Name: Ishita More
UID: 2019130039 (A)
TE Comps
Subject: Data Analytics Lab

Exp No.: 05

Aim:

Apriori Algorithm and Association rule mining with WEKA

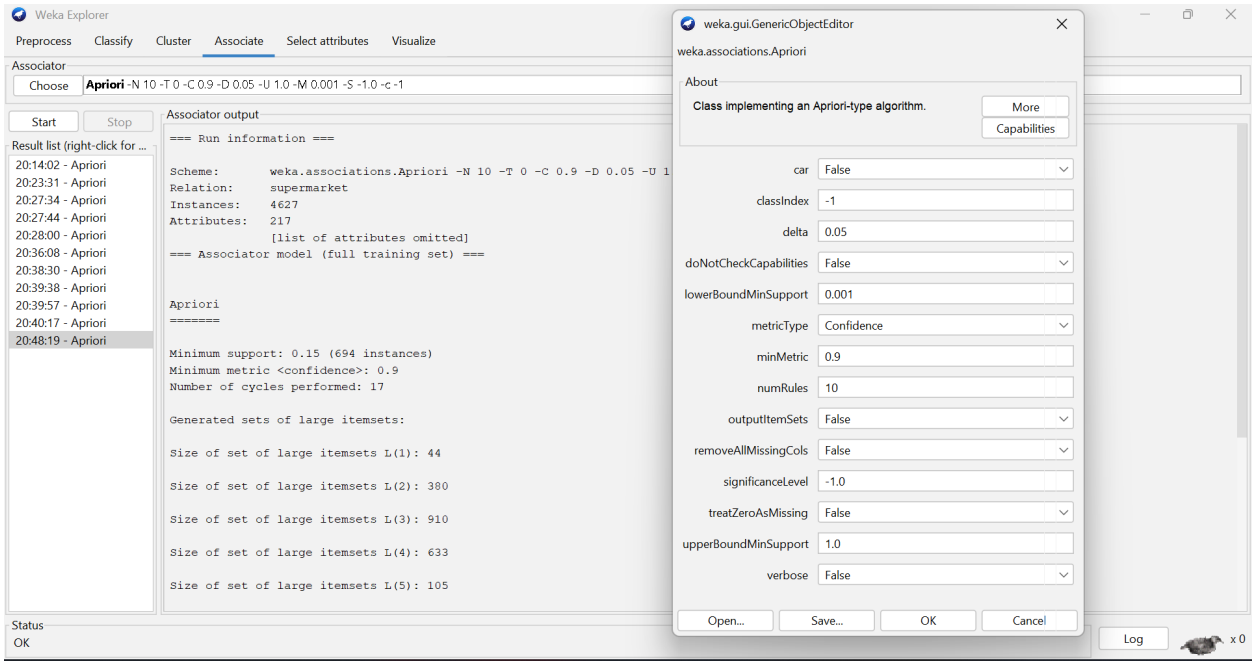
Objective;

Apply Apriori Algorithm to given dataset

Association Rule Mining with WEKA

Task

Consider dataset “Groceries” and apply apriori algorithm on it. What are the first 5 rules generated when the min support is 0.001 (0.1%) and min confidence is 0.9 (90%) .

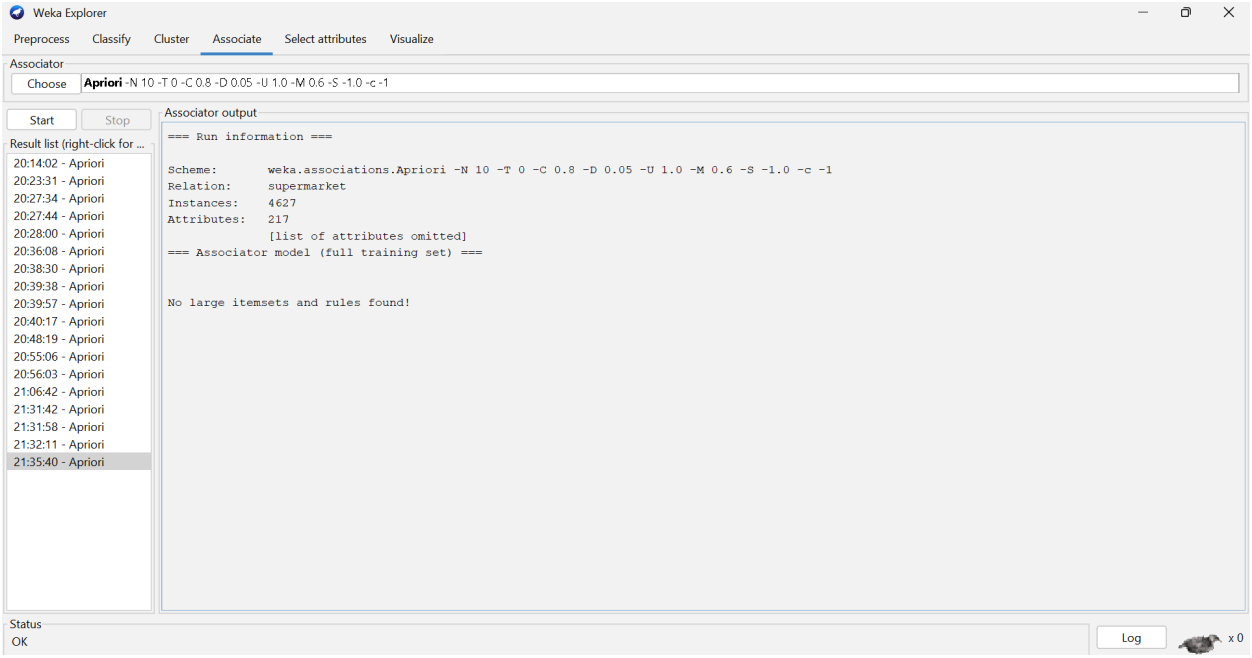


Best rules found:

1. biscuits=t frozen foods=t 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 815 ==> 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 779 <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)
10. frozen foods=t fruit=t total=high 969 ==> bread and cake=t 877 <conf:(0.91)> lift:(1.26) lev:(0.04) [179] conv:(2.92)

Exercise 1: Basic association rule creation manually

The 'database' below has four transactions. What association rules can be found in this set, if the minimum support (i.e coverage) is 60% and the minimum confidence (i.e. accuracy) is 80% ?



Exercise 2: Input file generation and Initial experiments with Weka's association rule discovery.

