

# ΕΡΓΑΣΙΑ “WEKA” ΦΟΥΡΚΙΩΤΗΣ ΚΩΝΣΤΑΝΤΙΝΟΣ ΠΑΝΑΓΙΩΤΗΣ

Παράδειγμα στην Προεπεξεργασία Δεδομένων ( bank-data.csv)

1<sup>ο</sup> βήμα – Διαγραφή Id

Attributes

All None Invert Pattern

No.	Name
1	<input type="checkbox"/> age
2	<input type="checkbox"/> sex
3	<input type="checkbox"/> region
4	<input type="checkbox"/> income
5	<input type="checkbox"/> married
6	<input checked="" type="checkbox"/> children
7	<input type="checkbox"/> car
8	<input type="checkbox"/> save_act
9	<input type="checkbox"/> current_act
10	<input type="checkbox"/> mortgage
11	<input type="checkbox"/> pep

Remove

2<sup>ο</sup> βήμα – Διακριτοποίηση Children σε 4 ομάδες

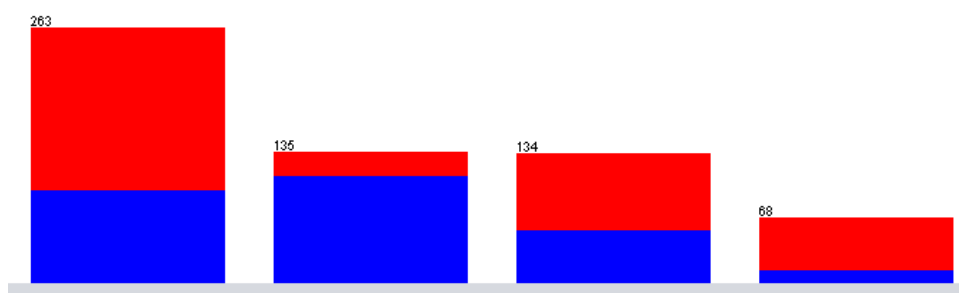
Name: children  
Missing: 0 (0%)

Distinct: 4

Type: Nominal  
Unique: 0 (0%)

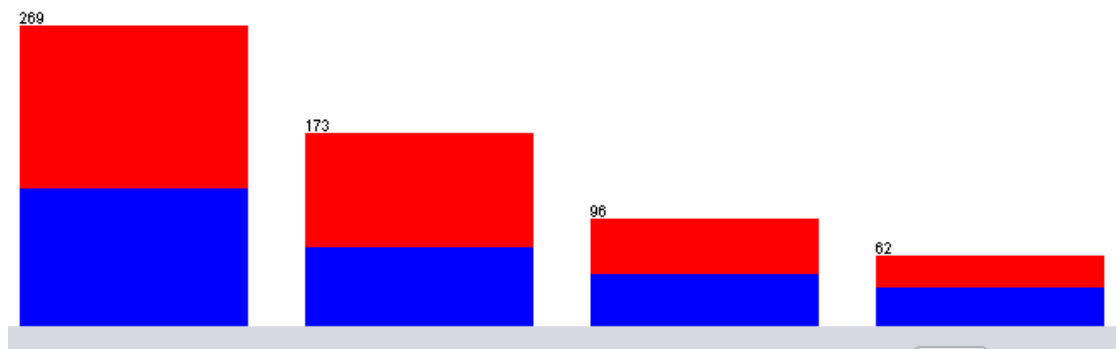
No.	Label	Count	Weight
1	'(-inf-0.75]'	263	263.0
2	'(0.75-1.5]'	135	135.0
3	'(1.5-2.25]'	134	134.0
4	'(2.25-inf]'	68	68.0

Class: pep (Nom) Visualize All



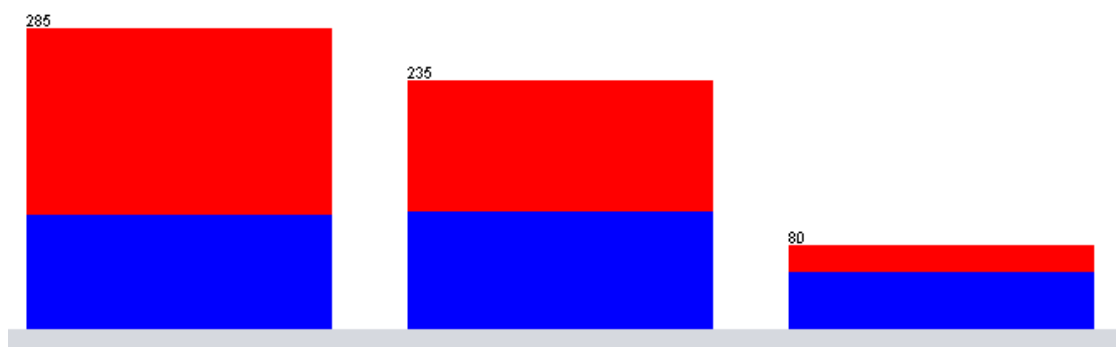
Name: region		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 4	
No.	Label	Count	Weight
1	INNER_CITY	269	269.0
2	TOWN	173	173.0
3	RURAL	96	96.0
4	SUBURBAN	62	62.0

Class: pep (Nom) Visualize All



Name: income		Type: Nominal	
Missing: 0 (0%)		Unique: 0 (0%)	
		Distinct: 3	
No.	Label	Count	Weight
1	'(-inf-24386.173333]'	285	285.0
2	'(24386.173333-43758.136667]'	235	235.0
3	'(43758.136667-inf]'	80	80.0

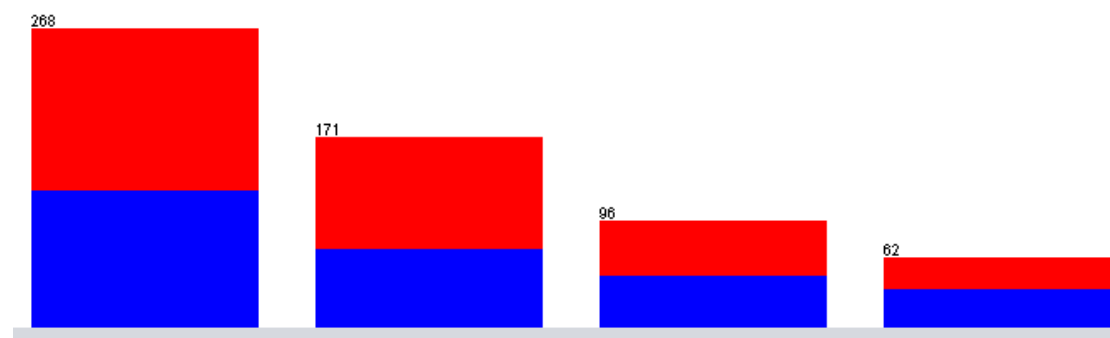
Class: pep (Nom) Visualize All



### 3<sup>ο</sup> βήμα - Ελλιπή δεδομένα (missing values)

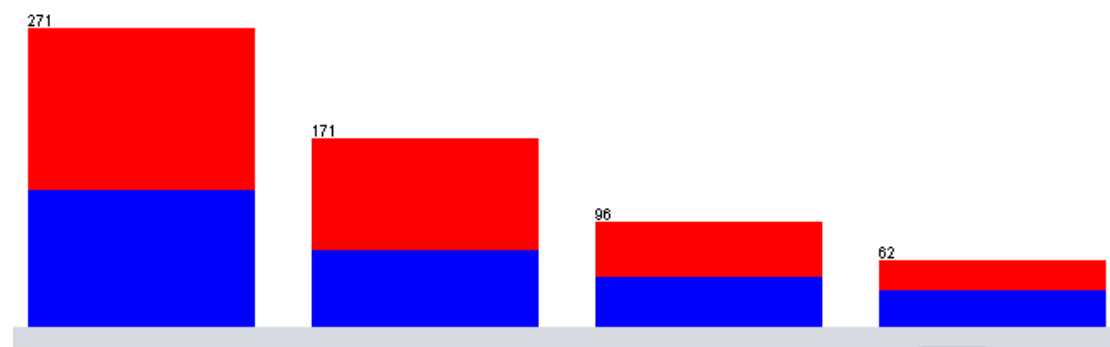
Name: region		Type: Nominal	
Missing: 3 (1%)		Distinct: 4	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	INNER_CITY	268	268.0
2	TOWN	171	171.0
3	RURAL	96	96.0
4	SUBURBAN	62	62.0

Class: pep (Nom) Visualize All



Name: region		Type: Nominal	
Missing: 0 (0%)		Distinct: 4	
		Unique: 0 (0%)	
No.	Label	Count	Weight
1	INNER_CITY	271	271.0
2	TOWN	171	171.0
3	RURAL	96	96.0
4	SUBURBAN	62	62.0

Class: pep (Nom) Visualize All



## Άσκηση 1η στην Κατηγοριοποίηση Δεδομένων στο Weka (analcatdata\_bankruptcy.arff)

### 1<sup>ο</sup> Βήμα - REMOVE COMPANY

No.	Name
1	<input checked="" type="checkbox"/> WC/TA
2	<input type="checkbox"/> RE/TA
3	<input type="checkbox"/> EBIT/TA
4	<input type="checkbox"/> S/TA
5	<input type="checkbox"/> BVE/BVL
6	<input type="checkbox"/> Bankrupt

