training set	SIZA OL CHA CLA	e: 5	3								
Supplied test set Set											
Cross-validation Folds 10	Time taken to b	uild model	: 0.06 se	conds							
Percentage split % 66	=== Evaluation	on trainin	g set ===								
More options											
	Time taken to t	est model	on traini	ing data: 0.	01 secon	ds					
(Nom) y	✓ === Summary ===										
Start Stop											
Result list (right-click for options)	Incorrectly Class	ssified Inst	ances stances	230		6,9803	8				
20:57:30 - trees.J48	Kappa statistic			0.58	3						
	Mean absolute e	rror		0.11	91						
	Root mean squar	ed error		0.24	4						
	Relative absolu	te error		59.54	8						
	Root relative s	quared err	or	77.19	73 %						
	Total Number of	Instances		3295							
	=== Detailed Ac	curacy By	Class ===								
		TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
		0.984	0.496	0.940	0.984	0.962	0.602	0.822	0.958	no	
		0.504	0.016	0.803	0.504	0.619	0.602	0.822	0.584	yes	
	Weighted Avg.	0.930	0.442	0.924	0.930	0.923	0.602	0.822	0.916		
	=== Confusion M	atrix ===									
	a b <-	- classifi	ed as								
	2878 46	a = no									
	184 187	b = yes									
tatue											
OK									Log	-	A

The classification result on testing set:

🥥 Weka Explorer									-		×
Preprocess Classify Cluster A Classifier	ssociate Select attribu	utes Vis	ualize								
Choose J48 -C 0.25 -M 2											
Test options	Classifier output										
Use training set Supplied test set Set	Time taken to bu	uild model	: 0.02 se	conds							
Cross-validation Folds 10 Percentage split % 66	=== Evaluation o	n test se	t ===								
More options	Time taken to te	st model	on suppli	ed test set	: 0.01 s	econds					
(Nom) y	✓ === Summary ===										
Start Stop	Correctly Classi	fied Inst	ances	744		90.2913	8				
esult list (right-click for options) 10:57:30 - trees.J48 10:59:35 - misc.InputMappedClassifier	Kappa statistic Mean absolute er Root mean square Relative absolut Root relative sq Total Number of === Detailed Acc	ror ed error uared err Instances	Class ===	0.37 0.14 0.28 76.59 97.78 824	67 4 99 53 % 48 %	511007					
	Weighted Avg. Confusion Ma a b < 714 30 a = 50 30 b =	TP Rate 0.960 0.375 0.903 ttrix === classified no yes	FP Rate 0.625 0.040 0.568	Precision 0.935 0.500 0.892	Recall 0.960 0.375 0.903	F-Measure 0.947 0.429 0.897	MCC 0.381 0.381 0.381	ROC Area 0.714 0.714 0.714	PRC Area 0.945 0.287 0.881	Class no yes	
tatus DK									Log		~

12-13:

lassifier							
Choose J48 -C 0.25 -M 20							
est options	Classifier o	utput					
Use training set	Size of	the tree : 53					
Supplied test set Set		-					
Cross-validation Folds 10	Attribut	weka.gui.GenericObjectEditor		×			
Percentage split % 66	Model at	weka.classifiers.trees.J48					
More options		About					
, interesting the second	(nominal	Class for concepting a prupod or up	an and Cd	Maria			
Nom) y 🗸 🗸	(nominal	Class for generating a pruned or un	sruned C4.	More			
Start Stan	(nominal			Capabilities			
Start	(numeric	hatchfine	100				
asun iist (right-click for options)	(numeric	oatchsize	100				
16:42:54 - misc.inputMappedClassifier	(nominal	binarySplits	False	~)			
	(numeric	collapseTree	True	~			
	(numeric	compactive					
	(numeric	confidenceFactor	0.25				
	(numeric	debug	False	~			
		dellar the state of the state	Falsa				
		doNotCheckCapabilities	False	~			
	Time tak	do Not Make Split Point Actual Value	False	~			
	Bval	minNumOhi	20				
	Time tak						
		numDecimalPlaces	2				
	Sunn	numFolds	3				
	Correctl						
	Incorrec	reducedErrorPruning	False	Y			
	Kappa st	saveInstanceData	False	~			
	Root nea	rees	1				
	Relative	seed	•				
	Root rel	subtreeRaising	True	~			
	LOCAL NO	unpruned	False	~			
	Deta	arthrende					
		useLaplace	False	~	ROC Area	PRC Area	Class
		useMDLcorrection	True	~	0.631	0.944	no
	Madada				0.631	0.201	yes
	weighted	Open Save	ОК	Cancel	0.031	0.009	
	Conf	USION MACTIX		<u>//</u>			
		d allocated as					
	97 2 1	a = no					
	6 2 1	b = yes					

The classification result on training set:

assifier													
Choose J48 -C	0.25 -M 4	20											
st options			Classifier output										
Use training set			2156 OL CHA CIA	9: I	T								
Supplied test set		Set											
Cross-validation	Folds	10	Time taken to b	uild model	: 0.01 se	conds							
Percentage split	%	66	=== Evaluation	on trainin	a set ===								
More o	ptions				-								
			Time taken to t	est model	on traini	ng data: 0	seconds						
lom) y			Summary										
Start		Stop											
sult list (right-click	for ontic	() ()	Correctly Class	ified Inst	ances	3013		91.4416	8				
:57:30 - trees 148	ioi opin	/////	Kappa statistic	SOLLIOU IN	scances	0.54	38	0.5504					
0:59:35 - misc Innu	tManner	Classifier	Mean absolute e	rror		0.13	36						
1:01:38 - trees 148	imappee	relassiner	Root mean squar	ed error		0.25	84						
			Relative absolu	te error		66.77	22 %						
			Root relative s	quared err	or	81.75	15 %						
			Total Number of	Instances		3295							
			=== Detailed Ac	curacy By	Class ===								
				TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
				0.961	0.450	0.944	0.961	0.952	0.546	0.815	0.956	no	
				0.550	0.039	0.639	0.550	0.591	0.546	0.815	0.488	yes	
			Weighted Avg.	0.914	0.404	0.910	0.914	0.912	0.546	0.815	0.904		
			Confusion M	atrix									
			a b <-	- classifi	ed as								
			2809 115	a = no									
			167 204	b = yes									

The classification result on testing set:

Weka Explorer								-		×
Preprocess Classify Cluster A	ssociate Select attributes	Visualize								
Classifier										
Choose J48 -C 0.25 -M 20										
Test options	Classifier output									
 Use training set 										
Supplied test set Set	Time taken to build me	odel: 0.02 s	econds							
Ocross-validation Folds 10										
O Percentage split % 66	=== Evaluation on test	t set ===								
More options	Time taken to test mod	iel on suppl	ied test set	: 0.01 s	econds					
(Nom) y	<pre>v === Summary ===</pre>									
Start Stop	Correctly Classified	Instances	733		88.9563	8				
Result list (right-click for options)	Incorrectly Classified	d Instances	91		11.0437	8				
20:57:30 - trees.J48	Mean absolute error		0.35	72						
20:59:35 - misc.InputMappedClassifier	Root mean squared erro	or	0.27	96						
21:01:38 - trees.J48	Relative absolute erro	or	78.28	175 %						
21:02:11 - Misc.InputMappedClassifier	Total Number of Instan	lces	824	JZ 8						
	=== Detailed Accuracy	By Class ==	-							
	TP R	ate FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
	0.94	L 0.588	0.937	0.941	0.939	0.359	0.725	0.946	no	
	Weighted Avg. 0.89	0.536	0.429	0.890	0.420	0.359	0.725	0.880	Yes	
	Confusion Matrix -									
	Confusion Matrix -									
	a b < classin	fied as								
	47 33 b = yes									
Status										
ОК								Log	-40	× x (

14:

In this case, model with minimum number of instances per leaf 2 has higher accuracy. So save this model.

Weka Explorer										-		\times
Preprocess Class	sify Cluster Asso	ociate Select attr	ibutes Vis	ualize								
Classifier												
Choose J48 -C	0.25 -M 20											
Test options		Classifier output										
Use training set		Size of the tr	ee: :	00								
Ose training set												
Supplied test set	set	Time taken to	build model	• 0.06 ec	conde							
 Cross-validation 	Folds 10	TIME CURCH CO	burra moder	. 0.00 50	condo							
 Percentage split 	A classifier errors A classif	=== Evaluation	on trainir	ng set ===								
More o	ptions	mine token to	teat model	on traini	na data. O	01	i.e.					
		Time taken to	test model	on traini	ng data: 0.	.01 second	15					
(Nom) y	~	=== Summary ==	=									
Start	Stop				0.045							
Result list (right-click	(for options)	Incorrectly Clas	assified Inst	stances	230		6,9803	8				
19:12:32 - trees ***	ss-validation Folds 10 centage split % 66 More options Time tai Time tai		ç		0.58	33						
19:12:58 - misc.	View in main window		error		0.11	191						
19:13:11 - trees	View in separate wind	low	red error		0.24	14						
19:13:31 - misc.	Save result buffer		squared err	or	77.19	973 %						
	Delete result buffer(s)		f Instances		3295							
	Load model		ccuracy By	Class ===								
	Save model											
	Re-evaluate model on	o current test set	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class	
	Re-apply this model's	configuration	0.504	0.498	0.803	0.504	0.619	0.602	0.822	0.584	yes	
	Visualize classifier erro	ors	0.930	0.442	0.924	0.930	0.923	0.602	0.822	0.916		
	Visualize tree		Aatrix ===									
	Visualize margin curve	e										
	Visualize threshold cu	rve >	- classifi a = no	ed as								
	Cost/Benefit analysis	>	b = yes									
	Visualize cost curve	>										
L	ŋ	10										
Status										Log		- ×0
OK										LUG	100	r **

Age grouping (discretization):

Filter.unsupervised.attribute.MathExpression, enter the expression:

ifelse(A>29.666667,ifelse(A>41.333333,ifelse(A>53,ifelse(A>64.66666,ifelse(A>76.333333,6,5),4),3),2),1)



NumericToNominal:

Open	file Open UR	L Ope	n DB	Gener	rate	Undo	Edit	Save	<u>.</u>
lter									
Choose	NumericToNominal -R 1							Apply	Sto
urrent relat	ion				Selected attrib	bute			
Relation: b nstances: 2	oank-new-clients-weka.filter 2000	s.unsupervised.at	Attribute Sum of weight	s: 21 s: 2000	Name: age Missing: 0 (0	0%) Dis	tinct: 6	Type: Numeric Unique: 0 (0%)	
ttributes	🙆 weka gui GenericObie	ctEditor						×	
All	• weka.gai.Geneneooje	citation						~	
7.01	weka.filters.unsupervised.a	ttribute.NumericToN	lominal						
No.	About								
1	A filter for turning numeric	attributes into nomina	l ones.					More	
2 🗌								Capabilities	
3									
4	attributeIndices	1							
6									
7	debug	False							
8	doNotCheckCapabilities	False						Visi	ualiz
9 🗌									
10 🗌	invertselection	False						~	
11 🗔									
12 🗌	Open		Save			ОК	Cance	2	
13	pdays								
14	previous					469			
16	emp var rate								
17 🗆	cons.price.idx				342				
18 🗌	cons.conf.idx						212	1	
14								53	

15:

	Open file	Open URL	Of	pen DB	Gene	erate		Undo		Edit	!	Save
ilte:	r Numer	icToNominal -R 1									Apply	Sto
Curn Re	ent relation elation: bank-ne tances: 2000	w-clients-weka.filters.	.unsupervised	Attri Sum of we	butes: 21 eights: 2000	Selected a Name: Missing:	attribute age 0 (0%)	2	Distinct: 6	ι	Type: Nomir Jnique: 0 (0%)	nal
ttri	ibutes					No.		Label		Count	v	Neight
	All	None	Invert	Pi	attern	1	1 2		342 891		342 891	
IN	1 200		Name			3	3		469		469	
	2 ioh					5	5		53		53	
	3 marital 4educat 5default 6housin 7loan	ion : g				6	6		33		33	
	8 contac	t				Class: y (S	tr)				~	Visualize
	10 day_of	week					89	1				
	12 campa	ign										
	14 previor	JS										
	15 Doutco	me							469			
	16 🗌 emp.va	ar.rate										
	17 cons.p	rice.idx onf.idx				342				212		

Rename Attribute:

1: (-inf-29.666667], 2: (29.6666667-41.333333], 3: (41.333333-53], 4: (53-64.6666667], 5: (64.6666667-76.333333], 6: (76.333333-inf)

Open file	Open URL	Open DB		Generate		Undo	Ec	lit	Save
er									
hoose Rena	meNominalValues -R 1 -N	*1: (-inf-29.6666667],	2: (29.666667	7-41.333333], 3: (41.3	33333-53], 4: (53-64.66	6667], 5: (64.6	566667-76.3 A	pply S
rent relation Relation: bank- stances: 2000	new-clients-weka.filters.uns	supervised Sur	Attributes: n of weights:	21 Selecte 2000 Missin	datt ne: ag ig: 0	ribute ge (0%) Dist	tinct: 6	Type: N Unique: 0	Nominal) (0%)
ributes				No.		Label	Co	unt	Weight
All	None	Invert	Pattern		1 (* 2 (*	-inf-29.666667] 29.666667-41.333333]	342 891	342 891	
No.		Name			3 (4	41.333333-53]	469	469	
1 🗌 age					4 (53-64.666667]	212	212	
4 educ 5 defa 6 hous 7 loan	ation ult ing								
8 cont 9 mon 10 day_ 11 dura 12 camp 13 pday 14 prev 15 pout	act th of_week baign s ous come			Class: y	(Nor	n) 591			∨ Visuali
16 emp 17 cons 18 cons	var.rate price.idx .conf.idx	ve		342		469	212		

Dummy variable creation for "job" and "education":

No.	Name
1	age
2	job=technician
3	job=blue-collar
4]job=management
5	job=admin.
6	job=services
7	job=housemaid
8	job=self-employed
9	job=unknown
10	job=unemployed
11 🗌	job=entrepreneur
12 🗌	job=retired
13 🗌	job=student
14	marital
15	education=professional.course
16	education=basic.4y
17 🗌	education=high.school
18 🗌	education=university.degree
19 🗌	education=basic.9y
20	education=unknown
21 🗌	education=basic.6y
22	education=illiterate
23	default

Variable selection:

Chose those attributes selected in phase 1 Q7.

Open file Open URL	Open DB	Gen	erate	Undo	Edit		Save
ter							
Choose NominalToBinary -R 2,4						Apply	Stop
irrent relation Relation: bank-new-clients-weka.filters.uns nstances: 2000	upervised Attri Sum of we	ibutes: 14 eights: 2000	Selected a Name: Missing:	attribute age 0 (0%) Di	stinct: 6	Type: Nomi Unique: 0 (0%	nal)
tributes			No.	Label	Count	1	Weight
All None	Invert Pa Name	attern	1 2 3	(-inf-29.666667] (29.666667-41.333333) (41.333333-53] (53-64.666667)	342] 891 469 212	342 891 469 212	
2Job=retired 3contact 4month 5duration 6pdays 7previous			6	(04.00007-76.33333 (76.333333-inf)	33	33	
8 poutcome 9 emp.var.rate 10 cons.price.idx 11 cons.conf.idx 12 euribor3m 13 nr.employed 14 y			Class: y (N	991 469		~	Visualize A
Remov	re		342		212	53	33

Duration normalization:

Expression applied: (A-0)/(3643-0)

The point is that the statistics should remain identical with that applied in bank-additional.csv

	Open URL	Open	DB Gene	Senerate Undo Edit Save					
Filter									
Choose Math	Expression -E (A-0)/(36	43-0) -V -R 5				App	oly Stop		
Current relation Relation: bank-r Instances: 2000	new-clients-weka.filters.	unsupervised	Attributes: 14 Sum of weights: 2000	Selected attribute Name: duration Missing: 0 (0%)	Distinct: 839	Type: Nu Unique: 375	meric 5 (19%)		
Attributes				Statis	tic	Value			
All	None	Invert	Pattern	Minimum	C	.001			
				Maximum	C	.997			
No.		Name		Mean	C	.106			
6 pday: 7 previ 8 poute	s ous come			Class: y (Nom)			Visualize		
9 emp. 10 cons.	var.rate price.idx			_					
11 🗌 cons.	conf.idx								
12 🗌 eurib	or3m								
13 🔄 nr.em 14 🗌 y	pioyed								
				96					

Standardize other numeric values:

Take the "pdays" as an example, the applied math expression should be:

(A-960.422)/191.923, the mean and std values are chosen from original bank additional dataset in phase 1.

Open fi	e	Open URL	Open DB	Gen	erate	Undo	Edit		Save
lter									
Choose M	athExpress	ion -E (A-960.422)/1	91.923 -V -R 6					Apply	Stop
urrent relatio Relation: ba Instances: 20	n nk-new-clie 00	nts-weka.filters.unsu	pervised Sum c	Attributes: 14 of weights: 2000	Selected attribut Name: pdays Missing: 0 (0%)	e Distinc	t: 22	Type: Num Unique: 5 (0%	eric 6)
ttributes						Statistic		Value	
All No.		None	Invert	Pattern	Minimum Maximum Mean		-5.004 0.201 -0.433		
4 _ n 5 _ d 6 _ p 7 _ p	nonth Juration days revious								- 10 P
8	outcome mp.var.rate ons.price.idx ons.conf.idx uribor3m r.employed	¢.			Class: y (Str)			~	Visualize .
					245				

Apply StringToNominal to target variable "y" (Optional):



Save it in .arff format.

16:

Process shown in the specification.