



Vidyavardhini's College of Engineering & Technology

Department of Computer Engineering

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer

Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1

Cluster mode

☐ Use training set

☐ Supplied test set Set...

☒ Percentage split % 66

☐ Classes to clusters evaluation (Num) Spending Score (1-100) v

☒ Store clusters for visualization

Ignore attributes

Start Stop

Result list (right-click for options)

04:15:06 - SimpleKMeans

04:49:24 - SimpleKMeans

Clusterer output

Number of iterations: 6

Within cluster sum of squared errors: 34.04028922797856

Initial starting points (random):

Cluster 0: 55,Female,50,43,45

Cluster 1: 64,Female,54,47,59

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data (132.0)	Cluster# 0 (77.0)	1 (55.0)
CustomerID	104.1439	97.9091	112.8727
Genre	Female	Female	Male
Age	38.7955	38.9481	38.5818
Annual Income (k\$)	62.0455	59.7273	65.2909
Spending Score (1-100)	50.5152	50.8701	50.0182

Time taken to build model (percentage split) : 0 seconds

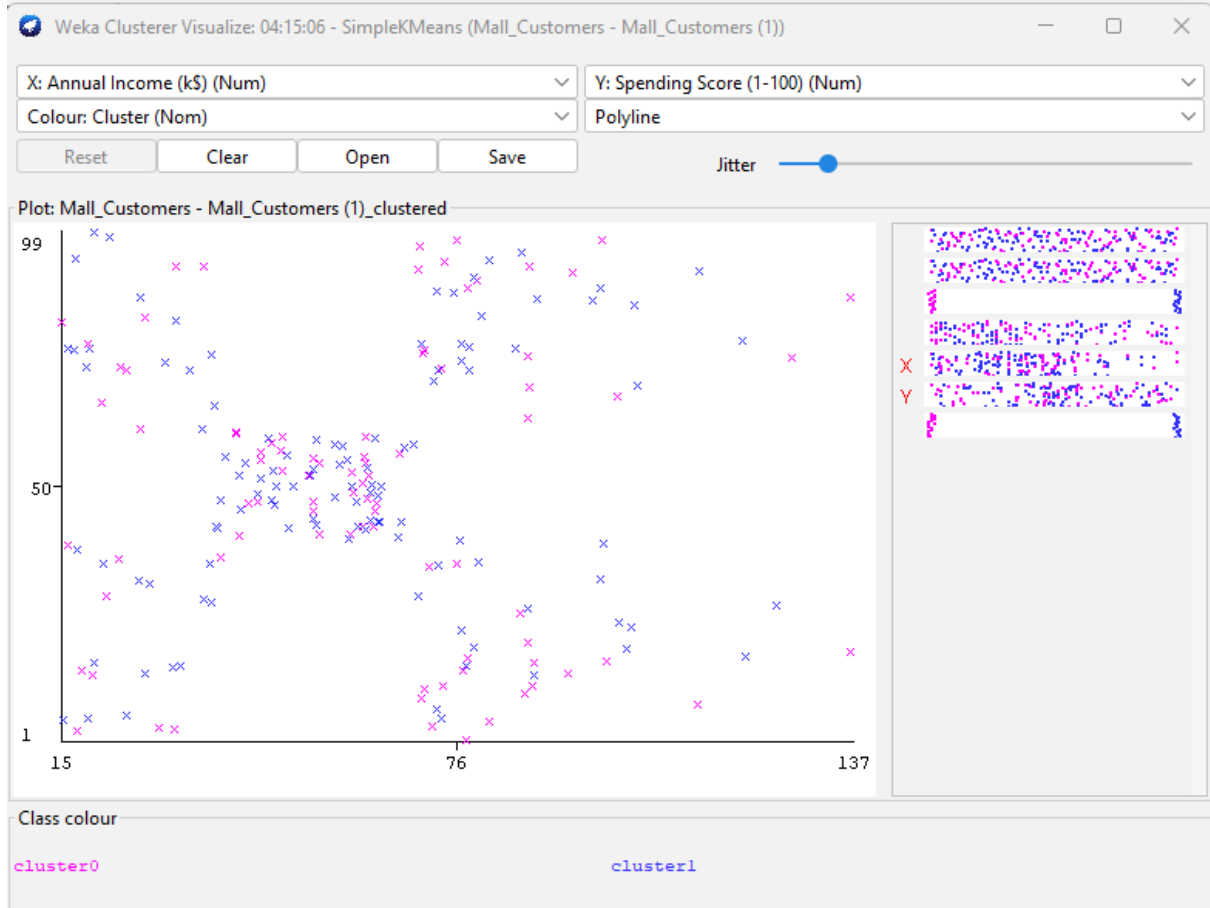
Clustered Instances

0 35 (51%)

1 33 (49%)

Status OK

Log





2. Decision Tree Induction using WEKA

A decision tree is a flowchart like tree structure, where each internal node(non-leaf node) denotes a test on an attribute,each branch represents an outcome of the test,and each leaf node (or terminal node) holds a class label. The topmost node in a tree is the root node