### 2<sup>ο</sup> βήμα – J48 (CROSS VALIDATION)

Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===  
=== Summary ===

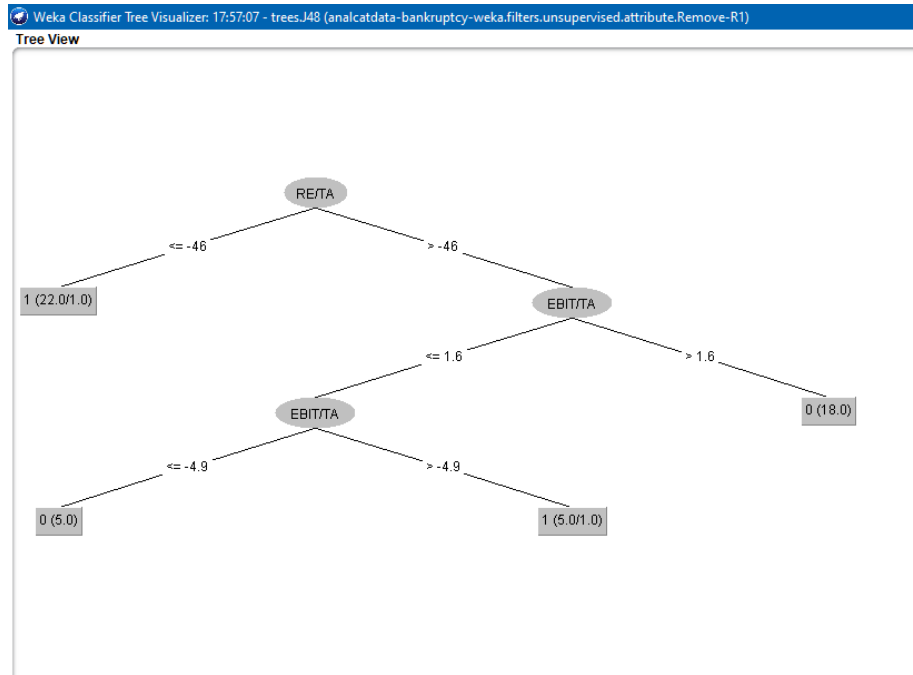
Correctly Classified Instances	39	78	%
Incorrectly Classified Instances	11	22	%
Kappa statistic	0.56		
Mean absolute error	0.2273		
Root mean squared error	0.4358		
Relative absolute error	45.2735 %		
Root relative squared error	86.7635 %		
Total Number of Instances	50		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,680	0,120	0,850	0,680	0,756	0,572	0,785	0,786	0
	0,880	0,320	0,733	0,880	0,800	0,572	0,785	0,692	1
Weighted Avg.	0,780	0,220	0,792	0,780	0,778	0,572	0,785	0,739	

=== Confusion Matrix ===

```
a b  <-- classified as
17 8 | a = 0
 3 22 | b = 1
```



### 3<sup>ο</sup> βήμα Perceptron (CROSS VALIDATION)

☒ Cross-validation Folds 10  
☐ Percentage split % 66  
 More options...

(Nom) Bankrupt  
 Start Stop

Result list (right-click for options)  
 17:54:28 - trees.J48  
 18:01:45 - functions.MultilayerPerceptron

Time taken to build model: 0.12 seconds

=== Stratified cross-validation ===  
 === Summary ===

Correctly Classified Instances	45	90	%
Incorrectly Classified Instances	5	10	%
Kappa statistic	0.8		
Mean absolute error	0.1256		
Root mean squared error	0.2746		
Relative absolute error	25.0159 %		
Root relative squared error	54.6702 %		
Total Number of Instances	50		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,880	0,080	0,917	0,880	0,898	0,801	0,965	0,964	0
	0,920	0,120	0,885	0,920	0,902	0,801	0,965	0,971	1
Weighted Avg.	0,900	0,100	0,901	0,900	0,900	0,801	0,965	0,967	

=== Confusion Matrix ===

```

a b  <-- classified as
22 3 | a = 0
 2 23 | b = 1
  
```

## 4<sup>ο</sup> βήμα BayesNet (CROSSVALIDATION)

Percentage split % 66

More options...

(Nom) Bankrupt

Start Stop

Result list (right-click for options)

- 17:54:28 - trees.J48
- 18:01:45 - functions.MultilayerPerceptron
- 18:05:16 - bayes.BayesNet

```
==== Stratified cross-validation ====
==== Summary ====
Correctly Classified Instances      44      88 %
Incorrectly Classified Instances    6      12 %
Kappa statistic                    0.76
Mean absolute error                0.1427
Root mean squared error            0.3198
Relative absolute error            28.4282 %
Root relative squared error        63.6768 %
Total Number of Instances          50

==== Detailed Accuracy By Class ====
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,880    0,120    0,880    0,880    0,880    0,760    0,933    0,894    0
          0,880    0,120    0,880    0,880    0,880    0,760    0,933    0,948    1
Weighted Avg.    0,880    0,120    0,880    0,880    0,880    0,760    0,933    0,921

==== Confusion Matrix ====
  a  b  <-- classified as
22  3  |  a = 0
 3 22  |  b = 1
```

ΑΡΑ ΑΠΟ ΤΟΥΣ ΕΞΗΣ ΕΛΕΓΧΟΥΣ CROSSVALIDATION ΕΠΙΛΕΓΩ ΤΟ ΠΙΟ ΣΗΜΑΝΤΙΚΟ-ΜΕΓΑΛΥΤΕΡΟ ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ:

J48                      ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 78%

**Perceptron            ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 90% ΑΡΑ ΘΑ ΕΠΙΛΕΞΩ ΑΥΤΟ ΠΟΥ ΕΙΝΑΙ ΤΟ ΠΕΡΙΣΣΟΤΕΡΟ ΣΩΣΤΟ**

BayesNet                ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 88%

## 5<sup>ο</sup> βήμα J48 (UseTrainingSet)

Test options

☒ Use training set

☐ Supplied test set    Set...

☐ Cross-validation    Folds 10

☐ Percentage split    % 66

More options...

(Nom) Bankrupt

Start Stop

Result list (right-click for options)

- 17:54:28 - trees.J48
- 18:01:45 - functions.MultilayerPerceptron
- 18:05:16 - bayes.BayesNet
- 18:08:41 - trees.J48

```
Classifier output

Time taken to build model: 0 seconds

==== Evaluation on training set ====

Time taken to test model on training data: 0 seconds

==== Summary ====
Correctly Classified Instances      48      96 %
Incorrectly Classified Instances    2       4 %
Kappa statistic                    0.92
Mean absolute error                0.0702
Root mean squared error            0.1873
Relative absolute error            14.0364 %
Root relative squared error        37.4651 %
Total Number of Instances          50

==== Detailed Accuracy By Class ====
          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0,920    0,000    1,000    0,920    0,958    0,923    0,974    0,974    0
          1,000    0,080    0,926    1,000    0,962    0,923    0,974    0,950    1
Weighted Avg.    0,960    0,040    0,963    0,960    0,960    0,923    0,974    0,962

==== Confusion Matrix ====
  a  b  <-- classified as
23  2  |  a = 0
 0 25  |  b = 1
```

6ο βήμα Perceptron (UseTrainingSet)

Test options

☒ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds

10

☐ Percentage split

%

66

More options...

(Nom) Bankrupt

Start

Stop

Result list (right-click for options)

17:54:28 - trees.J48

18:01:45 - functions.MultilayerPerceptron

18:05:16 - bayes.BayesNet

18:08:41 - trees.J48

18:11:54 - functions.MultilayerPerceptron

Classifier output

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances

47

94

%

Incorrectly Classified Instances

3

6

%

Kappa statistic

0.88

Mean absolute error

0.0913

Root mean squared error

0.2006

Relative absolute error

18.2595 %

Root relative squared error

40.124 %

Total Number of Instances

50

=== Detailed Accuracy By Class ===

TP Rate

FP Rate

Precision

Recall

F-Measure

MCC

ROC Area

PRC Area

Class

1,000

0,120

0,893

1,000

0,943

0,886

0,987

0,986

0

0,880

0,000

1,000

0,880

0,936

0,886

0,987

0,990

1

Weighted Avg.

0,940

0,060

0,946

0,940

0,940

0,886

0,987

0,988

=== Confusion Matrix ===

a

b

<-- classified as

25

0

|

a = 0

3

22

|

b = 1

7ο βήμα BayesNet (UseTrainingSet)

Test options

☒ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds

10

☐ Percentage split

%

66

More options...

(Nom) Bankrupt

Start

Stop

Result list (right-click for options)

17:54:28 - trees.J48

18:01:45 - functions.MultilayerPerceptron

18:05:16 - bayes.BayesNet

18:08:41 - trees.J48

18:11:54 - functions.MultilayerPerceptron

18:12:17 - bayes.BayesNet

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances

47

94

%

Incorrectly Classified Instances

3

6

%

Kappa statistic

0.88

Mean absolute error

0.0846

Root mean squared error

0.2148

Relative absolute error

16.9145 %

Root relative squared error

42.9613 %

Total Number of Instances

50

=== Detailed Accuracy By Class ===

TP Rate

FP Rate

Precision

Recall

F-Measure

MCC

ROC Area

PRC Area

Class

1,000

0,120

0,893

1,000

0,943

0,886

0,981

0,969

0

0,880

0,000

1,000

0,880

0,936

0,886

0,981

0,974

1

Weighted Avg.

0,940

0,060

0,946

0,940

0,940

0,886

0,981

0,971

=== Confusion Matrix ===

a

b

<-- classified as

25

0

|

a = 0

3

22

|

b = 1

ΑΡΑ ΑΠΟ ΤΟΥΣ ΕΞΗΣ ΕΛΕΓΧΟΥΣ UseTrainingSet ΕΠΙΛΕΓΩ ΤΟ ΠΙΟ ΣΗΜΑΝΤΙΚΟ-ΜΕΓΑΛΥΤΕΡΟ ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ:

**J48 ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 96% ΑΡΑ ΘΑ ΕΠΙΛΕΞΩ ΑΥΤΟ ΠΟΥ ΕΙΝΑΙ ΤΟ ΠΕΡΙΣΣΟΤΕΡΟ ΣΩΣΤΟ**

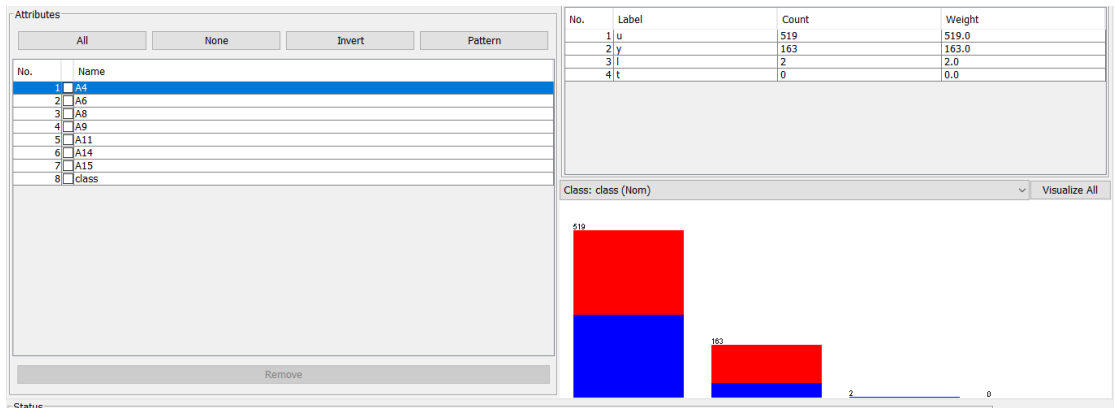
Perceptron ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 94%

BayesNet ΠΟΣΟΣΤΟ ΕΠΙΤΥΧΙΑΣ 94%

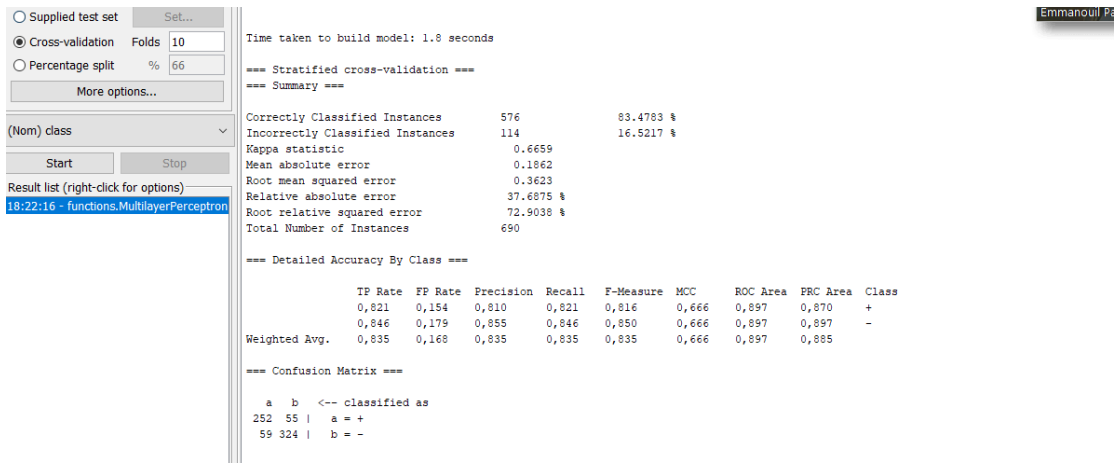


Άσκηση 2η στην Κατηγοριοποίηση Δεδομένων στο Weka(credit-approval.arff)

1<sup>ο</sup> Βήμα



2<sup>ο</sup> βήμα (perceptron) crossvalidation 83,4783%



## Άσκηση 3η στην Κατηγοριοποίηση Δεδομένων στο Weka (weather-nominal.arff)

### J48-Δένδρο Αποφάσεων (cross-validation)

Classifier

Choose **J48 -C 0.25 -M 2**

Test options

☐ Use training set

☐ Supplied test set

☒ Cross-validation Folds

☐ Percentage split %

(Nom) play

Result list (right-click for options)

09:03:04 - trees.J48

Classifier output

Size of the tree : 8

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	7	50	%
Incorrectly Classified Instances	7	50	%
Kappa statistic	-0.0426		
Mean absolute error	0.4167		
Root mean squared error	0.5984		
Relative absolute error	87.5	%	
Root relative squared error	121.2987	%	
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

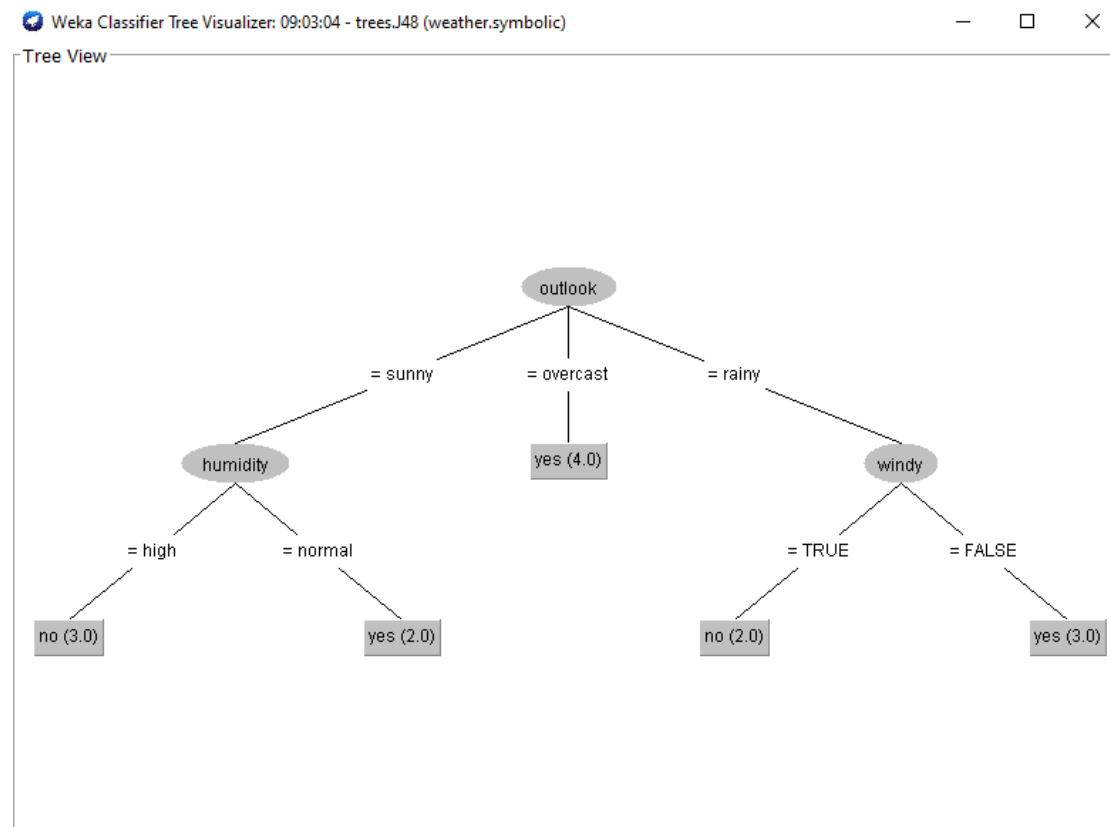
	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
	0,556	0,600	0,625	0,556	0,588	-0,043	0,633	0,758	yes
	0,400	0,444	0,333	0,400	0,364	-0,043	0,633	0,457	no
Weighted Avg.	0,500	0,544	0,521	0,500	0,508	-0,043	0,633	0,650	

=== Confusion Matrix ===

a b <-- classified as

5 4 | a = yes

3 2 | b = no



## Multilayer Perceptron (cross-validation)

Classifier

Choose **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

☐ Use training set

☐ Supplied Test on the same set that the classifier is trained on

☒ Cross-validation Folds

☐ Percentage split %

More options...

(Nom) play

Start Stop

Result list (right-click for options)

09:03:04 - trees.J48

09:05:20 - functions.MultilayerPerceptron

Classifier output

Input

Time taken to build model: 0.07 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	10	71.4286 %
Incorrectly Classified Instances	4	28.5714 %
Kappa statistic	0.3778	
Mean absolute error	0.287	
Root mean squared error	0.5268	
Relative absolute error	60.2616 %	
Root relative squared error	106.7798 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
	0,778	0,400	0,778	0,778	0,778	0,378	0,778	0,885	yes
	0,600	0,222	0,600	0,600	0,600	0,378	0,778	0,698	no
Weighted Avg.	0,714	0,337	0,714	0,714	0,714	0,378	0,778	0,818	

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

2 3 | b = no

Status

OK

Log

## Bayes (cross-validation)

Classifier

Choose **BayesNet** -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds

☐ Percentage split %

More options...

(Nom) play

Start Stop

Result list (right-click for options)

09:03:04 - trees.J48

09:05:20 - functions.MultilayerPerceptron

09:06:22 - bayes.BayesNet

Classifier output

LogScore AIC: -78.56181240647145

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	8	57.1429 %
Incorrectly Classified Instances	6	42.8571 %
Kappa statistic	-0.0244	
Mean absolute error	0.415	
Root mean squared error	0.4909	
Relative absolute error	87.1501 %	
Root relative squared error	99.5104 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
	0,778	0,800	0,636	0,778	0,700	-0,026	0,622	0,715	yes
	0,200	0,222	0,333	0,200	0,250	-0,026	0,622	0,591	no
Weighted Avg.	0,571	0,594	0,528	0,571	0,539	-0,026	0,622	0,671	

=== Confusion Matrix ===

a b <-- classified as

7 2 | a = yes

4 1 | b = no

Status

OK

Log

## J48-Δένδρο Αποφάσεων (training set)

Classifier

Choose J48 -C 0.25 -M 2

Test options

☒ Use training set

☐ Supplied test set Set...

☐ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) play

Start Stop

Result list (right-click for options)

09:03:04 - trees.J48

09:05:20 - functions.MultilayerPerceptron

09:06:22 - bayes.BayesNet

09:07:45 - trees.J48

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	14	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	14		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	FRC Area	Class
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	yes
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	no
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

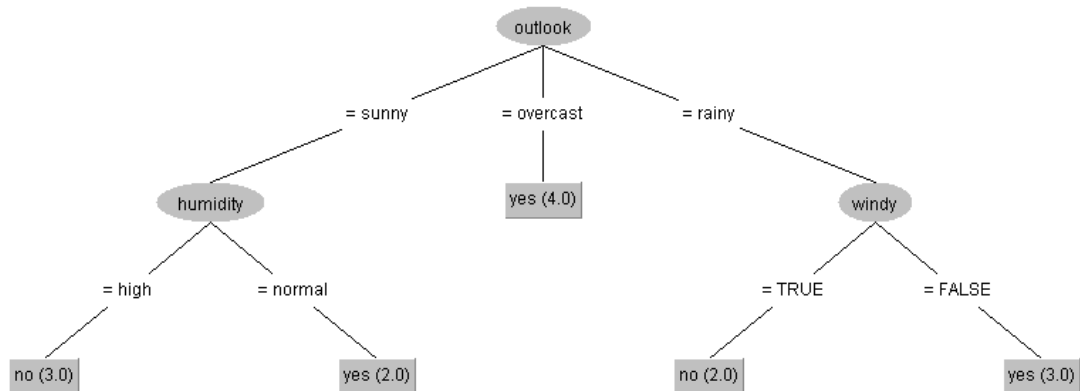
0 5 | b = no

Status OK

Log x 0

Weka Classifier Tree Visualizer: 09:07:45 - trees.J48 (weather.symbolic)

Tree View



## Multilayer Perceptron (training set)

Classifier

Choose **MultilayerPerceptron** -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a

Test options

☒ Use training set

☐ Supplied test set

☐ Cross-validation Folds

☐ Percentage split %

(Nom) play

Result list (right-click for options)

- 09:03:04 - trees.J48
- 09:05:20 - functions.MultilayerPerceptron
- 09:06:22 - bayes.BayesNet
- 09:07:45 - trees.J48
- 09:09:31 - functions.MultilayerPerceptron**

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	14	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0.0245	
Root mean squared error	0.0354	
Relative absolute error	5.2713 %	
Root relative squared error	7.3945 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===


	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	yes
0,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	1,000	no
Weighted Avg.	1,000	0,000	1,000	1,000	1,000	1,000	1,000	1,000	

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

0 5 | b = no

Status OK   x 0

## Bayes (training set)

Classifier

Choose **BayesNet** -D -Q weka.classifiers.bayes.net.search.local.K2 -- -P 1 -S BAYES -E weka.classifiers.bayes.net.estimate.SimpleEstimator -- -A 0.5

Test options

☒ Use training set

☐ Supplied test set

☐ Cross-validation Folds

☐ Percentage split %

(Nom) play

Result list (right-click for options)

- 09:03:04 - trees.J48
- 09:05:20 - functions.MultilayerPerceptron
- 09:06:22 - bayes.BayesNet
- 09:07:45 - trees.J48
- 09:09:31 - functions.MultilayerPerceptron
- 09:10:13 - bayes.BayesNet**

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances	13	92.8571 %
Incorrectly Classified Instances	1	7.1429 %
Kappa statistic	0.8372	
Mean absolute error	0.2615	
Root mean squared error	0.3242	
Relative absolute error	56.3272 %	
Root relative squared error	67.6228 %	
Total Number of Instances	14	

=== Detailed Accuracy By Class ===


	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1,000	0,200	0,900	1,000	0,947	0,849	0,849	0,933	0,963	yes
0,800	0,000	1,000	0,800	0,889	0,849	0,849	0,933	0,925	no
Weighted Avg.	0,929	0,129	0,936	0,929	0,926	0,849	0,933	0,949	

=== Confusion Matrix ===

a b <-- classified as

9 0 | a = yes

1 4 | b = no

Status OK   x 0

Έχοντας εξερευνήσει σε αυτή τη νέα βάση(weather\_nominal) όπως και στην προηγούμενη άσκηση διαφορετικούς τύπους αλγορίθμων κατηγοριοποίησης J48,PERCEPTRON και BayesNET, παρατηρούμε ότι ο Perceptron μας δίνει την μεγαλύτερη ακρίβεια αποτελεσμάτων με 71.42%,ενώ η μέθοδος η οποία χρησιμοποιείται είναι η (cross-validation).

Αντιθέτως, εάν χρησιμοποιήσουμε την μέθοδο (training set) προβλέπεται ακρίβεια 100% με τους αλγορίθμους Trees J48 και Perceptron.

Συμπερασματικά παρατηρώ ότι ο Perceptron, αυτό το νευρωνικό δίκτυο, μας δίνει τις υψηλότερες τιμές και στις 2 μεθόδους που έκανα χρήση, δηλαδή σε αυτή της Cross Validtion και του Training Set.

## Άσκηση στη Συσταδοποίηση Δεδομένων με το Weka (Wholesale CustomersData.csv)

### Βήμα 1ο

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize | Auto-WEKA | Interactive Parallel Coordinates Plot | RConsole

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **None** [Apply] [Stop]

**Current relation**  
Relation: Wholesale customers data  
Instances: 440  
Attributes: 8  
Sum of weights: 440

**Attributes**  
[All] [None] [Invert] [Pattern]

No.	Name
1	<input checked="" type="checkbox"/> Channel
2	<input type="checkbox"/> Region
3	<input type="checkbox"/> Fresh
4	<input type="checkbox"/> Milk
5	<input type="checkbox"/> Grocery
6	<input type="checkbox"/> Frozen
7	<input type="checkbox"/> Detergents_Paper
8	<input type="checkbox"/> Delicassen

[Remove]

**Selected attribute**  
Name: Channel  
Missing: 0 (0%)  
Distinct: 2  
Type: Numeric  
Unique: 0 (0%)

Statistic	Value
Minimum	1
Maximum	2
Mean	1.323
StdDev	0.468

Class: Delicassen (Num) [Visualize All]

298 105 37

1 0 1.5 2

**Status**  
Transferred Wholesale customers data into R as "rdata" [Log] x0

### Βήμα 2ο

**Cluster mode**

☒ Use training set  
☐ Supplied test set [Set...]  
☐ Percentage split % 66  
☐ Classes to clusters evaluation  
(Num) Delicassen  
☒ Store clusters for visualization

[Ignore attributes]

[Start] [Stop]

**Result list (right-click for options)**

18:34:57 - SimpleKMeans

**Clusterer output**

Final cluster centroids:

Attribute	Full Data (440.0)	Cluster#		
		0 (298.0)	1 (105.0)	2 (37.0)
Channel	1.3227	1	2	2
Region	2.5432	2.5101	3	1.5135
Fresh	12000.2977	13475.5604	9831.5048	6273.1351
Milk	5796.2659	3451.7248	10981.0095	9965.8649
Grocery	7951.2773	3962.1376	15953.8095	17370.1351
Frozen	3071.9318	3748.2517	1513.2	2048.2432
Detergents_Paper	2881.4932	790.5604	6899.2381	8320.2703
Delicassen	1524.8705	1415.9564	1826.2095	1546.9189