@relation exercise

@attribute exista {TRUE, FALSE}

@attribute existb {TRUE, FALSE}

@attribute existc {TRUE, FALSE}

@attribute existd {TRUE, FALSE}

@attribute existe {TRUE, FALSE}

@attribute existk {TRUE, FALSE}

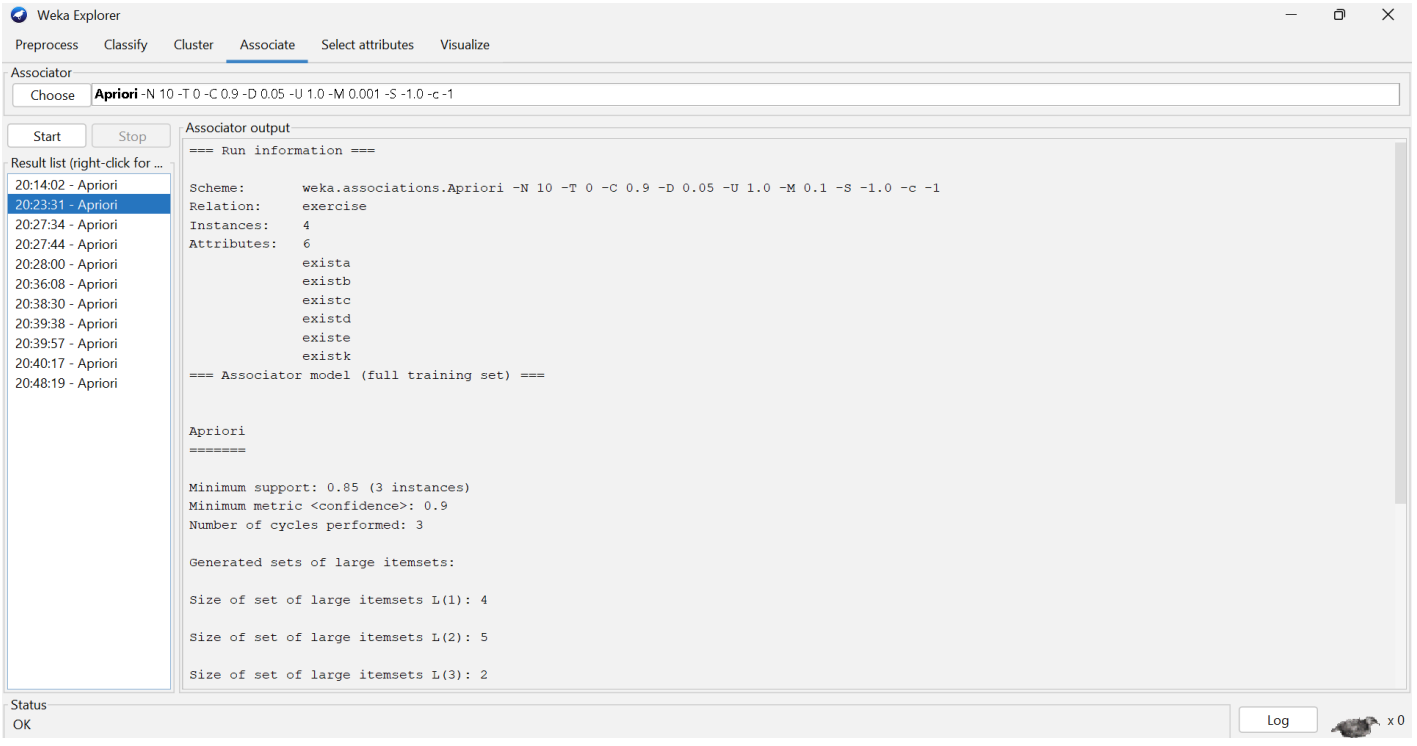
@data

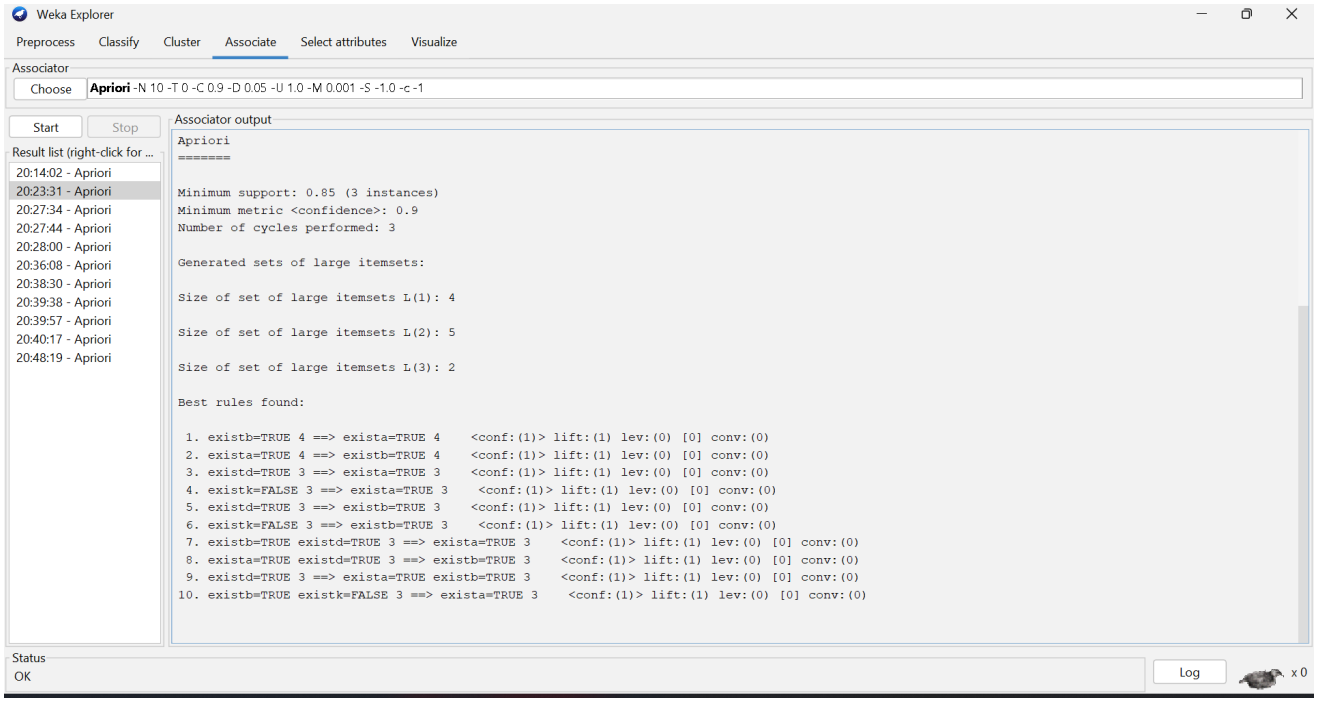
TRUE,TRUE,FALSE,TRUE,FALSE,TRUE

TRUE,TRUE,TRUE,TRUE,TRUE,FALSE

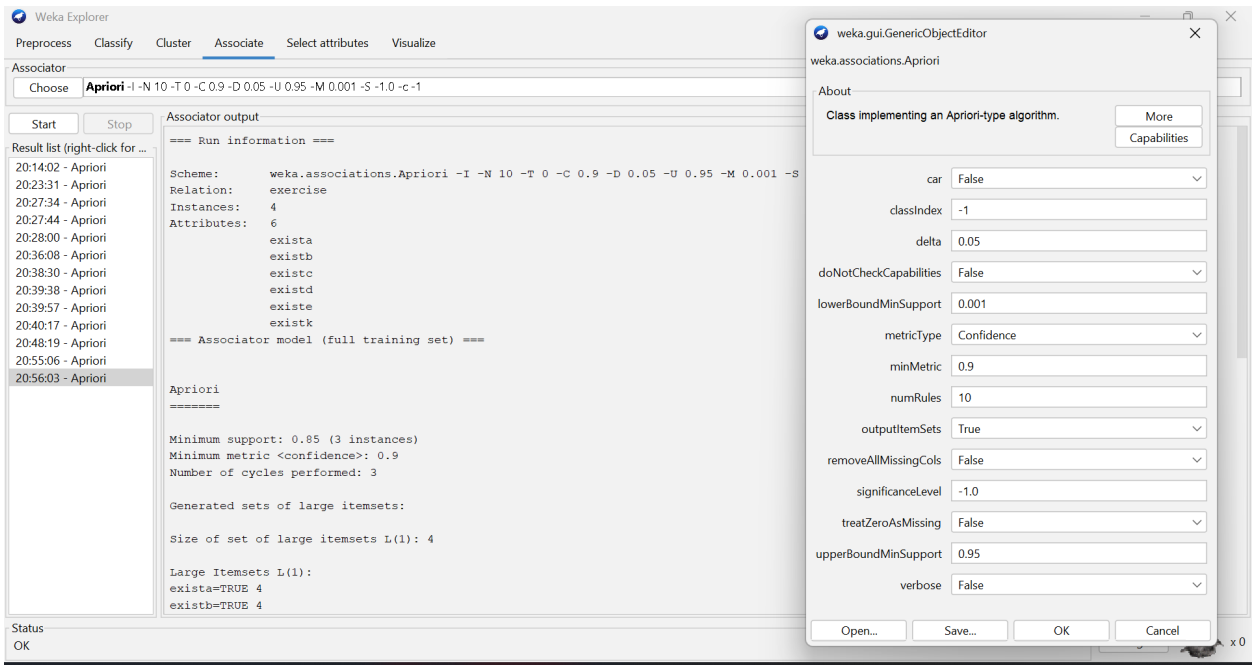
TRUE,TRUE,TRUE,FALSE,TRUE,FALSE

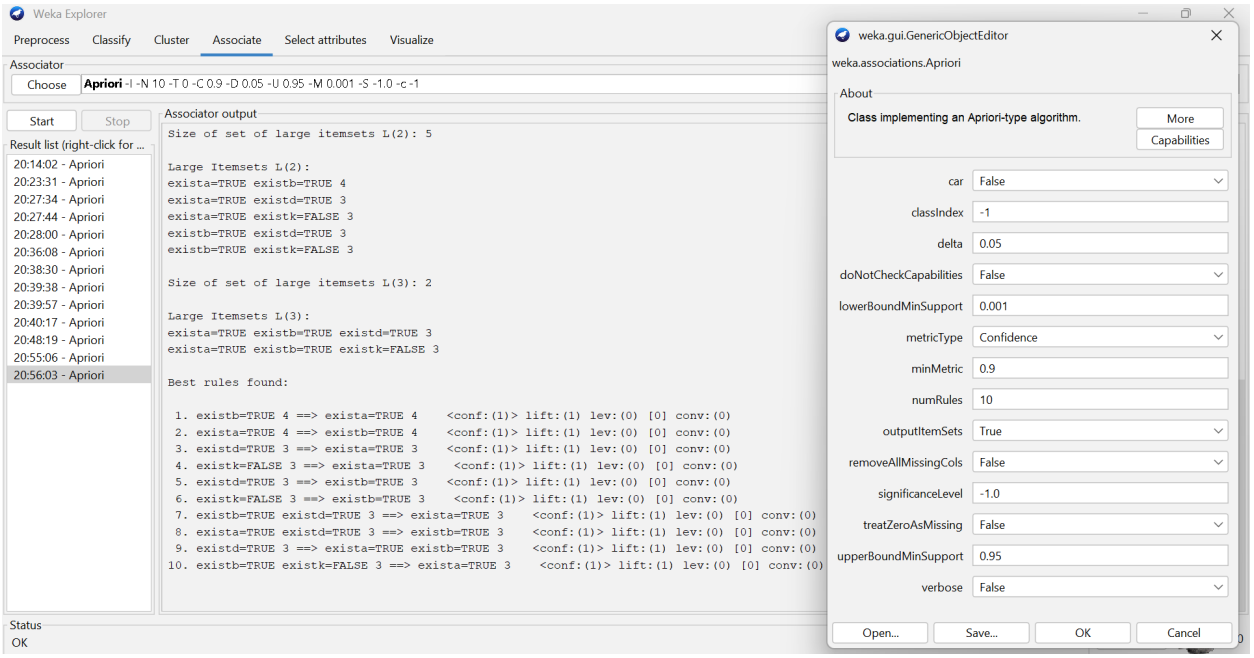
TRUE,TRUE,FALSE,TRUE,FALSE,FALSE



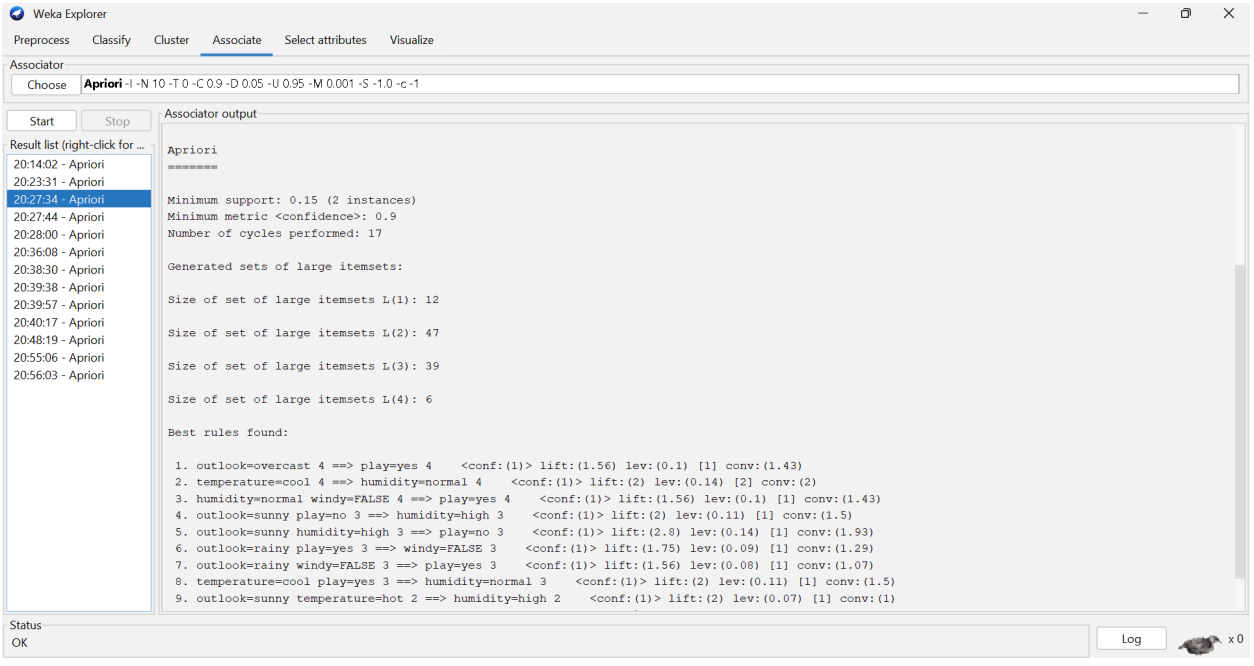

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
Associator output
Apriori
=======
Minimum support: 0.85 (3 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3
Generated sets of large itemsets:
Size of set of large itemsets L(1): 4
Size of set of large itemsets L(2): 5
Size of set of large itemsets L(3): 2
Best rules found:
1. existb=TRUE 4 ==> exista=TRUE 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
2. exista=TRUE 4 ==> existb=TRUE 4 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
3. existd=TRUE 3 ==> exista=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
4. existk=FALSE 3 ==> exista=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
5. existd=TRUE 3 ==> existb=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
6. existk=FALSE 3 ==> existb=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
7. existb=TRUE existd=TRUE 3 ==> exista=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
8. exista=TRUE existd=TRUE 3 ==> existb=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
9. existd=TRUE 3 ==> exista=TRUE existb=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
10. existb=TRUE existk=FALSE 3 ==> exista=TRUE 3 <conf:(1)> lift:(1) lev:(0) [0] conv:(0)
Status
OK
Log
x 0

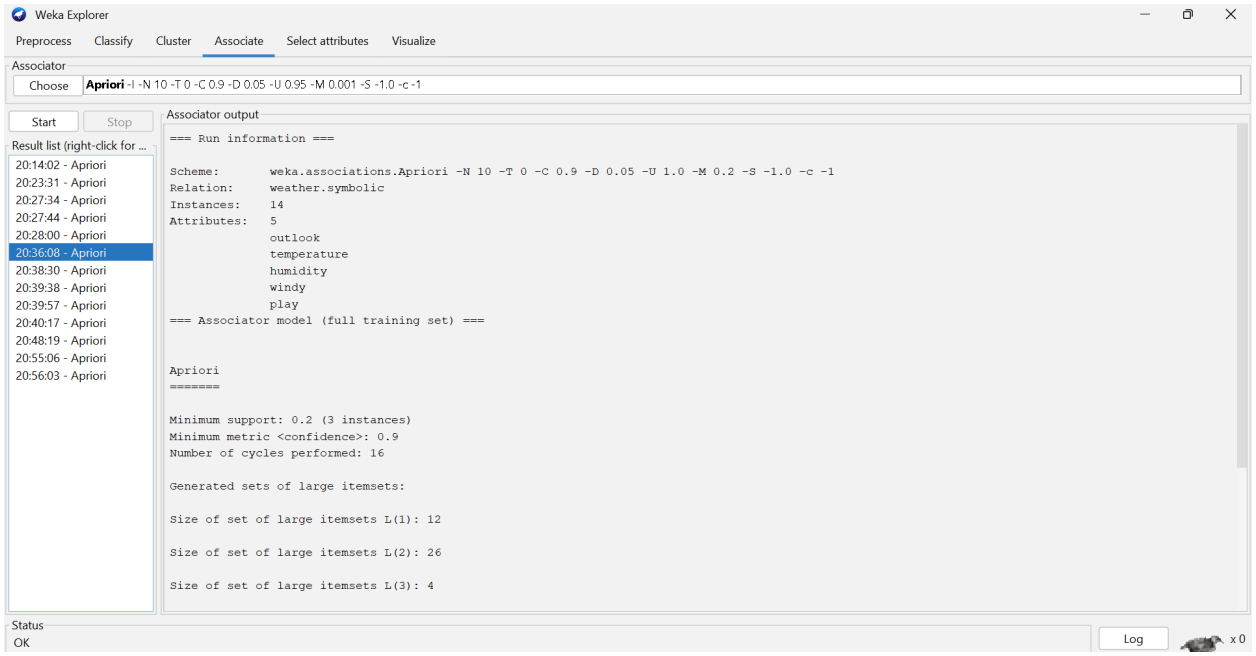
Update

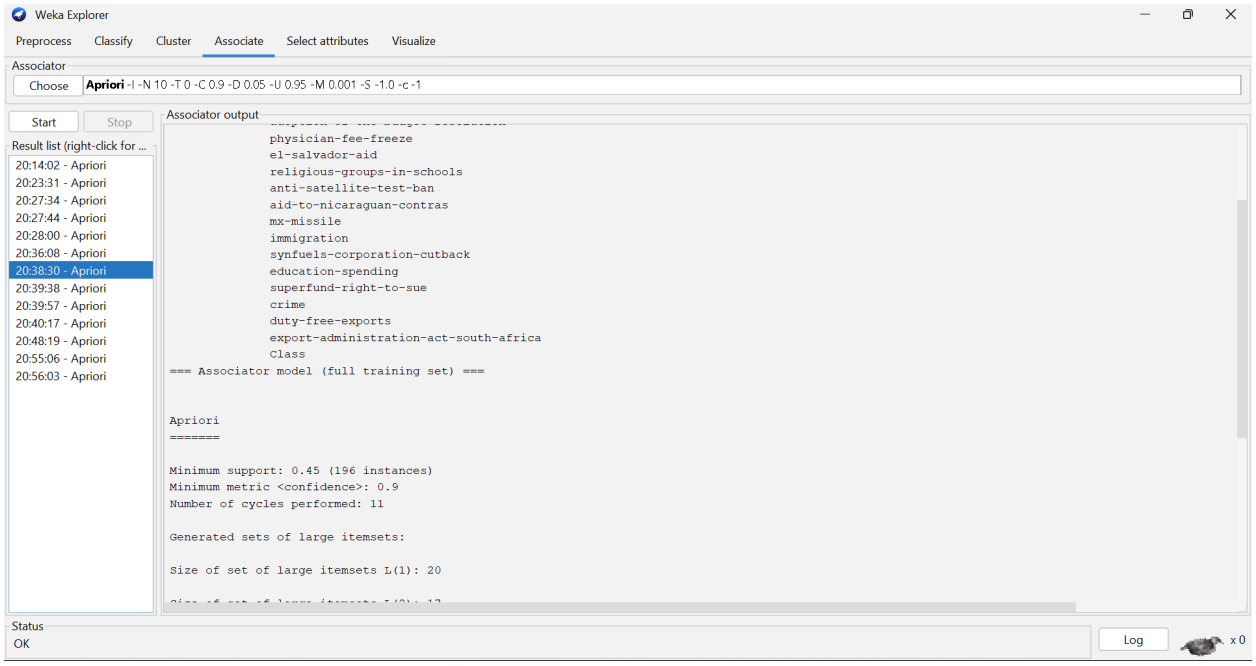

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
20:55:06 - Apriori
20:56:03 - Apriori
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S
Relation: exercise
Instances: 4
Attributes: 6
exista
existb
existc
existd
existe
existk
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.85 (3 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 3
Generated sets of large itemsets:
Size of set of large itemsets L(1): 4
Large Itemsets L(1):
exista=TRUE 4
existb=TRUE 4
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car False
classIndex -1
delta 0.05
doNotCheckCapabilities False
lowerBoundMinSupport 0.001
metricType Confidence
minMetric 0.9
numRules 10
outputItemSets True
removeAllMissingCols False
significanceLevel -1.0
treatZeroAsMissing False
upperBoundMinSupport 0.95
verbose False
Open...
Save...
OK
Cancel
Status
OK
x 0



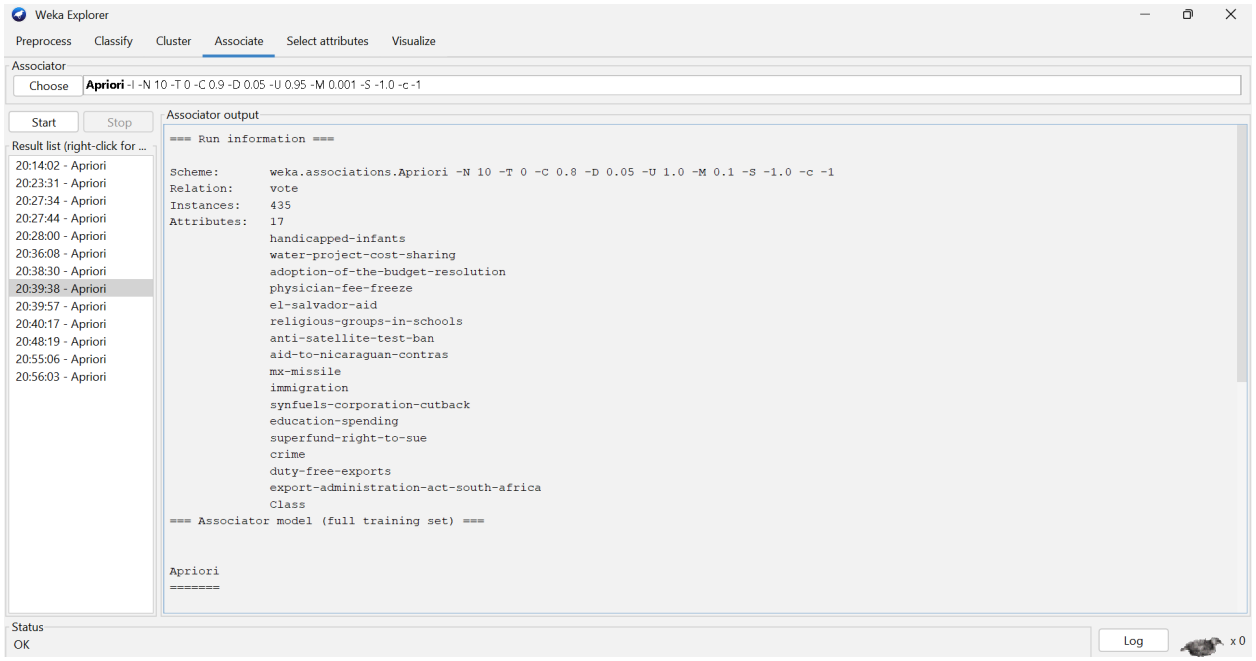
Exercise 3: Mining Association Rule with WEKA Explorer – Weather dataset

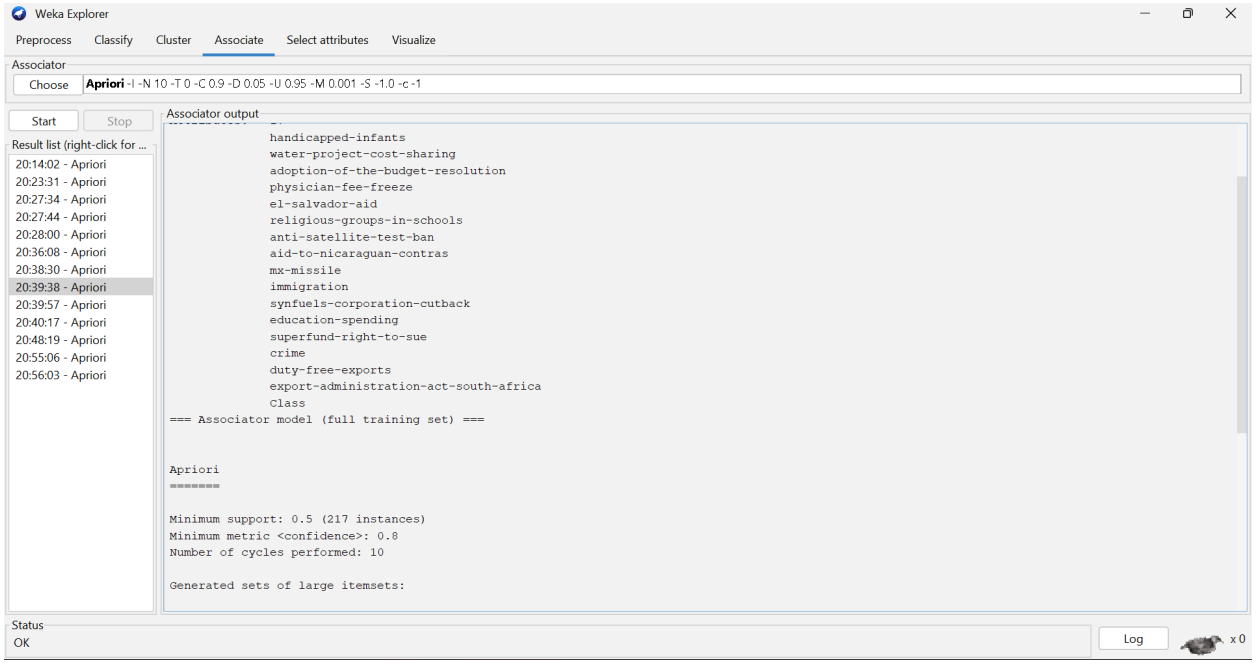


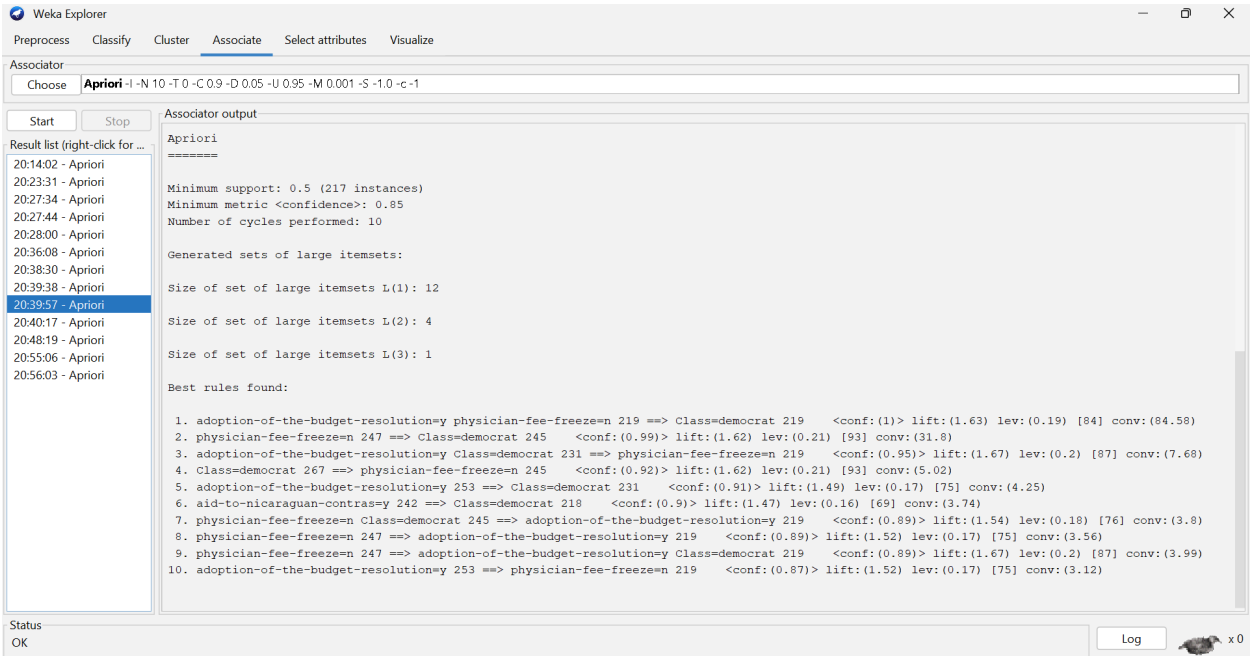

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
20:55:06 - Apriori
20:56:03 - Apriori
Associator output
=== Run information ===
Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1
Relation: weather.symbolic
Instances: 14
Attributes: 5
outlook
temperature
humidity
windy
play
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.2 (3 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 16
Generated sets of large itemsets:
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 26
Size of set of large itemsets L(3): 4
Status
OK
Log
x 0

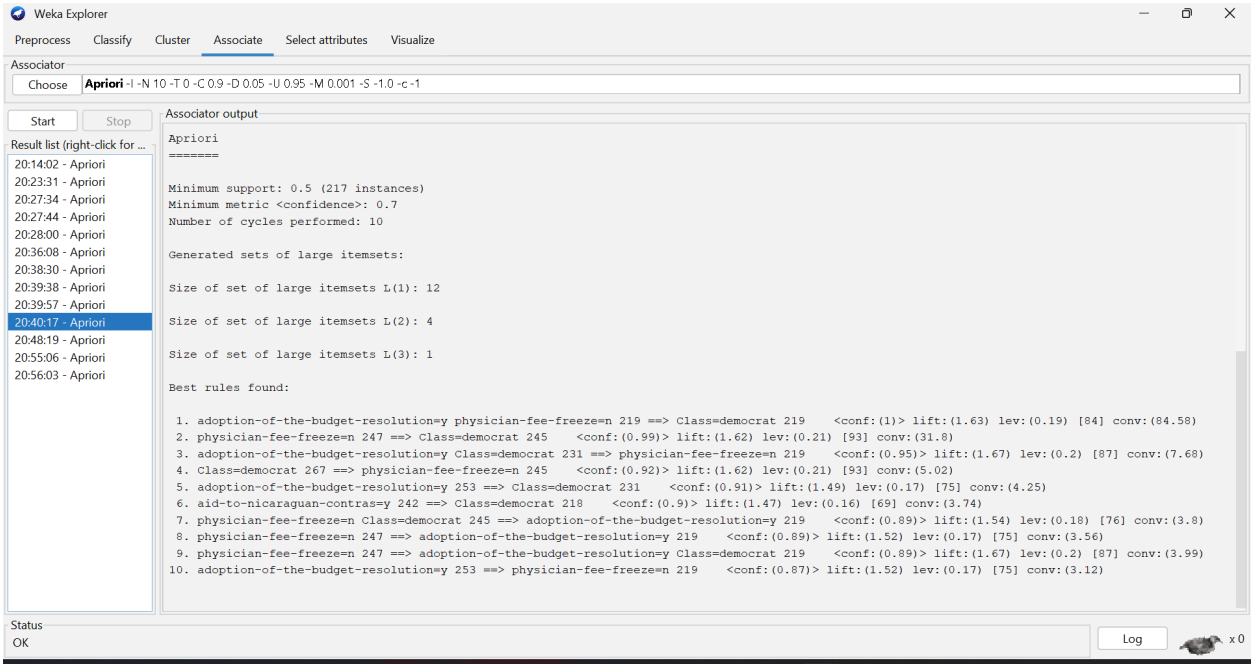

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
20:55:06 - Apriori
20:56:03 - Apriori
Associator output
physician-fee-freeze
el-salvador-aid
religious-groups-in-schools
anti-satellite-test-ban
aid-to-nicaraguan-contras
mx-missile
immigration
synfuels-corporation-cutback
education-spending
superfund-right-to-sue
crime
duty-free-exports
export-administration-act-south-africa
Class
=== Associator model (full training set) ===
Apriori
=======
Minimum support: 0.45 (196 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11
Generated sets of large itemsets:
Size of set of large itemsets L(1): 20
Status
OK
Log
x 0

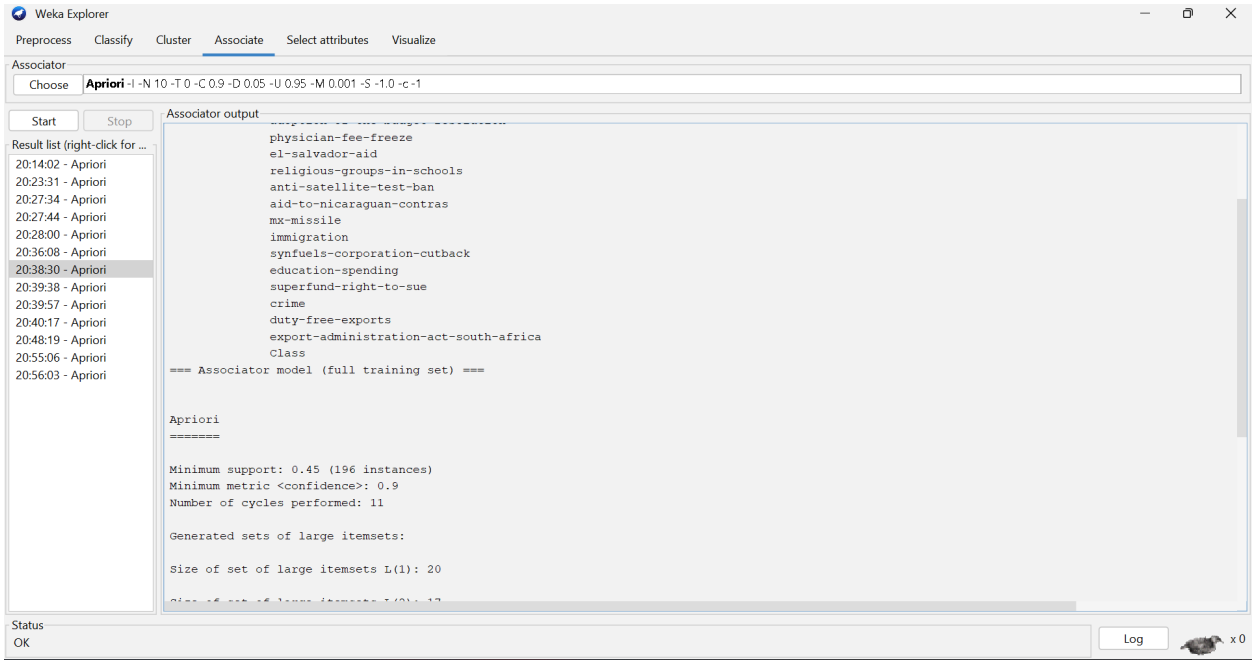
Exercise 4: Mining Association Rule with WEKA Explorer – Vote






Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
20:55:06 - Apriori
20:56:03 - Apriori
Associator output
Apriori
=======
Minimum support: 0.5 (217 instances)
Minimum metric <confidence>: 0.85
Number of cycles performed: 10
Generated sets of large itemsets:
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1
Best rules found:
1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 <conf:(1)> lift:(1.63) lev:(0.19) [84] conv:(84.58)
2. physician-fee-freeze=n 247 ==> Class=democrat 245 <conf:(0.99)> lift:(1.62) lev:(0.21) [93] conv:(31.8)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219 <conf:(0.95)> lift:(1.67) lev:(0.2) [87] conv:(7.68)
4. Class=democrat 267 ==> physician-fee-freeze=n 245 <conf:(0.92)> lift:(1.62) lev:(0.21) [93] conv:(5.02)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231 <conf:(0.91)> lift:(1.49) lev:(0.17) [75] conv:(4.25)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218 <conf:(0.9)> lift:(1.47) lev:(0.16) [69] conv:(3.74)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219 <conf:(0.89)> lift:(1.54) lev:(0.18) [76] conv:(3.8)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219 <conf:(0.89)> lift:(1.52) lev:(0.17) [75] conv:(3.56)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219 <conf:(0.89)> lift:(1.67) lev:(0.2) [87] conv:(3.99)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219 <conf:(0.87)> lift:(1.52) lev:(0.17) [75] conv:(3.12)
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 0.95 -M 0.001 -S -1.0 -c -1
Start
Stop
Result list (right-click for ...
20:14:02 - Apriori
20:23:31 - Apriori
20:27:34 - Apriori
20:27:44 - Apriori
20:28:00 - Apriori
20:36:08 - Apriori
20:38:30 - Apriori
20:39:38 - Apriori
20:39:57 - Apriori
20:40:17 - Apriori
20:48:19 - Apriori
20:55:06 - Apriori
20:56:03 - Apriori
Associator output
Apriori
=======
Minimum support: 0.5 (217 instances)
Minimum metric <confidence>: 0.7
Number of cycles performed: 10
Generated sets of large itemsets:
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 4
Size of set of large itemsets L(3): 1
Best rules found:
1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democrat 219 <conf:(1)> lift:(1.63) lev:(0.19) [84] conv:(84.58)
2. physician-fee-freeze=n 247 ==> Class=democrat 245 <conf:(0.99)> lift:(1.62) lev:(0.21) [93] conv:(31.8)
3. adoption-of-the-budget-resolution=y Class=democrat 231 ==> physician-fee-freeze=n 219 <conf:(0.95)> lift:(1.67) lev:(0.2) [87] conv:(7.68)
4. Class=democrat 267 ==> physician-fee-freeze=n 245 <conf:(0.92)> lift:(1.62) lev:(0.21) [93] conv:(5.02)
5. adoption-of-the-budget-resolution=y 253 ==> Class=democrat 231 <conf:(0.91)> lift:(1.49) lev:(0.17) [75] conv:(4.25)
6. aid-to-nicaraguan-contras=y 242 ==> Class=democrat 218 <conf:(0.9)> lift:(1.47) lev:(0.16) [69] conv:(3.74)
7. physician-fee-freeze=n Class=democrat 245 ==> adoption-of-the-budget-resolution=y 219 <conf:(0.89)> lift:(1.54) lev:(0.18) [76] conv:(3.8)
8. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y 219 <conf:(0.89)> lift:(1.52) lev:(0.17) [75] conv:(3.56)
9. physician-fee-freeze=n 247 ==> adoption-of-the-budget-resolution=y Class=democrat 219 <conf:(0.89)> lift:(1.67) lev:(0.2) [87] conv:(3.99)
10. adoption-of-the-budget-resolution=y 253 ==> physician-fee-freeze=n 219 <conf:(0.87)> lift:(1.52) lev:(0.17) [75] conv:(3.12)
Status
OK
Log
x 0



When support is 0.5, the confidence level is 0.7, then no of cycles are 10
When support is 0.5, the confidence level is 0.8, then no of cycles are 10
When support is 0.5, the confidence level is 0.85, then no of cycles are 10
When support is 0.45, the confidence level is 0.9, then no of cycles are 11

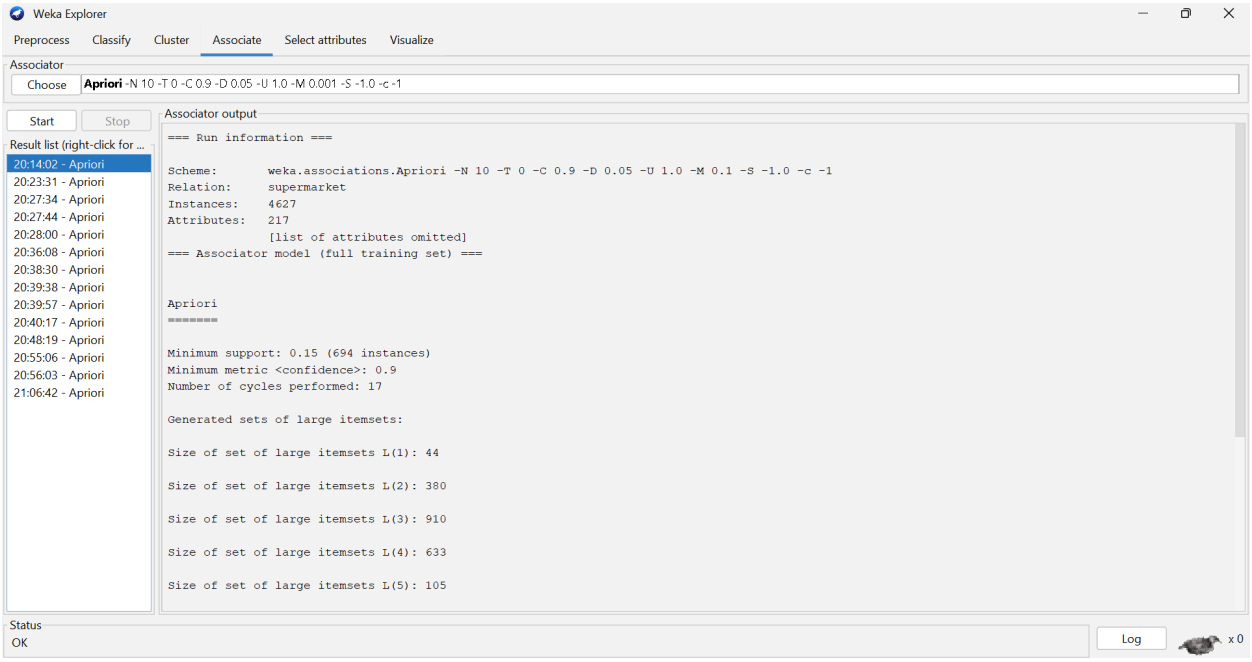
Therefore, when the confidence level is increased by a certain level, support decreases, and the number of cycles increase.

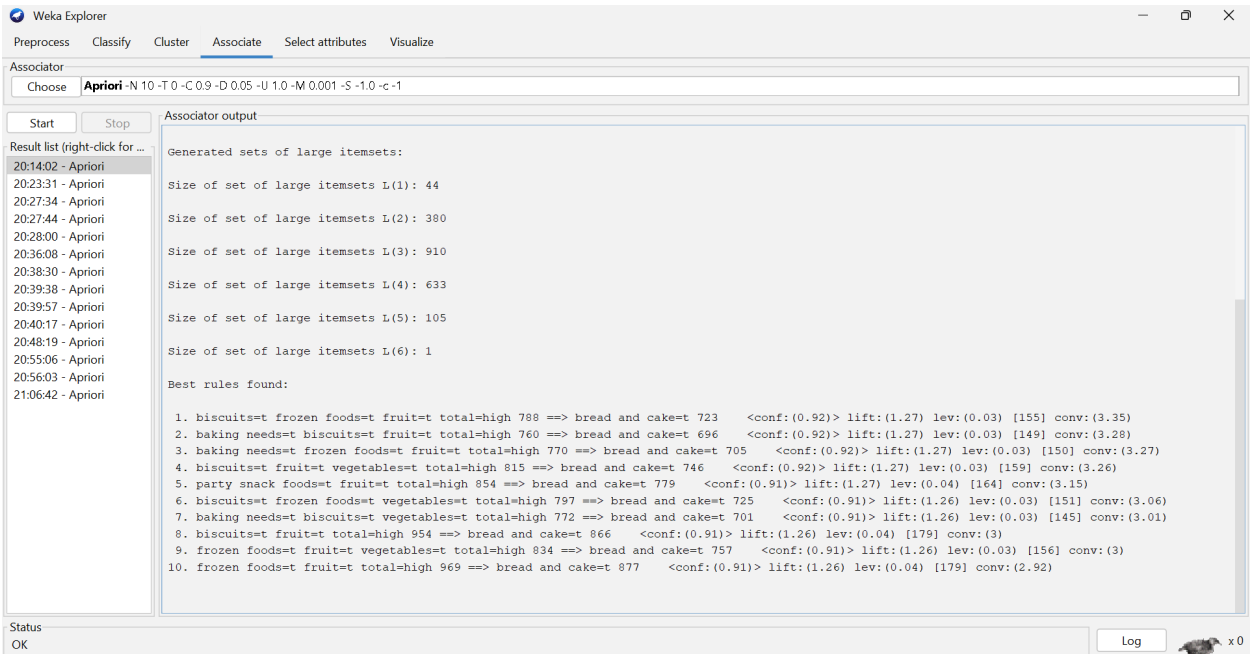
Exercise 5: Let's run Apriori on another real-world dataset.

Load data at Preprocess tab. Click the Open file button to bring up a standard dialog through which you can select a file. Choose the supermarket.arff file. To see the original dataset, click the Edit button, a viewer window opens with dataset loaded.

To do market basket analysis in Weka, each transaction is coded as an instance of which the attributes represent the items in the store. Each attribute has only one value: If a particular transaction does not contain it (i.e., the customer did not buy that item), this is coded as a missing value.

Task 1. Experiment with Apriori and investigate the effect of the various parameters described before. Prepare a brief oral presentation on the main findings of your investigation.





Apriori algorithm is a sequence of steps to be followed to find the most frequent itemset in the given database. This data mining technique follows the join and the prune steps iteratively until the most frequent itemset is achieved. The primary objective of the apriori algorithm is to create the association rule between different objects. The association rule describes how two or more objects are related to one another. Apriori algorithm is also called frequent pattern mining.

In this dataset, after association 10 rules rae generated when minimum support is 0.15 and confidence is 0.9. The number of cycles perform are 17 respectively.

Here, we have the biscuits, frozen foods, fruit which has the high of 788 which gives bread and cake, We have the 92% confidence and there's 100% support. Then we have baking needs, biscuits, fruits which is total of 760 and then we have the bread and cake. Similarly, we have baking needs, frozen foods, fruits which is total of 770 and then we have the bread and cake. we have biscuits, fruits, vegetables which is total of 815 and then we have the bread and cake. we have party snack foods, fruits which is total of 854 and then we have the bread and cake. we have the biscuits, frozen foods, vegetables which has the high of 797 which gives bread and cake and etc.