Example:-

Outlook	Temperature	Humidity	Windy	Class
sunny	hot	high	false	N
sunny	hot	high	true	N
overcast	hot	high	false	P
rain	mild	high	false	P
rain	cool	normal	false	P
rain	cool	normal	true	N
overcast	cool	normal	true	P
sunny	mild	high	false	N
sunny	cool	normal	false	P
rain	mild	normal	false	P
sunny	mild	normal	true	P
overcast	mild	high	true	P
overcast	hot	normal	false	P
rain	mild	high	true	N

Code:

@relation weather

@attribute outlook {sunny, overcast, rainy} @attribute temperature real

@attribute humidity real

@attribute windy {TRUE, FALSE} @attribute play {yes, no}

@data sunny,85,85,FALSE,no

sunny,80,90,TRUE,no overcast,83,86,FALSE,yes rainy,70,96,FALSE,yes

rainy,68,80,FALSE,yes rainy,65,70,TRUE,no overcast,64,65,TRUE,yes

sunny,72,95,FALSE,no sunny,69,70,FALSE,yes rainy,75,80,FALSE,yes

sunny,75,70,TRUE,yes

overcast,72,90,TRUE,yes overcast,81,75,FALSE,yes rainy,71,91,TRUE,no

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'J48 -C 0.25 -M 2'. The test options are set to 'Cross-validation' with 'Folds' set to 10. The classifier output is displayed, showing the following statistics:

Correctly Classified Instances	5034	61.9645 %
Incorrectly Classified Instances	3090	38.0355 %
Kappa statistic	0.4921	
Mean absolute error	0.0987	
Root mean squared error	0.2233	
Relative absolute error	46.2973 %	
Root relative squared error	68.4233 %	
Total Number of Instances	8124	