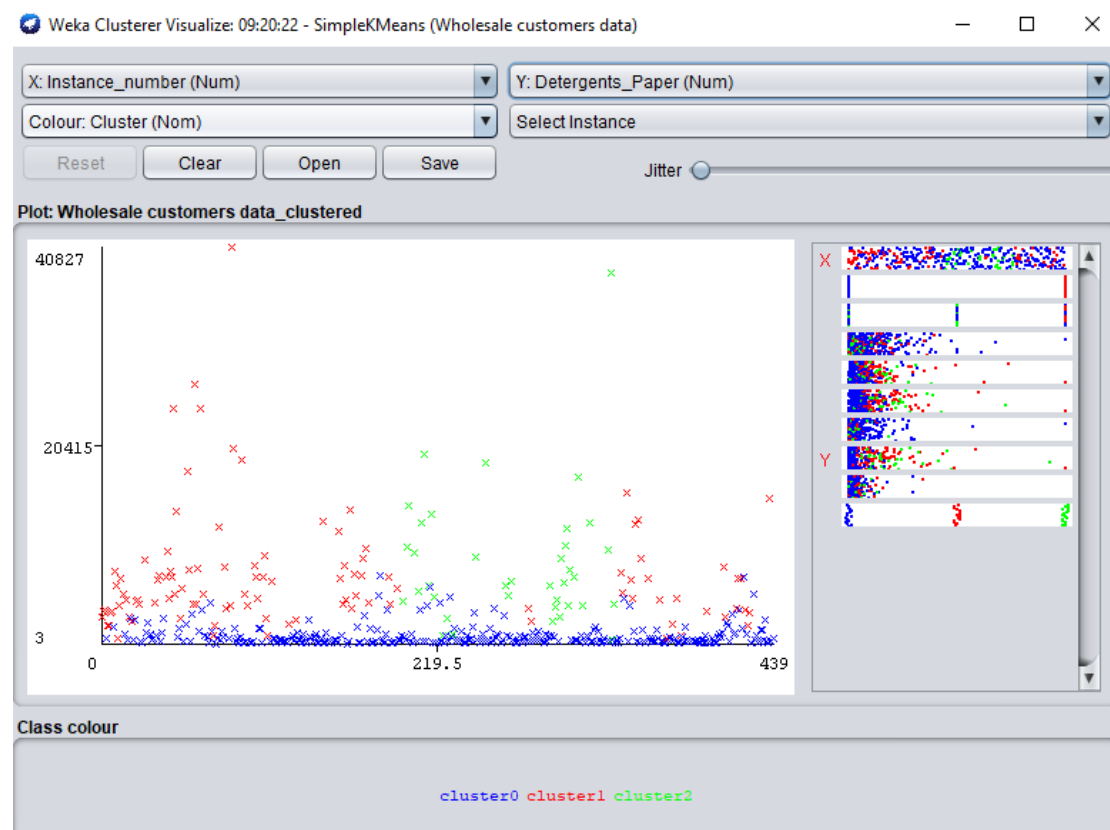
Time taken to build model (full training data) : 0.03 seconds

=== Model and evaluation on training set ===

Clustered Instances

Cluster#	Instances	Percentage
0	298	( 68%)
1	105	( 24%)
2	37	( 8%)

### Βήμα 3<sup>ο</sup>





## Άσκηση 1η στους κανόνες συσχέτισης(supermarket.arff)

### Βήμα 1<sup>ο</sup>

```
Result list (right-click for op... 18:47:51 - Apriori
bread and cake=t milk-cream=t fruit=t vegetables=t total=high 725
baking needs=t biscuits=t frozen foods=t milk-cream=t fruit=t 715
baking needs=t biscuits=t frozen foods=t milk-cream=t vegetables=t 705
baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t 796
baking needs=t biscuits=t milk-cream=t fruit=t vegetables=t 753
baking needs=t frozen foods=t milk-cream=t fruit=t vegetables=t 778
juice-sat-cord-ms=t biscuits=t frozen foods=t fruit=t vegetables=t 698
biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t 707
biscuits=t frozen foods=t milk-cream=t fruit=t vegetables=t 760

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

Large Itemsets L(6):
bread and cake=t baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t 716

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)
```

BEST RULES FOUND: Παρατηρώ το εξής για τους κανόνες συσχέτισης με το περισσότερο confidence:

1) Οι καταναλωτές που έχουν αγοράσει μπισκότο, κατεψυγμένο φαγητό , φρούτα και έχουν πληρώσει μεγάλο σύνολο, ανέρχονται σε 788 άτομα. Από αυτά τα 788 άτομα, τα 723 φαίνεται να έχουν αγοράσει ψωμί και κέϊκ.

2) Οι καταναλωτές που έχουν αγοράσει προϊόντα για baking , μπισκότα, φρούτα και έκανα υψηλό λογαριασμό είναι 760 άτομα. Από αυτά τα 760 άτομα φαίνεται πως αγόρασαν ψωμί και κέϊκ τα 696 άτομα.

3) Οι καταναλωτές που έχουν αγοράσει προϊόντα για baking, κατεψυγμένα, φρούτα και έχουν πληρώσει υψηλό λογαριασμό είναι 770 άτομα. Από αυτά τα 770 άτομα φαίνεται πως αγόρασαν τα 705 άτομα ψωμί και κέϊκ.

...

..

..

9) Οι καταναλωτές που έχουν αγοράσει προϊόντα για κατεψυγμένα, φρούτα, λαχανικά και έχουν πληρώσει υψηλό λογαριασμό είναι 834 άτομα. Από αυτά τα 834 άτομα φαίνεται πως αγόρασαν τα 757 άτομα ψωμί και κέϊκ.

10) Οι καταναλωτές που έχουν αγοράσει προϊόντα για κατεψυγμένα και φρούτα, έχουν πληρώσει υψηλό λογαριασμό είναι 969 άτομα. Από αυτά τα 969 άτομα φαίνεται πως αγόρασαν τα 877 άτομα ψωμί και κέϊκ.

## 1<sup>ο</sup> βήμα

Start

Stop

Result list (right-click for op...  
18:47:51 - Apriori  
18:59:12 - Apriori

=====

Minimum support: 0.5 (2314 instances)  
Minimum metric <confidence>: 0.9  
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Large Itemsets L(1):  
bread and cake=t 3330  
baking needs=t 2795  
juice-sat-cord-ms=t 2463  
biscuits=t 2605  
frozen foods=t 2717  
party snack foods=t 2330  
milk-cream=t 2939  
fruit=t 2962  
vegetables=t 2961  
total=low 2948

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

Large Itemsets L(2):  
bread and cake=t milk-cream=t 2337  
bread and cake=t fruit=t 2325

Best rules found:

**Απάντηση:** Δεν βρίσκεται κάποιος κανόνας, διότι ο συνδυασμός οι τιμές υποστήριξης και εμπιστοσύνης είναι υπερβολικά μεγάλες, εξίσου και ο συνδυασμός τους, και δεν υπάρχουν τέτοιοι κανόνες.

## 2<sup>ο</sup> βήμα

☐ Store output for visualization

Start

Stop

Result list (right-click for op...  
18:47:51 - Apriori  
18:59:12 - Apriori  
19:01:03 - Apriori

Associator output

Minimum metric <confidence>: 0.7  
Number of cycles performed: 10

Generated sets of large itemsets:

Size of set of large itemsets L(1): 10

Large Itemsets L(1):  
bread and cake=t 3330  
baking needs=t 2795  
juice-sat-cord-ms=t 2463  
biscuits=t 2605  
frozen foods=t 2717  
party snack foods=t 2330  
milk-cream=t 2939  
fruit=t 2962  
vegetables=t 2961  
total=low 2948

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

Large Itemsets L(2):  
bread and cake=t milk-cream=t 2337  
bread and cake=t fruit=t 2325

Best rules found:

1. milk-cream=t 2939 ==> bread and cake=t 2337 <conf:(0.8)> lift:(1.1) lev:(0.05) [221] conv:(1.37)  
2. fruit=t 2962 ==> bread and cake=t 2325 <conf:(0.78)> lift:(1.09) lev:(0.04) [193] conv:(1.3)  
3. bread and cake=t 3330 ==> milk-cream=t 2337 <conf:(0.7)> lift:(1.1) lev:(0.05) [221] conv:(1.22)

### Απάντηση:

- Βρίσκονται 3 κανόνες. Ο 1<sup>ος</sup> μας αναφέρει από τους 2939 καταναλωτές που αποτελούν το σύνολο των καταναλωτών που αγοράζουν κρέμα γάλακτος, το 80% από το σύνολο των καταναλωτών που αγοράζουν κρέμα γάλακτος που είναι 2337 άτομα, παράλληλα αγοράζουν ψωμί και κέικ.
- Ο 2<sup>ος</sup> κανόνας μας αναφέρει πως οι 2962 καταναλωτές που αποτελούν το σύνολο των καταναλωτών που αγοράζουν φρούτα, το 78% από το σύνολο των καταναλωτών που αγοράζουν φρούτα που είναι 2325 άτομα, παράλληλα αγοράζουν ψωμί και κέικ.
- Ο 3<sup>ος</sup> κανόνας μας λέει πως οι 3330 καταναλωτές που αποτελούν το σύνολο των καταναλωτών που αγοράζουν ψωμί και κέικ, το 70% από το σύνολο των καταναλωτών που αγοράζουν ψωμί και κέικ που είναι 2337 άτομα, παράλληλα αγοράζουν και κρέμα γάλακτος.

### 3<sup>ο</sup> βήμα

Associator

Choose **Apriori** -I -N 10 -T 0 -C 0.5 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 21

☐ Store output for visualization

Start Stop

Result list (right-click for op...)

18:47:51 - Apriori  
18:59:12 - Apriori  
19:01:03 - Apriori  
07:43:35 - Apriori

Associator output

```
0 465
frozen foods=t milk-cream=t fruit=t vegetables=t 1057
0 497
frozen foods=t fruit=t vegetables=t total=high 834
0 463

Size of set of large itemsets L(5): 4

Large Itemsets L(5):
bread and cake=t baking needs=t biscuits=t frozen foods=t vegetables=t 848
0 463
bread and cake=t baking needs=t biscuits=t fruit=t vegetables=t 896
0 469
bread and cake=t baking needs=t frozen foods=t fruit=t vegetables=t 910
0 472
bread and cake=t biscuits=t frozen foods=t fruit=t vegetables=t 929
0 492

Best rules found:

1. sauces-gravy-pkle=t margarine=t total=high 767 ==> canned vegetables=t 467 conf:(0.61)
2. biscuits=t margarine=t total=high 856 ==> canned vegetables=t 512 conf:(0.6)
3. biscuits=t sauces-gravy-pkle=t total=high 864 ==> canned vegetables=t 514 conf:(0.59)
4. bread and cake=t biscuits=t tissues-paper prd=t total=high 788 ==> canned vegetables=t 467 conf:(0.59)
5. bread and cake=t sauces-gravy-pkle=t frozen foods=t total=high 791 ==> canned vegetables=t 468 conf:(0.59)
6. sauces-gravy-pkle=t tissues-paper prd=t total=high 830 ==> canned vegetables=t 491 conf:(0.59)
7. biscuits=t sauces-gravy-pkle=t margarine=t 877 ==> canned vegetables=t 516 conf:(0.59)
8. biscuits=t tissues-paper prd=t total=high 891 ==> canned vegetables=t 522 conf:(0.59)
9. frozen foods=t margarine=t total=high 870 ==> canned vegetables=t 509 conf:(0.59)
10. tissues-paper prd=t margarine=t total=high 816 ==> canned vegetables=t 475 conf:(0.58)
```

Οι κανόνες που έχουν προκύψει περιγράφουν τη συσχέτιση της αγοράς κονσερβοποιημένων λαχανικών με την αγορά άλλων συγκεκριμένων προϊόντων.

Best rules found:

1<sup>ος</sup> κανόνας: Αυτοί που αγοράζουν sauces-gravy-pkle, μαργαρίνη και έχουν πληρώσει μεγάλο λογαριασμό είναι 767 άτομα, ενώ αγοράζουν και λαχανικά σε κονσέρβα το 61% , δηλαδή οι 467 από τους 767.

2<sup>ος</sup> κανόνας: Αυτοί που αγοράζουν μπισκότα, μαργαρίνη και έχουν πληρώσει μεγάλο λογαριασμό είναι 856 άτομα, ενώ αγοράζουν και λαχανικά σε κονσέρβα το 60% , δηλαδή οι 512 από τους 856.

3<sup>ος</sup> κανόνας: Αυτοί που αγοράζουν μπισκότα, sauces-gravy-pkle και έχουν πληρώσει υψηλό λογαριασμό είναι 864 άτομα, ενώ από αυτούς τους 864 οι 514 που αποτελούν το 59% έχουν αγοράσει και λαχανικά σε κονσέρβα.

..

..

10<sup>ος</sup> κανόνας: Αυτοί που αγοράζουν tissues-paper , μαργαρίνη και έχουν πληρώσει υψηλό λογαριασμό είναι 816, ενώ το 58% δηλαδή 475 καταναλωτές, έχουν αγοράσει και συσκευασμένα λαχανικά.

#### 4<sup>ο</sup> βήμα:

Associator

Choose Apriori -I -N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -A -c 217

☐ Store output for visualization

Start Stop

Result list (right-click for op...)

18:47:51 - Apriori  
18:59:12 - Apriori  
19:01:03 - Apriori  
07:43:35 - Apriori  
07:33:34 - Apriori

Associator output

Size of set of large itemsets L(6): 9

Large Itemsets L(6):

bread and cake=t baking needs=t biscuits=t frozen foods=t fruit=t vegetables=t 716  
1 521  
bread and cake=t baking needs=t biscuits=t milk-cream=t fruit=t vegetables=t 668  
1 467  
bread and cake=t baking needs=t frozen foods=t tissues-paper prd=t fruit=t vegetables=t 612  
1 476  
bread and cake=t baking needs=t frozen foods=t milk-cream=t fruit=t vegetables=t 690  
1 480  
bread and cake=t juice-sat-cord-ms=t biscuits=t frozen foods=t fruit=t vegetables=t 629  
1 465  
bread and cake=t biscuits=t sauces-gravy-pkle=t frozen foods=t fruit=t vegetables=t 596  
1 468  
bread and cake=t biscuits=t frozen foods=t party snack foods=t fruit=t vegetables=t 640  
1 478  
bread and cake=t biscuits=t frozen foods=t tissues-paper prd=t fruit=t vegetables=t 598  
1 475  
bread and cake=t biscuits=t frozen foods=t milk-cream=t fruit=t vegetables=t 688  
1 486

Best rules found:

1. baking needs=t biscuits=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t 574 ==> total=high 470 conf:(0.82)
2. bread and cake=t biscuits=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t 600 ==> total=high 491 conf:(0.82)
3. bread and cake=t baking needs=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t 620 ==> total=high 506 conf:(0.82)
4. bread and cake=t baking needs=t biscuits=t sauces-gravy-pkle=t tissues-paper prd=t 595 ==> total=high 483 conf:(0.81)
5. bread and cake=t biscuits=t sauces-gravy-pkle=t tissues-paper prd=t vegetables=t 583 ==> total=high 469 conf:(0.8)
6. bread and cake=t sauces-gravy-pkle=t frozen foods=t tissues-paper prd=t vegetables=t 610 ==> total=high 490 conf:(0.8)

Προκύπτουν έξι κανόνες, οι οποίοι συσχετίζουν συγκεκριμένα προϊόντα με πωλήσεις υψηλής αξίας.

1<sup>ος</sup> κανόνας: Οι καταναλωτές που αγοράζουν baking needs,μπισκότα,sauces-gravy-pkle, κατεψυγμένα, tissues και prd είναι 574, ενώ από αυτούς τους 574 οι 470 που αποτελούν το 82%, πραγματοποιούν επίσης και υψηλό λογαριασμό σε αγορές

## Άσκηση 2<sup>η</sup> Κανόνες Συσχέτισης(vote.arff)

### Βήμα 1<sup>ο</sup>,2<sup>ο</sup>,3<sup>ο</sup>

Apriori

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

☐ Store output for visualization

Start Stop

Result list (right-click for op...)

08:00:59 - Apriori

Apriori output

=====  
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  
Size of set of large itemsets L(2): 17  
Size of set of large itemsets L(3): 6  
Size of set of large itemsets L(4): 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. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-contras=y 198 ==> Class=democrat 198 <conf:(1)> lift:(1.63) lev:(0.18) [76] con  
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210 <conf:(1)> lift:(1.62) lev:(0.19) [80] conv:(40.74)  
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 <conf:(1)> lift:(1.62) lev:(0.18) [77] conv:(39.01)  
5. physician-fee-freeze=n 247 ==> Class=democrat 245 <conf:(0.99)> lift:(1.62) lev:(0.21) [93] conv:(31.8)  
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 <conf:(0.98)> lift:(1.77) lev:(0.2) [85] conv:(22.18)  
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 <conf:(0.98)> lift:(1.76) lev:(0.2) [88] conv:(18.46)  
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198 <conf:(0.98)> lift:(1.72) lev:(0.19) [82]  
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 <conf:(0.97)> lift:(1.57) lev:(0.17) [71] conv:(9.85)  
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210 <conf:(0.96)> lift:(1.7) lev:(0.2) [86] conv:(10.47)

Status  
OK

Log x 0

### Βήμα 4<sup>ο</sup>

Προκύπτουν 10 κανόνες που συσχετίζουν τον τρόπο ψηφοφορίας και το κόμμα από το οποίο προέρχεται το μέλος του Κογκρέσου. Οι 4 πρώτοι κανόνες όπου και έχουν confidence = 1, , δηλαδή όλοι σημαίνει ότι 100% συσχετίζεται ο τρόπος ψηφοφορίας και το κόμμα από το οποίο προέρχεται το μέλος του Κογκρέσου.

Με ένα πιο ενδελεχή έλεγχο έχουμε τις εξής αναλύσεις των κανόνων:

**1<sup>ος</sup> κανόνας:** Όποιος βουλευτής του Κογκρέσου έχει υπερψηφίσει την υιοθέτηση του προϋπολογισμού και καταψήφισε το πάγωμα αμοιβών υγειονομικού, φαίνεται πως προέρχεται από το Δημοκρατικό κόμμα (219/219 και αντίστοιχα το 100%).

**2<sup>ος</sup> κανόνας:** Όποιος βουλευτής του Κογκρέσου έχει υπερψηφίσει την υιοθέτηση του προϋπολογισμού, καταψήφισε το πάγωμα αμοιβών υγειονομικού και παράλληλα υπερψήφισε την βοήθεια στην Νικαράγουα, φαίνεται πως προέρχεται από το Δημοκρατικό κόμμα (198/198 και αντίστοιχα το 100%).

**3<sup>ος</sup> κανόνας:** Όποιος βουλευτής του Κογκρέσου καταψήφισε το πάγωμα αμοιβών υγειονομικού προσωπικού και υπερψήφισε την οικονομική βοήθεια σε αντικυβερνητικές ομάδες στη Νικαράγουα, φαίνεται πως προέρχεται σχεδόν 100% από το Δημοκρατικό κόμμα (210/211).

**4<sup>ος</sup> κανόνας:** Όποιος βουλευτής καταψήφισε το πάγωμα αμοιβών υγειονομικού προσωπικού και καταψήφισε την χρηματοδότηση εκπαιδευτικών προγραμμάτων, φαίνεται ότι προέρχεται 100% προέρχεται από το Δημοκρατικό κόμμα (201/202).

**5<sup>ος</sup> κανόνας:** Όποιος βουλευτής καταψήφισε το πάγωμα αμοιβών των υγειονομικών, φαίνεται ότι προέρχεται κατά 99% από το δημοκρατικό κόμμα (245/247).

- .
- .
- .
- .

**10<sup>ος</sup> κανόνας:** Όποιος βουλευτής υπερψήφισε την βοήθεια σε αντικυβερνητικές ομάδες της Νικαράγουα και προέρχεται από το Δημοκρατικό κόμμα, φαίνεται τότε πρόκειται να το πάγωμα αμοιβών των υγειονομικών ( 210/218, δηλαδή 96%).

## Βήμα 4<sup>ο</sup>,5<sup>ο</sup>

Associator

Choose **Apriori** -N 101 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S 1.0 -c 1

☒ Store output for visualization

Associator output

Start Stop

Result list (right-click for op...)

08:00:59 - Apriori

08:19:32 - Apriori

Associator output

73. mx-missile=y Class=democrat 188 ==> physician-fee-freeze=n 179 <conf:(0.95)> lift:(1.68) lev:(0.17) [72] conv:(8.13)

74. mx-missile=y Class=democrat 188 ==> aid-to-nicaraguan-contras=y 179 <conf:(0.95)> lift:(1.71) lev:(0.17) [74] conv:(8.34)

75. adoption-of-the-budget-resolution=y el-salvador-aid=n 188 ==> physician-fee-freeze=n aid-to-nicaraguan-contras=y 179 <conf:(0.95)> lift:(1.96) lev:(0.2) [75] conv:(8.38)

76. el-salvador-aid=n anti-satellite-test-ban=y 188 ==> aid-to-nicaraguan-contras=y Class=democrat 179 <conf:(0.95)> lift:(1.9) lev:(0.19) [84] conv:(9.38)

77. adoption-of-the-budget-resolution=y el-salvador-aid=n 188 ==> physician-fee-freeze=n aid-to-nicaraguan-contras=y Class=democrat 179 <conf:(0.95)> lift:(1.96) lev:(0.2) [75] conv:(8.38)

78. el-salvador-aid=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 184 ==> physician-fee-freeze=n 175 <conf:(0.95)> lift:(1.67) lev:(0.16) [70] conv:(7.68)

79. 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)

80. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y Class=democrat 197 <conf:(0.95)> lift:(1.89) lev:(0.21) [92] conv:(8.65)

81. anti-satellite-test-ban=y aid-to-nicaraguan-contras=y Class=democrat 189 ==> el-salvador-aid=n 179 <conf:(0.95)> lift:(1.98) lev:(0.2) [88] conv:(8.97)

82. el-salvador-aid=n anti-satellite-test-ban=y 188 ==> physician-fee-freeze=n 178 <conf:(0.95)> lift:(1.67) lev:(0.16) [71] conv:(7.39)

83. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 185 ==> el-salvador-aid=n 175 <conf:(0.95)> lift:(1.98) lev:(0.2) [86] conv:(8.38)

84. el-salvador-aid=y mx-missile=n 184 ==> religious-groups-in-schools=y 174 <conf:(0.95)> lift:(1.51) lev:(0.14) [58] conv:(6.27)

85. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y Class=democrat 184 ==> adoption-of-the-budget-resolution=y 174 <conf:(0.95)> lift:(1.96) lev:(0.2) [75] conv:(8.38)

86. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y Class=democrat 184 ==> physician-fee-freeze=n aid-to-nicaraguan-contras=y 174 <conf:(0.95)> lift:(1.96) lev:(0.2) [75] conv:(8.38)

87. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y Class=democrat 184 ==> el-salvador-aid=n 174 <conf:(0.95)> lift:(1.98) lev:(0.2) [86] conv:(8.38)

88. el-salvador-aid=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 184 ==> physician-fee-freeze=n Class=democrat 174 <conf:(0.95)> lift:(1.68) lev:(0.17) [72] conv:(8.13)

89. anti-satellite-test-ban=y Class=democrat 200 ==> aid-to-nicaraguan-contras=y 189 <conf:(0.94)> lift:(1.7) lev:(0.18) [77] conv:(7.39)

90. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y 215 ==> Class=democrat 203 <conf:(0.94)> lift:(1.54) lev:(0.16) [71] conv:(6.39)

91. education-spending=n Class=democrat 213 ==> physician-fee-freeze=n 201 <conf:(0.94)> lift:(1.66) lev:(0.18) [80] conv:(7.08)

92. physician-fee-freeze=n anti-satellite-test-ban=y Class=democrat 195 ==> aid-to-nicaraguan-contras=y 184 <conf:(0.94)> lift:(1.7) lev:(0.17) [75] conv:(7.2)

93. physician-fee-freeze=n aid-to-nicaraguan-contras=y Class=democrat 210 ==> adoption-of-the-budget-resolution=y 198 <conf:(0.94)> lift:(1.62) lev:(0.17) [75] conv:(6.84)

94. el-salvador-aid=n anti-satellite-test-ban=y 188 ==> physician-fee-freeze=n Class=democrat 177 <conf:(0.94)> lift:(1.67) lev:(0.16) [71] conv:(6.84)

95. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> physician-fee-freeze=n 192 <conf:(0.94)> lift:(1.66) lev:(0.18) [76] conv:(6.78)

96. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 187 ==> Class=democrat 176 <conf:(0.94)> lift:(1.53) lev:(0.14) [69] conv:(6.27)

97. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 185 ==> adoption-of-the-budget-resolution=y 174 <conf:(0.94)> lift:(1.62) lev:(0.17) [75] conv:(7.2)

98. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 185 ==> adoption-of-the-budget-resolution=y Class=democrat 174 <conf:(0.94)> lift:(1.62) lev:(0.17) [75] conv:(7.2)

99. physician-fee-freeze=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 185 ==> el-salvador-aid=n Class=democrat 174 <conf:(0.94)> lift:(2.05) lev:(0.2) [86] conv:(8.38)

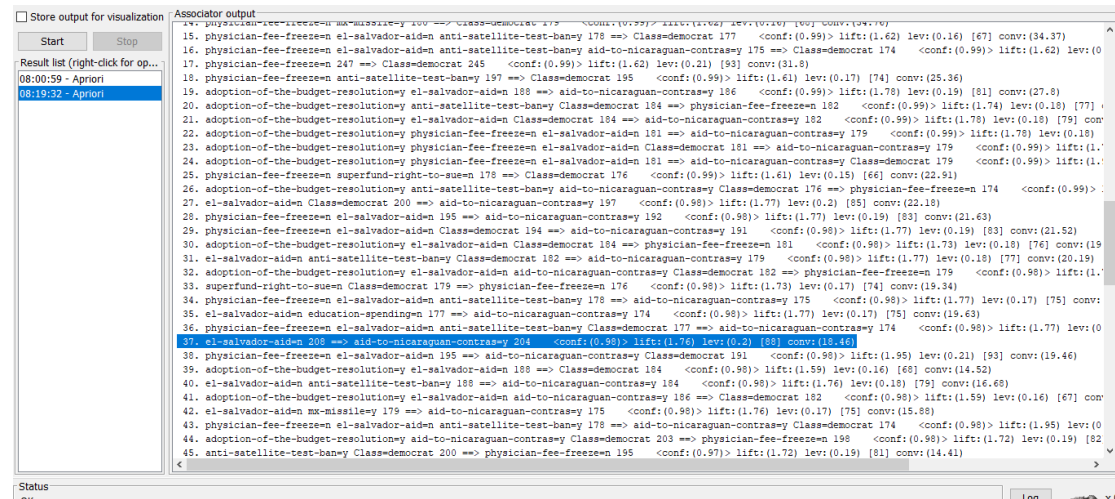
100. physician-fee-freeze=n anti-satellite-test-ban=y 197 ==> aid-to-nicaraguan-contras=y 185 <conf:(0.94)> lift:(1.69) lev:(0.17) [75] conv:(6.72)

101. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> adoption-of-the-budget-resolution=y 198 <conf:(0.94)> lift:(1.61) lev:(0.17) [75] conv:(6.31)

Παρατηρώ στο 4<sup>ο</sup> βήμα πως ο αριθμός των κανόνων γίνεται να έχει τιμές και μεγαλύτερες του 100, όμως εκείνοι οι κανόνες με το μεγαλύτερο Confidence πάντα θα είναι οι καλύτεροι κανόνες.

Οι 7 πρώτοι κανόνες αποτελούνται από Confidence 1 και αντίστοιχα το 100% του συνόλου, ενώ οι κανόνες 8 και 9 έχουν Confidence 1 αλλά είναι (ο 8<sup>ος</sup> 210/211) και αντίστοιχα (ο 9<sup>ος</sup> 201/202).

## Βήμα 6<sup>ο</sup>



Associator output

Start Stop

Result list (right-click for op...)

08:00:59 - Apriori

08:19:32 - Apriori

15. physician-fee-freeze=n el-salvador-aid=n anti-satellite-test-ban=y 178 ==> Class=democrat 177 <conf:(0.99)> lift:(1.62) lev:(0.16) [67] conv:(34.37)

16. physician-fee-freeze=n el-salvador-aid=n anti-satellite-test-ban=y aid-to-nicaraguan-contras=y 178 ==> Class=democrat 174 <conf:(0.99)> lift:(1.62) lev:(0.16) [67] conv:(34.37)

17. physician-fee-freeze=n 247 ==> Class=democrat 245 <conf:(0.99)> lift:(1.62) lev:(0.21) [83] conv:(31.8)

18. physician-fee-freeze=n anti-satellite-test-ban=y 197 ==> Class=democrat 195 <conf:(0.99)> lift:(1.61) lev:(0.17) [74] conv:(25.36)

19. adoption-of-the-budget-resolution=y el-salvador-aid=n 188 ==> aid-to-nicaraguan-contras=y 186 <conf:(0.99)> lift:(1.78) lev:(0.19) [81] conv:(27.8)

20. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y Class=democrat 184 ==> physician-fee-freeze=n 182 <conf:(0.99)> lift:(1.74) lev:(0.18) [77] conv:(26.8)

21. adoption-of-the-budget-resolution=y el-salvador-aid=n Class=democrat 184 ==> aid-to-nicaraguan-contras=y 182 <conf:(0.99)> lift:(1.78) lev:(0.18) [79] conv:(27.8)

22. adoption-of-the-budget-resolution=y physician-fee-freeze=n el-salvador-aid=n 181 ==> aid-to-nicaraguan-contras=y 179 <conf:(0.99)> lift:(1.78) lev:(0.18) [77] conv:(26.8)

23. adoption-of-the-budget-resolution=y physician-fee-freeze=n el-salvador-aid=n Class=democrat 181 ==> aid-to-nicaraguan-contras=y 179 <conf:(0.99)> lift:(1.78) lev:(0.18) [77] conv:(26.8)

24. adoption-of-the-budget-resolution=y physician-fee-freeze=n el-salvador-aid=n 181 ==> aid-to-nicaraguan-contras=y Class=democrat 179 <conf:(0.99)> lift:(1.78) lev:(0.18) [77] conv:(26.8)

25. physician-fee-freeze=n superfund-right-to-sue=n 178 ==> Class=democrat 176 <conf:(0.98)> lift:(1.61) lev:(0.15) [66] conv:(22.91)

26. adoption-of-the-budget-resolution=y anti-satellite-test-ban=y aid-to-nicaraguan-contras=y Class=democrat 176 ==> physician-fee-freeze=n 174 <conf:(0.99)> lift:(1.78) lev:(0.19) [81] conv:(27.8)

27. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 <conf:(0.98)> lift:(1.77) lev:(0.2) [85] conv:(22.18)

28. physician-fee-freeze=n el-salvador-aid=n 195 ==> aid-to-nicaraguan-contras=y 192 <conf:(0.98)> lift:(1.77) lev:(0.19) [83] conv:(21.63)

29. physician-fee-freeze=n el-salvador-aid=n Class=democrat 194 ==> aid-to-nicaraguan-contras=y 191 <conf:(0.98)> lift:(1.77) lev:(0.19) [83] conv:(21.63)

30. adoption-of-the-budget-resolution=y el-salvador-aid=n Class=democrat 184 ==> physician-fee-freeze=n 181 <conf:(0.98)> lift:(1.73) lev:(0.18) [76] conv:(21.52)

31. el-salvador-aid=n anti-satellite-test-ban=y Class=democrat 182 ==> aid-to-nicaraguan-contras=y 179 <conf:(0.98)> lift:(1.77) lev:(0.18) [77] conv:(20.19)

32. adoption-of-the-budget-resolution=y el-salvador-aid=n aid-to-nicaraguan-contras=y Class=democrat 182 ==> physician-fee-freeze=n 179 <conf:(0.98)> lift:(1.77) lev:(0.18) [77] conv:(20.19)

33. superfund-right-to-sue=n Class=democrat 179 ==> physician-fee-freeze=n 176 <conf:(0.98)> lift:(1.73) lev:(0.17) [74] conv:(19.34)

34. physician-fee-freeze=n el-salvador-aid=n anti-satellite-test-ban=y 178 ==> aid-to-nicaraguan-contras=y 175 <conf:(0.98)> lift:(1.77) lev:(0.17) [75] conv:(19.34)

35. el-salvador-aid=n education-spending=n 177 ==> aid-to-nicaraguan-contras=y 174 <conf:(0.98)> lift:(1.77) lev:(0.17) [75] conv:(19.34)

36. physician-fee-freeze=n el-salvador-aid=n anti-satellite-test-ban=y Class=democrat 177 ==> aid-to-nicaraguan-contras=y 174 <conf:(0.98)> lift:(1.77) lev:(0.17) [75] conv:(19.34)

37. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 <conf:(0.98)> lift:(1.76) lev:(0.2) [88] conv:(19.46)

38. physician-fee-freeze=n el-salvador-aid=n 195 ==> aid-to-nicaraguan-contras=y Class=democrat 191 <conf:(0.98)> lift:(1.95) lev:(0.21) [93] conv:(19.46)

39. adoption-of-the-budget-resolution=y el-salvador-aid=n 188 ==> Class=democrat 184 <conf:(0.98)> lift:(1.59) lev:(0.16) [68] conv:(14.52)

40. el-salvador-aid=n anti-satellite-test-ban=y 188 ==> aid-to-nicaraguan-contras=y 184 <conf:(0.98)> lift:(1.76) lev:(0.18) [79] conv:(16.68)

41. adoption-of-the-budget-resolution=y el-salvador-aid=n aid-to-nicaraguan-contras=y 186 ==> Class=democrat 182 <conf:(0.98)> lift:(1.59) lev:(0.16) [68] conv:(14.52)

42. el-salvador-aid=n mx-missile=y 179 ==> aid-to-nicaraguan-contras=y 175 <conf:(0.98)> lift:(1.76) lev:(0.17) [75] conv:(15.88)

43. physician-fee-freeze=n el-salvador-aid=n anti-satellite-test-ban=y 178 ==> aid-to-nicaraguan-contras=y Class=democrat 174 <conf:(0.98)> lift:(1.95) lev:(0.21) [93] conv:(19.46)

44. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat 203 ==> physician-fee-freeze=n 198 <conf:(0.98)> lift:(1.72) lev:(0.19) [82] conv:(14.41)

45. anti-satellite-test-ban=y Class=democrat 200 ==> physician-fee-freeze=n 195 <conf:(0.97)> lift:(1.72) lev:(0.19) [81] conv:(14.41)

Status

Run

Παρατηρώ πως από τον κανόνα 37 προκύπτει με confidence 0.98 ή 98% (δηλαδή ότι το 98%, 204/208), ότι ένας βουλευτής στο Κογκρέσο ο οποίος έχει καταψηφίσει το νομοσχέδιο για βοήθεια στο Ελ Σαλβαδόρ, πρόκειται να υπερψηφίσει το νομοσχέδιο της οικονομικής βοήθειας σε αντικυβερνητικές ομάδες της Νικαράγουα.

## Βήμα 7<sup>ο</sup>

Παρατηρούμε ότι στους 10 καλύτερους κανόνες η κλάση ισούται με Democrat. Εύλογα συμπεραίνουμε ότι ο τρόπος ψηφοφορίας των Democrat στο Κογκρέσο, είναι πιο εύκολα προβλέψιμος από εκείνο των Ρεπουμπλικάνων , ενώ είναι άξιο να αναφερθεί ότι υπάρχει και ισχυρή συσχέτιση confidence από 0.99 έως και 1,άρα συσχετίζεται σημαντικά.