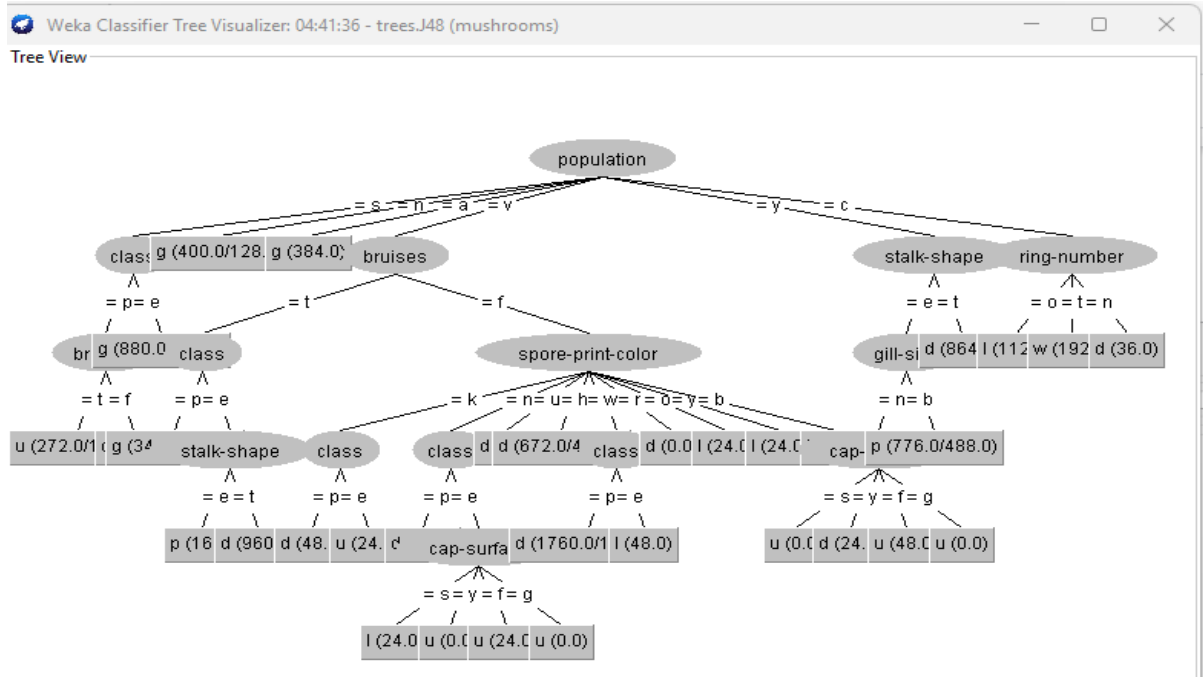
Below the statistics, there is a section for 'Detailed Accuracy By Class' and a 'Confusion Matrix'.

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
u	0.465	0.018	0.548	0.465	0.503	0.483	0.982	0.676	u
g	0.681	0.132	0.650	0.681	0.665	0.541	0.937	0.852	g
m	0.277	0.020	0.346	0.277	0.308	0.287	0.972	0.371	m
d	0.743	0.123	0.793	0.743	0.767	0.628	0.925	0.907	d
p	0.333	0.135	0.288	0.333	0.309	0.187	0.832	0.356	p
w	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	w
l	0.490	0.062	0.473	0.490	0.481	0.421	0.936	0.634	l
Weighted Avg.	0.620	0.109	0.629	0.620	0.623	0.512	0.922	0.759	

=== Confusion Matrix ===

	a	b	c	d	e	f	g	<-- Classified as
171	197	0	0	0	0	0	0	a = u
141	1463	153	108	283	0	0	0	b = g
0	211	81	0	0	0	0	0	c = m
0	160	0	2338	426	0	224	0	d = d
0	221	0	311	381	0	231	0	e = p
0	0	0	0	0	192	0	0	f = w
0	0	0	192	232	0	408	0	g = l



3. Apriori Algorithm using WEKA

In this current world, globalization is the main feature of any environment. Everyone has to be update, fast and forward and information is the main element for it. For survival in this world it's the basic need to use and to store the information means to prepare a proper database or dataset to analyze. Using and storing the database is not an issue, but finding the relevant dataset or to analyze the meaningful dataset for a particular aspect, from the junkyard of the database is very big problem in analysis of a specific part of the database. To solve this problem the concept of data mining is used to abstracts the desirable information. Useful information from the large databases has been extracted in the form of the association rules. There are many algorithms have been developed to extract the association rules from the large databases. Apriori algorithm is the most popular algorithm to extract the association rules from the databases.

TID	Items
1	A,B,C,D,G,H
2	A,B,C,D,E,F,H
3	B,C,D,E,H
4	B,E,G,H
5	A,B,D,E,G,H
6	A,C,F,G,H
7	B,D,E,G,H
8	A,C,D,E,G,H
9	B,C,D,E,H
10	A,C,E,F,H
11	C,E,H
12	A,D,E,F,H
13	B,C,E,F,H
14	A,B,C,F,H
15	A,B,E,F,H



Example

CODE:

```
@relation TEST_ITEM_TRANS @attribute A {TRUE, FALSE} @attribute B {TRUE, FALSE} @attribute C {TRUE, FALSE} @attribute D {TRUE, FALSE} @attribute E {TRUE, FALSE} @attribute F {TRUE, FALSE} @attribute G {TRUE, FALSE} @attribute H {TRUE, FALSE} @data
TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,TRUE,TRUE
TRUE,TRUE,TRUE,TRUE,TRUE,TRUE,FALSE,TRUE
FALSE,TRUE,TRUE,TRUE,FALSE,FALSE,FALSE,TRUE
FALSE,TRUE,FALSE,FALSE,TRUE,FALSE,TRUE,TRUE
TRUE,TRUE,FALSE,TRUE,TRUE,FALSE,TRUE,TRUE
TRUE,FALSE,TRUE,FALSE,FALSE,TRUE,TRUE,TRUE
FALSE,TRUE,FALSE,TRUE,TRUE,FALSE,TRUE,TRUE
TRUE,FALSE,TRUE,TRUE,TRUE,FALSE,TRUE,TRUE
FALSE,TRUE,TRUE,TRUE,TRUE,FALSE,FALSE,TRUE
TRUE,FALSE,TRUE,FALSE,TRUE,TRUE,FALSE,TRUE
FALSE,FALSE,TRUE,FALSE,TRUE,FALSE,FALSE,TRUE
TRUE,FALSE,FALSE,TRUE,TRUE,TRUE,FALSE,TRUE
FALSE,TRUE,TRUE,FALSE,TRUE,TRUE,FALSE,TRUE
TRUE,TRUE,TRUE,FALSE,FALSE,TRUE,FALSE,TRUE TRUE,TRUE,FALSE,FALSE,
TRUE,TRUE,FALSE,TRUE
```

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Associator

Choose Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Start Stop

Result list (right-click for ...)

16:04:55 - Apriori

Associator output

Apriori

Minimum support: 0.15 (694 instances)

Minimum metric <confidence>: 0.9

Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 44

Size of set of large itemsets L(2): 380

Size of set of large itemsets L(3): 910

Size of set of large itemsets L(4): 633

Size of set of large itemsets L(5): 105

Size of set of large itemsets L(6): 1

Best rules found:

1. biscuits=frozen foods= fruit=t total=high 788 ==> bread and cake=t 723 <conf:(0.92)> lift:(1.27) lev:(0.03) [155] conv:(3.35)
2. baking needs=t biscuits=t fruit=t total=high 760 ==> bread and cake=t 696 <conf:(0.92)> lift:(1.27) lev:(0.03) [149] conv:(3.28)
3. baking needs=t frozen foods=t fruit=t total=high 770 ==> bread and cake=t 705 <conf:(0.92)> lift:(1.27) lev:(0.03) [150] conv:(3.27)
4. biscuits=t fruit=t vegetables=t total=high 915 ==> bread and cake=t 746 <conf:(0.92)> lift:(1.27) lev:(0.03) [159] conv:(3.26)
5. party snack foods=t fruit=t total=high 854 ==> bread and cake=t 775 <conf:(0.91)> lift:(1.27) lev:(0.04) [164] conv:(3.15)
6. biscuits=t frozen foods=t vegetables=t total=high 797 ==> bread and cake=t 725 <conf:(0.91)> lift:(1.26) lev:(0.03) [151] conv:(3.06)
7. baking needs=t biscuits=t vegetables=t total=high 772 ==> bread and cake=t 701 <conf:(0.91)> lift:(1.26) lev:(0.03) [145] conv:(3.01)
8. biscuits=t fruit=t total=high 954 ==> bread and cake=t 866 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(3)
9. frozen foods=t fruit=t vegetables=t total=high 834 ==> bread and cake=t 757 <conf:(0.91)> lift:(1.26) lev:(0.03) [156] conv:(3)

Status OK

Log x1

Conclusion: