

Fall 2017

CSE475\_Section: 01

Machine Learning

**Project Report**

**Topic**

Comparison between Decision Tree and SVM by using same dataset and  
 application of APriori algorithm for finding "Interesting Association Rules"

**Submitted To:**

Dr. Shamim Hasnat Ripon,

Associate Professor.

**Submitted By:**

|  |  |
| --- | --- |
| Name | ID |
| Kevin Stephen Biswas | 2014-2-60-091 |
| Md. Tanvir Aunjum | 2014-1-60-030 |
| Tasmia Tasmin | 2014-2-60-025 |

**Date of Report Submission:** 06/12/2017

**Comparison between Decision Tree algorithm C4.5 and SVM by using same dataset**

We have applied C4.5 or J48 algorithm on breast-cancer dataset for 20%, 50% and 100% as training data and remainder test data of the dataset to get better comparison. Like as breast-cancer dataset, we have also applied this algorithm on dataset car, credit-a, ecoli, glass, hepatitis as the same way.

We have also applied SVM algorithm on breast-cancer, car, credit-a, ecoli, glass, hepatitis dataset for 20%, 50% and 100% as training data and remainder test data of the dataset so that we can get better comparison result of SVM with C4.5 algorithm.

Firstly, we are giving the analysis result of these datasets which was done with Weka tool.

**C4.5 algorithm for Dataset breast-cancer.arff:**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: breast-cancer

Instances: 286

Attributes: 10

age

menopause

tumor-size

inv-nodes

node-caps

deg-malig

breast

breast-quad

irradiat

Class

=== Classifier model (full training set) ===

J48 pruned tree

------------------

node-caps = yes

| deg-malig = 1: recurrence-events (1.01/0.4)

| deg-malig = 2: no-recurrence-events (26.2/8.0)

| deg-malig = 3: recurrence-events (30.4/7.4)

node-caps = no: no-recurrence-events (228.39/53.4)

Number of Leaves : 4

Size of the tree : 6

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

**Result for 20% training data:**

Test mode: split 20.0% train, remainder test

=== Summary ===

Correctly Classified Instances 160 69.869 %

Incorrectly Classified Instances 69 30.131 %

Kappa statistic 0.0836

K&B Relative Info Score 2662.1458 %

K&B Information Score 24.1345 bits 0.1054 bits/instance

Class complexity | order 0 200.3618 bits 0.8749 bits/instance

Class complexity | scheme 5554.1551 bits 24.254 bits/instance

Complexity improvement- (Sf) -5353.7934 bits -23.379 bits/instance

Mean absolute error 0.3718

Root mean squared error 0.4588

Relative absolute error 87.2487 %

Root relative squared error 100.6294 %

Total Number of Instances 229

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.932 0.866 0.722 0.932 0.814 0.107 0.623 0.768 no-recurrence-events

0.134 0.068 0.450 0.134 0.207 0.107 0.623 0.379 recurrence-events

Weighted Avg. 0.699 0.632 0.643 0.699 0.636 0.107 0.623 0.654

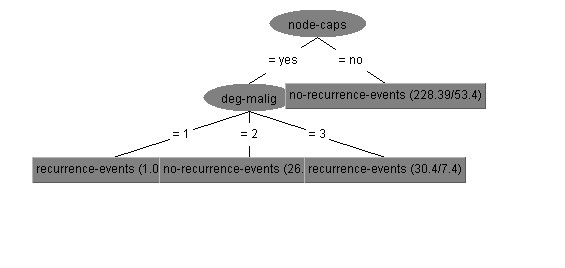
=== Confusion Matrix ===

a b <-- classified as

151 11 | a = no-recurrence-events

58 9 | b = recurrence-events

**Decision tree:**

****

**Result for 50% training data:**

Test mode: split 50.0% train, remainder test

=== Summary ===

Correctly Classified Instances 99 69.2308 %

Incorrectly Classified Instances 44 30.7692 %

Kappa statistic 0.1561

K&B Relative Info Score 2048.4564 %

K&B Information Score 17.2065 bits 0.1203 bits/instance

Class complexity | order 0 132.4327 bits 0.9261 bits/instance

Class complexity | scheme 131.476 bits 0.9194 bits/instance

Complexity improvement (Sf) 0.9567 bits 0.0067 bits/instance

Mean absolute error 0.3639

Root mean squared error 0.46

Relative absolute error 86.4811 %

Root relative squared error 97.1493 %

Total Number of Instances 143

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.938 0.809 0.703 0.938 0.804 0.198 0.656 0.754 no-recurrence-events

0.191 0.063 0.600 0.191 0.290 0.198 0.656 0.466 recurrence-events

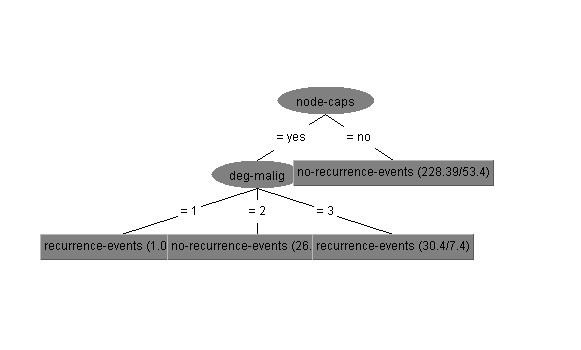
Weighted Avg. 0.692 0.563 0.669 0.692 0.635 0.198 0.656 0.660

=== Confusion Matrix ===

a b<-- classified as

90 6 | a = no-recurrence-events

38 9 | b = recurrence-events

**Decision tree:**

**Result for 100% training data:**

Test mode: evaluate on training data

=== Summary ===

Correctly Classified Instances 217 75.8741 %

Incorrectly Classified Instances 69 24.1259 %

Kappa statistic 0.2899

K&B Relative Info Score 4117.2496 %

K&B Information Score 36.2148 bits 0.1266 bits/instance

Class complexity | order 0 251.0655 bits 0.8779 bits/instance

Class complexity | scheme 227.0777 bits 0.794 bits/instance

Complexity improvement (Sf) 23.9879 bits 0.0839 bits/instance

Mean absolute error 0.3658

Root mean squared error 0.4269

Relative absolute error 87.4491 %

Root relative squared error 93.4017 %

Total Number of Instances 286

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.965 0.729 0.758 0.965 0.849 0.352 0.639 0.767 no-recurrence-events

0.271 0.035 0.767 0.271 0.400 0.352 0.639 0.461 recurrence-events

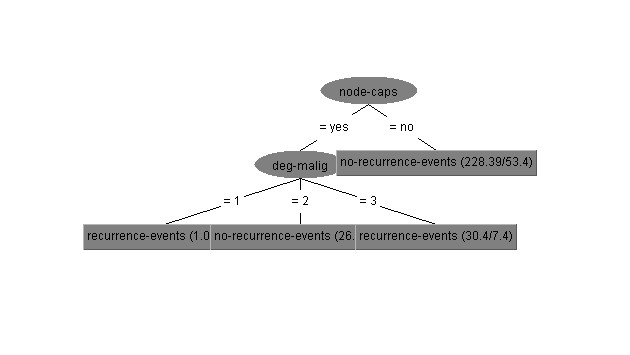
Weighted Avg. 0.759 0.523 0.760 0.759 0.716 0.352 0.639 0.676

=== Confusion Matrix ===

a b <-- classified as

194 7 | a = no-recurrence-events

62 23 | b = recurrence-events

**Decision tree:**

**SVM algorithm for Dataset glass.arff:**

=== Run information ===

Scheme: weka.classifiers.functions.SMO -C 1.0 -L 0.001 -P 1.0E-12 -N 0 -V -1 -W 1 -K "weka.classifiers.functions.supportVector.PolyKernel -E 1.0 -C 250007" -calibrator "weka.classifiers.functions.Logistic -R 1.0E-8 -M -1 -num-decimal-places 4"

Relation: Glass

Instances: 214

Attributes: 10

RI

Na

Mg

Al

Si

K

Ca

Ba

Fe

Type

=== Classifier model (full training set) ===

SMO

Kernel used:

Linear Kernel: K(x,y) = <x,y>

Classifier for classes: build wind float, build wind non-float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-1.1464 \* (normalized) RI

+ -0.4336 \* (normalized) Na

+ -2.5943 \* (normalized) Mg

+ 3.3669 \* (normalized) Al

+ -0.088 \* (normalized) Si

+ 0.5338 \* (normalized) K

+ -0.247 \* (normalized) Ca

+ -0.1905 \* (normalized) Ba

+ 1.0895 \* (normalized) Fe

+ 1.5558

Number of kernel evaluations: 1288 (58.572% cached)

Classifier for classes: build wind float, vehic wind float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.1061 \* (normalized) RI

+ -0.0297 \* (normalized) Na

+ 0.006 \* (normalized) Mg

+ -0.003 \* (normalized) Al

+ -0.0602 \* (normalized) Si

+ -0.0564 \* (normalized) K

+ 0.0706 \* (normalized) Ca

+ -0.0021 \* (normalized) Ba

+ -0.0039 \* (normalized) Fe

- 0.955

Number of kernel evaluations: 1239 (73.666% cached)

Classifier for classes: build wind float, vehic wind non-float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

- 1

Number of kernel evaluations: -1

Classifier for classes: build wind float, containers

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.4697 \* (normalized) RI

+ -0.5183 \* (normalized) Na

+ -2.4301 \* (normalized) Mg

+ 1.3939 \* (normalized) Al

+ -0.3768 \* (normalized) Si

+ 0.2087 \* (normalized) K

+ 0.7276 \* (normalized) Ca

+ 0.7054 \* (normalized) Ba

+ 0.218 \* (normalized) Fe

+ 0.3044

Number of kernel evaluations: 247 (74.109% cached)

Classifier for classes: build wind float, tableware

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.0407 \* (normalized) RI

+ 1.0768 \* (normalized) Na

+ -2.1971 \* (normalized) Mg

+ 0.317 \* (normalized) Al

+ 0.3163 \* (normalized) Si

+ -0.6108 \* (normalized) K

+ 0.402 \* (normalized) Ca

+ -0.0323 \* (normalized) Ba

+ -0.5961 \* (normalized) Fe

- 0.0594

Number of kernel evaluations: 363 (69.263% cached)

Classifier for classes: build wind float, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.0308 \* (normalized) RI

+ 0.5448 \* (normalized) Na

+ -1.9185 \* (normalized) Mg

+ 1.0234 \* (normalized) Al

+ -0.2241 \* (normalized) Si

+ 0.6455 \* (normalized) K

+ -0.4843 \* (normalized) Ca

+ 1.4847 \* (normalized) Ba

+ -0.2613 \* (normalized) Fe

+ 0.074

Number of kernel evaluations: 263 (67.611% cached)

Classifier for classes: build wind non-float, vehic wind float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.2453 \* (normalized) RI

+ -0.0029 \* (normalized) Na

+ 0.0744 \* (normalized) Mg

+ -0.1938 \* (normalized) Al

+ -0.1001 \* (normalized) Si

+ -0.1193 \* (normalized) K

+ 0.132 \* (normalized) Ca

+ 0.0475 \* (normalized) Ba

+ -0.0214 \* (normalized) Fe

- 0.9042

Number of kernel evaluations: 1228 (81.694% cached)

Classifier for classes: build wind non-float, vehic wind non-float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

- 1

Number of kernel evaluations: -1

Classifier for classes: build wind non-float, containers

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.6372 \* (normalized) RI

+ -0.2863 \* (normalized) Na

+ -1.1823 \* (normalized) Mg

+ 1.3352 \* (normalized) Al

+ 0.3711 \* (normalized) Si

+ 0.7632 \* (normalized) K

+ -0.2636 \* (normalized) Ca

+ -0.1507 \* (normalized) Ba

+ -0.2798 \* (normalized) Fe

- 0.7852

Number of kernel evaluations: 765 (73.354% cached)

Classifier for classes: build wind non-float, tableware

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.5382 \* (normalized) RI

+ 1.1992 \* (normalized) Na

+ -0.941 \* (normalized) Mg

+ 0.0371 \* (normalized) Al

+ 0.7911 \* (normalized) Si

+ -0.5435 \* (normalized) K

+ -0.3804 \* (normalized) Ca

+ -0.615 \* (normalized) Fe

- 0.947

Number of kernel evaluations: 551 (72.214% cached)

Classifier for classes: build wind non-float, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.7228 \* (normalized) RI

+ 1.1976 \* (normalized) Na

+ -1.6512 \* (normalized) Mg

+ 1.3999 \* (normalized) Al

+ 0.4716 \* (normalized) Si

+ 0.3555 \* (normalized) K

+ -1.1593 \* (normalized) Ca

+ 1.4598 \* (normalized) Ba

+ -0.9559 \* (normalized) Fe

- 0.6299

Number of kernel evaluations: 829 (69.216% cached)

Classifier for classes: vehic wind float, vehic wind non-float

BinarySMO

Machine linear: showing attribute weights, not support vectors.

- 1

Number of kernel evaluations: -1

Classifier for classes: vehic wind float, containers

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.4931 \* (normalized) RI

+ -0.5588 \* (normalized) Na

+ -2.23 \* (normalized) Mg

+ 0.9844 \* (normalized) Al

+ 0.0565 \* (normalized) Si

+ 0.2368 \* (normalized) K

+ 0.6183 \* (normalized) Ca

+ 0.7746 \* (normalized) Ba

+ 0.3525 \* (normalized) Fe

+ 0.2375

Number of kernel evaluations: 146 (80.429% cached)

Classifier for classes: vehic wind float, tableware

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.1957 \* (normalized) RI

+ 0.8582 \* (normalized) Na

+ -2.1432 \* (normalized) Mg

+ 0.2875 \* (normalized) Al

+ 0.5125 \* (normalized) Si

+ -0.543 \* (normalized) K

+ 0.3753 \* (normalized) Ca

+ -0.1765 \* (normalized) Fe

- 0.1305

Number of kernel evaluations: 91 (72.508% cached)

Classifier for classes: vehic wind float, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.4449 \* (normalized) RI

+ 0.6332 \* (normalized) Na

+ -2.0088 \* (normalized) Mg

+ 0.6381 \* (normalized) Al

+ -0.1315 \* (normalized) Si

+ 0.5692 \* (normalized) K

+ -0.3995 \* (normalized) Ca

+ 1.583 \* (normalized) Ba

+ -0.355 \* (normalized) Fe

+ 0.1505

Number of kernel evaluations: 152 (72.662% cached)

Classifier for classes: vehic wind non-float, containers

BinarySMO

Machine linear: showing attribute weights, not support vectors.

+ 1

Number of kernel evaluations: -1

Classifier for classes: vehic wind non-float, tableware

BinarySMO

Machine linear: showing attribute weights, not support vectors.

+ 1

Number of kernel evaluations: -1

Classifier for classes: vehic wind non-float, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

+ 1

Number of kernel evaluations: -1

Classifier for classes: containers, tableware

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.4152 \* (normalized) RI

+ 1.6242 \* (normalized) Na

+ 1.158 \* (normalized) Mg

+ -0.9549 \* (normalized) Al

+ 0.3889 \* (normalized) Si

+ -0.8119 \* (normalized) K

+ -0.6678 \* (normalized) Ca

+ -0.6767 \* (normalized) Ba

+ -0.3729 \* (normalized) Fe

- 0.7515

Number of kernel evaluations: 80 (77.901% cached)

Classifier for classes: containers, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

-0.6775 \* (normalized) RI

+ 2.0251 \* (normalized) Na

+ 0.1088 \* (normalized) Mg

+ 0.5969 \* (normalized) Al

+ 0.7321 \* (normalized) Si

+ -1.3744 \* (normalized) K

+ -1.2601 \* (normalized) Ca

+ 1.1996 \* (normalized) Ba

+ -0.805 \* (normalized) Fe

- 0.6823

Number of kernel evaluations: 305 (77.915% cached)

Classifier for classes: tableware, headlamps

BinarySMO

Machine linear: showing attribute weights, not support vectors.

0.1851 \* (normalized) RI

+ -0.4361 \* (normalized) Na

+ -0.602 \* (normalized) Mg

+ 1.5256 \* (normalized) Al

+ -0.4622 \* (normalized) Si

+ 0.7426 \* (normalized) K

+ -0.5264 \* (normalized) Ca

+ 1.4703 \* (normalized) Ba

+ 0.0381 \* (normalized) Fe

+ 0.3816

Number of kernel evaluations: 134 (71.61% cached)

Time taken to build model: 0.03 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

**Result for 20% training data:**

Test mode: split 20.0% train, remainder test

=== Summary ===

Correctly Classified Instances 88 51.462 %

Incorrectly Classified Instances 83 48.538 %

Kappa statistic 0.3046

K&B Relative Info Score 1851.1653 %

K&B Information Score 46.6825 bits 0.273 bits/instance

Class complexity | order 0 386.0298 bits 2.2575 bits/instance

Class complexity | scheme 344.4801 bits 2.0145 bits/instance

Complexity improvement (Sf) 41.5496 bits 0.243 bits/instance

Mean absolute error 0.2135

Root mean squared error 0.3164

Relative absolute error 96.3292 %

Root relative squared error 96.6967 %

Total Number of Instances 171

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

1.000 0.635 0.434 1.000 0.605 0.398 0.691 0.441 build wind float

0.154 0.066 0.588 0.154 0.244 0.142 0.595 0.433 build wind non-float

0.000 0.000 0.000 0.000 0.000 0.000 0.636 0.102 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.222 0.012 0.500 0.222 0.308 0.310 0.926 0.308 containers

0.200 0.000 1.000 0.200 0.333 0.442 0.758 0.303 tableware

0.826 0.007 0.950 0.826 0.884 0.870 0.943 0.840 headlamps

Weighted Avg. 0.515 0.235 0.549 0.515 0.436 0.331 0.699 0.455

=== Confusion Matrix ===

a b c d e f g <-- classified as

56 0 0 0 0 0 0 | a = build wind float

54 10 0 0 1 0 0 | b = build wind non-float

13 0 0 0 0 0 0 | c = vehic wind float

0 0 0 0 0 0 0 | d = vehic wind non-float

2 5 0 0 2 0 0 | e = containers

2 1 0 0 0 1 1 | f = tableware

2 1 0 0 1 0 19 | g = headlamps

**Result for 50% training data:**

Test mode: split 50.0% train, remainder test

=== Summary ===

Correctly Classified Instances 53 49.5327 %

Incorrectly Classified Instances 54 50.4673 %

Kappa statistic 0.307

K&B Relative Info Score 1223.9059 %

K&B Information Score 28.4908 bits 0.2663 bits/instance

Class complexity | order 0 232.1306 bits 2.1694 bits/instance

Class complexity | scheme 216.2156 bits 2.0207 bits/instance

Complexity improvement (Sf) 15.915 bits 0.1487 bits/instance

Mean absolute error 0.214

Root mean squared error 0.317

Relative absolute error 99.8829 %

Root relative squared error 97.6589 %

Total Number of Instances 107

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.871 0.513 0.409 0.871 0.557 0.334 0.715 0.417 build wind float

0.222 0.129 0.556 0.222 0.317 0.123 0.588 0.474 build wind non-float

0.000 0.000 0.000 0.000 0.000 0.000 0.650 0.091 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.600 0.059 0.333 0.600 0.429 0.412 0.927 0.274 containers

0.000 0.000 0.000 0.000 0.000 0.000 0.671 0.118 tableware

0.867 0.011 0.929 0.867 0.897 0.881 0.959 0.852 headlamps

Weighted Avg. 0.495 0.207 0.498 0.495 0.441 0.291 0.700 0.463

=== Confusion Matrix ===

a b c d e f g <-- classified as

27 4 0 0 0 0 0 | a = build wind float

30 10 0 0 5 0 0 | b = build wind non-float

5 2 0 0 0 0 0 | c = vehic wind float

0 0 0 0 0 0 0 | d = vehic wind non-float

1 1 0 0 3 0 0 | e = containers

1 1 0 0 1 0 1 | f = tableware

2 0 0 0 0 0 13 | g = headlamps

**Result for 100% training data:**

Test mode: evaluate on training data

=== Summary ===

Correctly Classified Instances 130 60.7477 %

Incorrectly Classified Instances 84 39.2523 %

Kappa statistic 0.4259

K&B Relative Info Score 3153.0098 %

K&B Information Score 70.3598 bits 0.3288 bits/instance

Class complexity | order 0 467.2716 bits 2.1835 bits/instance

Class complexity | scheme 424.1279 bits 1.9819 bits/instance

Complexity improvement (Sf) 43.1436 bits 0.2016 bits/instance

Mean absolute error 0.212

Root mean squared error 0.3138

Relative absolute error 100.1781 %

Root relative squared error 96.7111 %

Total Number of Instances 214

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.514 0.118 0.679 0.514 0.585 0.431 0.796 0.567 build wind float

0.868 0.471 0.504 0.868 0.638 0.390 0.722 0.502 build wind non-float

0.000 0.000 0.000 0.000 0.000 0.000 0.645 0.108 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.154 0.000 1.000 0.154 0.267 0.382 0.957 0.491 containers

0.111 0.000 1.000 0.111 0.200 0.327 0.624 0.181 tableware

0.862 0.011 0.926 0.862 0.893 0.877 0.962 0.879 headlamps

Weighted Avg. 0.607 0.207 0.629 0.607 0.564 0.435 0.783 0.529

=== Confusion Matrix ===

a b c d e f g <-- classified as

36 34 0 0 0 0 0 | a = build wind float

10 66 0 0 0 0 0 | b = build wind non-float

6 11 0 0 0 0 0 | c = vehic wind float

0 0 0 0 0 0 0 | d = vehic wind non-float

0 10 0 0 2 0 1 | e = containers

0 7 0 0 0 1 1 | f = tableware

1 3 0 0 0 0 25 | g = headlamps

**C4.5 algorithm for Dataset glass.arff:**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: Glass

Instances: 214

Attributes: 10

RI

Na

Mg

Al

Si

K

Ca

Ba

Fe

Type

=== Classifier model (full training set) ===

J48 pruned tree

------------------

Ba <= 0.27

| Mg <= 2.41

| | K <= 0.03

| | | Na <= 13.75: build wind non-float (3.0)

| | | Na > 13.75: tableware (9.0)

| | K > 0.03

| | | Na <= 13.49

| | | | RI <= 1.5241: containers (13.0/1.0)

| | | | RI > 1.5241: build wind non-float (3.0)

| | | Na > 13.49: build wind non-float (7.0/1.0)

| Mg > 2.41

| | Al <= 1.41

| | | RI <= 1.51707

| | | | RI <= 1.51596: build wind float (3.0)

| | | | RI > 1.51596

| | | | | Fe <= 0.12

| | | | | | Mg <= 3.54: vehic wind float (5.0)

| | | | | | Mg > 3.54

| | | | | | | RI <= 1.51667: build wind non-float (2.0)

| | | | | | | RI > 1.51667: vehic wind float (2.0)

| | | | | Fe > 0.12: build wind non-float (2.0)

| | | RI > 1.51707

| | | | K <= 0.23

| | | | | Mg <= 3.34: build wind non-float (2.0)

| | | | | Mg > 3.34

| | | | | | Si <= 72.64

| | | | | | | Na <= 14.01: build wind float (14.0)

| | | | | | | Na > 14.01

| | | | | | | | RI <= 1.52211

| | | | | | | | | Na <= 14.32: vehic wind float (3.0)

| | | | | | | | | Na > 14.32: build wind float (2.0)

| | | | | | | | RI > 1.52211: build wind float (3.0)

| | | | | | Si > 72.64: vehic wind float (3.0)

| | | | K > 0.23

| | | | | Mg <= 3.75

| | | | | | Fe <= 0.14

| | | | | | | RI <= 1.52043: build wind float (36.0)

| | | | | | | RI > 1.52043: build wind non-float (2.0/1.0)

| | | | | | Fe > 0.14

| | | | | | | Al <= 1.17: build wind non-float (5.0)

| | | | | | | Al > 1.17: build wind float (6.0/1.0)

| | | | | Mg > 3.75: build wind non-float (10.0)

| | Al > 1.41

| | | Si <= 72.49

| | | | Ca <= 8.28: build wind non-float (6.0)

| | | | Ca > 8.28: vehic wind float (5.0/1.0)

| | | Si > 72.49

| | | | RI <= 1.51732

| | | | | Fe <= 0.22: build wind non-float (30.0/1.0)

| | | | | Fe > 0.22

| | | | | | RI <= 1.51629: build wind float (2.0)

| | | | | | RI > 1.51629: build wind non-float (2.0)

| | | | RI > 1.51732

| | | | | RI <= 1.51789: build wind float (3.0)

| | | | | RI > 1.51789: build wind non-float (2.0)

Ba > 0.27

| Si <= 70.16: build wind non-float (2.0/1.0)

| Si > 70.16: headlamps (27.0/1.0)

Number of Leaves : 30

Size of the tree : 59

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

**Result for 20% training data:**

Test mode: split 20.0% train, remainder test

=== Summary ===

Correctly Classified Instances 98 57.3099 %

Incorrectly Classified Instances 73 42.6901 %

Kappa statistic 0.4298

K&B Relative Info Score 7436.9111 %

K&B Information Score 187.5433 bits 1.0967 bits/instance

Class complexity | order 0 386.0298 bits 2.2575 bits/instance

Class complexity | scheme 36656.1749 bits 214.3636 bits/instance

Complexity improvement (Sf) -36270.1451 bits -212.1061 bits/instance

Mean absolute error 0.133

Root mean squared error 0.3159

Relative absolute error 60.0011 %

Root relative squared error 96.5467 %

Total Number of Instances 171

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.714 0.278 0.556 0.714 0.625 0.414 0.765 0.572 build wind float

0.446 0.104 0.725 0.446 0.552 0.393 0.681 0.573 build wind non-float

0.231 0.038 0.333 0.231 0.273 0.229 0.766 0.207 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.778 0.105 0.292 0.778 0.424 0.433 0.836 0.239 containers

1.000 0.042 0.417 1.000 0.588 0.632 0.979 0.417 tableware

0.609 0.000 1.000 0.609 0.757 0.757 0.816 0.684 headlamps

Weighted Avg. 0.573 0.140 0.645 0.573 0.577 0.445 0.750 0.538

=== Confusion Matrix ===

a b c d e f g <-- classified as

40 7 2 0 4 3 0 | a = build wind float

21 29 4 0 8 3 0 | b = build wind non-float

8 1 3 0 0 1 0 | c = vehic wind float

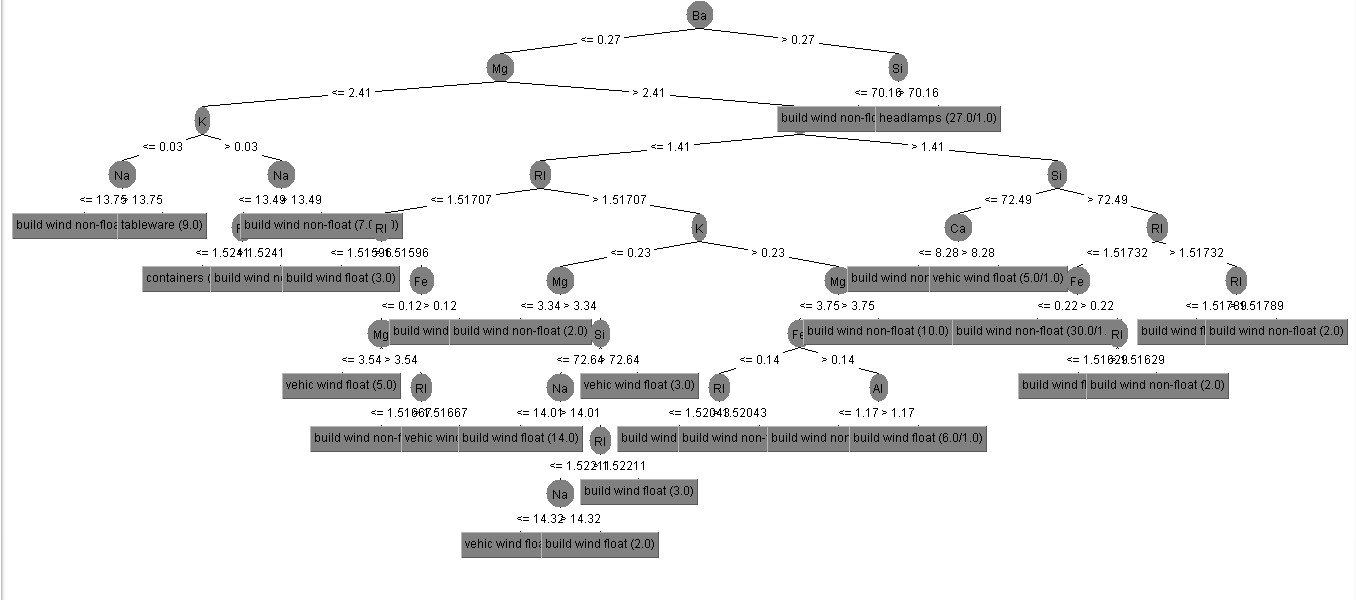
0 0 0 0 0 0 0 | d = vehic wind non-float

0 2 0 0 7 0 0 | e = containers

0 0 0 0 0 5 0 | f = tableware

3 1 0 0 5 0 14 | g = headlamps

**Decision tree:**

****

**Result for 50% training data:**

Test mode: split 50.0% train, remainder test

=== Summary ===

Correctly Classified Instances 65 60.7477 %

Incorrectly Classified Instances 42 39.2523 %

Kappa statistic 0.4755

K&B Relative Info Score 5368.778 %

K&B Information Score 124.9777 bits 1.168 bits/instance

Class complexity | order 0 232.1306 bits 2.1694 bits/instance

Class complexity | scheme 32262.0616 bits 301.5146 bits/instance

Complexity improvement (Sf) -32029.931 bits -299.3451 bits/instance

Mean absolute error 0.1211

Root mean squared error 0.3198

Relative absolute error 56.4994 %

Root relative squared error 98.5263 %

Total Number of Instances 107

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.516 0.132 0.615 0.516 0.561 0.407 0.698 0.486 build wind float

0.578 0.161 0.722 0.578 0.642 0.435 0.701 0.583 build wind non-float

0.429 0.100 0.231 0.429 0.300 0.249 0.709 0.159 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.600 0.020 0.600 0.600 0.600 0.580 0.964 0.529 containers

0.750 0.029 0.500 0.750 0.600 0.594 0.860 0.384 tableware

0.933 0.076 0.667 0.933 0.778 0.749 0.934 0.663 headlamps

Weighted Avg. 0.607 0.125 0.637 0.607 0.612 0.471 0.751 0.528

=== Confusion Matrix ===

a b c d e f g <-- classified as

16 8 6 0 0 0 1 | a = build wind float

7 26 4 0 2 3 3 | b = build wind non-float

2 1 3 0 0 0 1 | c = vehic wind float

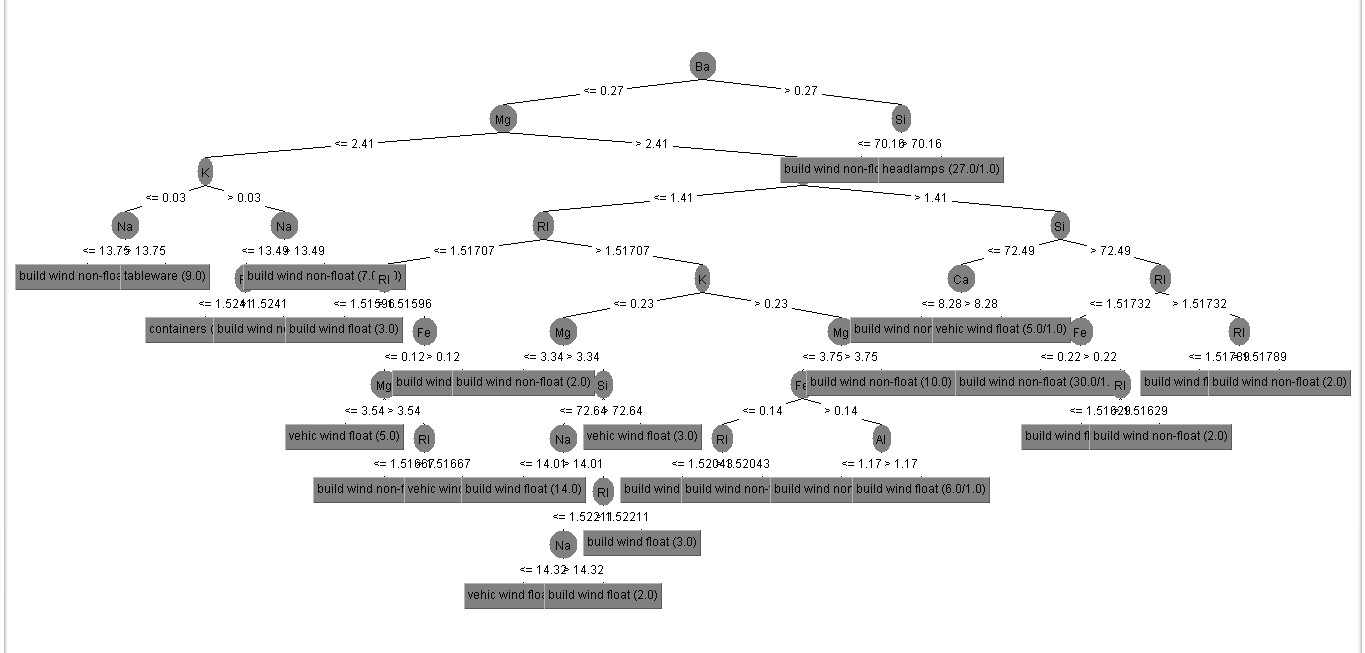
0 0 0 0 0 0 0 | d = vehic wind non-float

0 0 0 0 3 0 2 | e = containers

1 0 0 0 0 3 0 | f = tableware

0 1 0 0 0 0 14 | g = headlamps

**Decision tree:**

****

**Result for 100% training data:**

Test mode: evaluate on training data

=== Summary ===

Correctly Classified Instances 206 96.2617 %

Incorrectly Classified Instances 8 3.7383 %

Kappa statistic 0.9492

K&B Relative Info Score 19777.9255 %

K&B Information Score 441.3469 bits 2.0624 bits/instance

Class complexity | order 0 467.2716 bits 2.1835 bits/instance

Class complexity | scheme 33.2334 bits 0.1553 bits/instance

Complexity improvement (Sf) 434.0381 bits 2.0282 bits/instance

Mean absolute error 0.0169

Root mean squared error 0.092

Relative absolute error 8.005 %

Root relative squared error 28.364 %

Total Number of Instances 214

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class

0.971 0.007 0.986 0.971 0.978 0.968 0.994 0.989 build wind float

0.974 0.029 0.949 0.974 0.961 0.939 0.996 0.988 build wind non-float

1.000 0.005 0.944 1.000 0.971 0.969 0.999 0.987 vehic wind float

0.000 0.000 0.000 0.000 0.000 0.000 ? ? vehic wind non-float

0.923 0.005 0.923 0.923 0.923 0.918 0.997 0.919 containers

1.000 0.000 1.000 1.000 1.000 1.000 1.000 1.000 tableware

0.897 0.005 0.963 0.897 0.929 0.919 0.994 0.943 headlamps

Weighted Avg. 0.963 0.014 0.963 0.963 0.962 0.950 0.995 0.978

=== Confusion Matrix ===

a b c d e f g <-- classified as

68 1 0 0 0 0 1 | a = build wind float

1 74 1 0 0 0 0 | b = build wind non-float

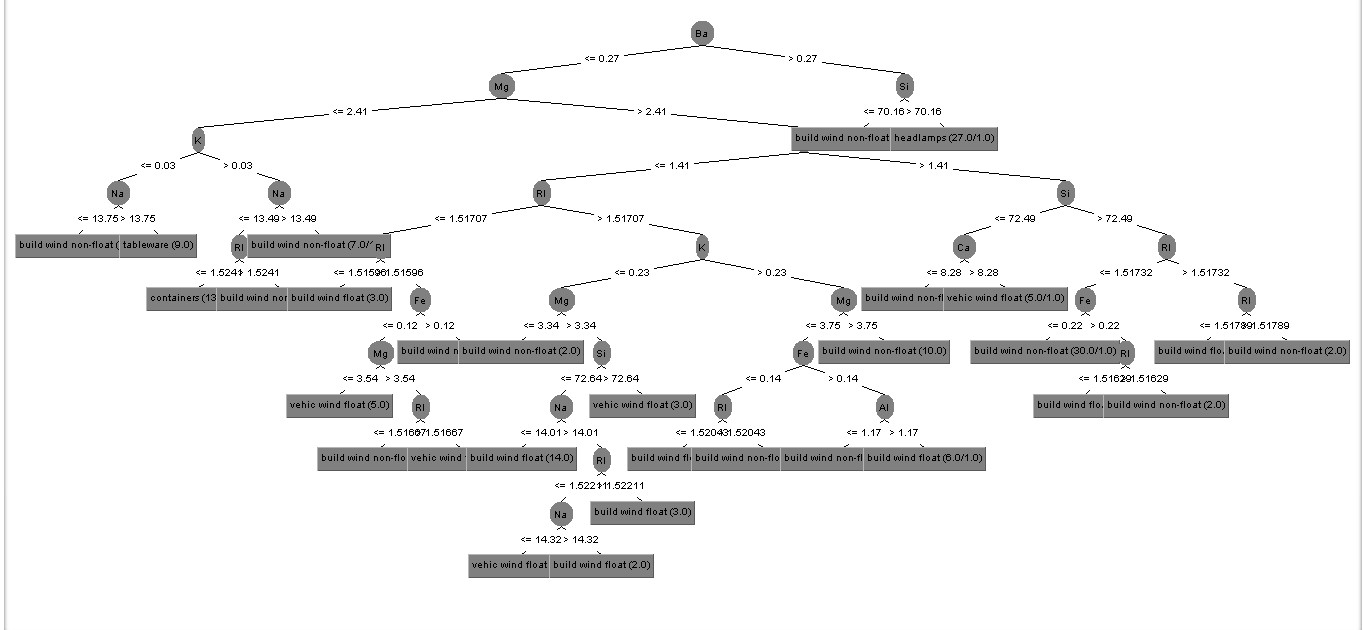
0 0 17 0 0 0 0 | c = vehic wind float

0 0 0 0 0 0 0 | d = vehic wind non-float

0 1 0 0 12 0 0 | e = containers

0 0 0 0 0 9 0 | f = tableware

0 2 0 0 1 0 26 | g = headlamps

**Decision tree:**

**Table: Mean Absolute Error**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Set | Sample Size | Algorithm | |
| C4.5 | SVM |
| breast-cancer.arff | 20% | 0.3718 | 0.3231 |
| 50% | 0.3639 | 0.2727 |
| 100% | 0.3658 | 0.2378 |
| car.arff | 20% | 0.0878 | 0.2607 |
| 50% | 0.0757 | 0.258 |
| 100% | 0.0248 | 0.2545 |
| credit-a.arff | 20% | 0.2038 | 0.1612 |
| 50% | 0.1963 | 0.2058 |
| 100% | 0.1564 | 0.1406 |
| ecoli.arff | 20% | 0.0779 | 0.192 |
| 50% | 0.0543 | 0.1904 |
| 100% | 0.0269 | 0.1892 |
| glass.arff | 20% | 0.133 | 0.2135 |
| 50% | 0.1211 | 0.214 |
| 100% | 0.0169 | 0.212 |
| hepatitis.arff | 20% | 0.3006 | 0.2419 |
| 50% | 0.2584 | 0.1299 |
| 100% | 0.1272 | 0.1161 |

**Table: Root Mean Squared Error**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Set | Sample Size | Algorithm | |
| C4.5 | SVM |
| breast-cancer.arff | 20% | 0.4588 | 0.5685 |
| 50% | 0.46 | 0.5222 |
| 100% | 0.4269 | 0.4876 |
| glass.arff | 20% | 0.3159 | 0.3164 |
| 50% | 0.3198 | 0.317 |
| 100% | 0.092 | 0.3138 |

**Table: Relative Absolute Error**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Set | Sample Size | Algorithm | |
| C4.5 | SVM |
| breast-cancer.arff | 20% | 87.2487 % | 75.8249 % |
| 50% | 86.4811 % | 64.8063 % |
| 100% | 87.4491 % | 56.8377 % |
| glass.arff | 20% | 60.0011 % | 96.3292 % |
| 50% | 56.4994 % | 99.8829 % |
| 100% | 8.005 % | 100.1781 % |

**Table: Root relative squared error**

|  |  |  |  |
| --- | --- | --- | --- |
| Data Set | Sample Size | Algorithm | |
| C4.5 | SVM |
| breast-cancer.arff | 20% | 100.6294 % | 124.6896 % |
| 50% | 97.1493 % | 110.2899 % |
| 100% | 93.4017 % | 106.691 % |
| glass.arff | 20% | 96.5467 % | 96.6967 % |
| 50% | 98.5263 % | 97.6589 % |
| 100% | 28.364 % | 96.7111 % |

**Some Graphical Representation for Glass dataset is Given to Visualize the Relation and Differences:**

**Pie Charts to Compare the Correctness of Instances:**

**Algorithm C4.5:**

**Correctness of Classified Instances**

**Figure: Result for 20% training data**

**Figure: Result for 50% training data**

**Figure: Result for 100% training data:**

**Bar Diagram for Tables:**

**Result for 20% training data:**

**For 100% training data:**

**APriori algorithm for dataset audiology.arff:**

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: audiology

Instances: 226

Attributes: 70

age\_gt\_60

air

airBoneGap

ar\_c

ar\_u

bone

boneAbnormal

bser

history\_buzzing

history\_dizziness

history\_fluctuating

history\_fullness

history\_heredity

history\_nausea

history\_noise

history\_recruitment

history\_ringing

history\_roaring

history\_vomiting

late\_wave\_poor

m\_at\_2k

m\_cond\_lt\_1k

m\_gt\_1k

m\_m\_gt\_2k

m\_m\_sn

m\_m\_sn\_gt\_1k

m\_m\_sn\_gt\_2k

m\_m\_sn\_gt\_500

m\_p\_sn\_gt\_2k

m\_s\_gt\_500

m\_s\_sn

m\_s\_sn\_gt\_1k

m\_s\_sn\_gt\_2k

m\_s\_sn\_gt\_3k

m\_s\_sn\_gt\_4k

m\_sn\_2\_3k

m\_sn\_gt\_1k

m\_sn\_gt\_2k

m\_sn\_gt\_3k

m\_sn\_gt\_4k

m\_sn\_gt\_500

m\_sn\_gt\_6k

m\_sn\_lt\_1k

m\_sn\_lt\_2k

m\_sn\_lt\_3k

middle\_wave\_poor

mod\_gt\_4k

mod\_mixed

mod\_s\_mixed

mod\_s\_sn\_gt\_500

mod\_sn

mod\_sn\_gt\_1k

mod\_sn\_gt\_2k

mod\_sn\_gt\_3k

mod\_sn\_gt\_4k

mod\_sn\_gt\_500

notch\_4k

notch\_at\_4k

o\_ar\_c

o\_ar\_u

s\_sn\_gt\_1k

s\_sn\_gt\_2k

s\_sn\_gt\_4k

speech

static\_normal

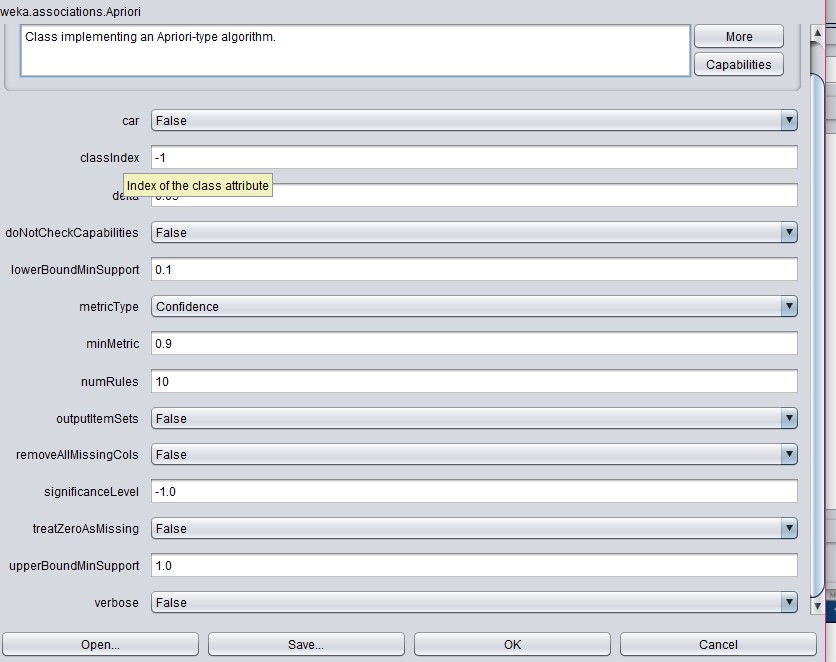
tymp

viith\_nerve\_signs

wave\_V\_delayed

waveform\_ItoV\_prolonged

class



**Figure: APriori Configuration for dataset audiology.arff**

**APriori algorithm for dataset supermarket.arff:**

=== Run information ===

Scheme: weka.associations.Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1

Relation: supermarket

Instances: 4627

Attributes: 217

[list of attributes omitted]

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

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 itemsetsL(1): 44

Size of set of large itemsetsL(2): 380

Size of set of large itemsetsL(3): 910

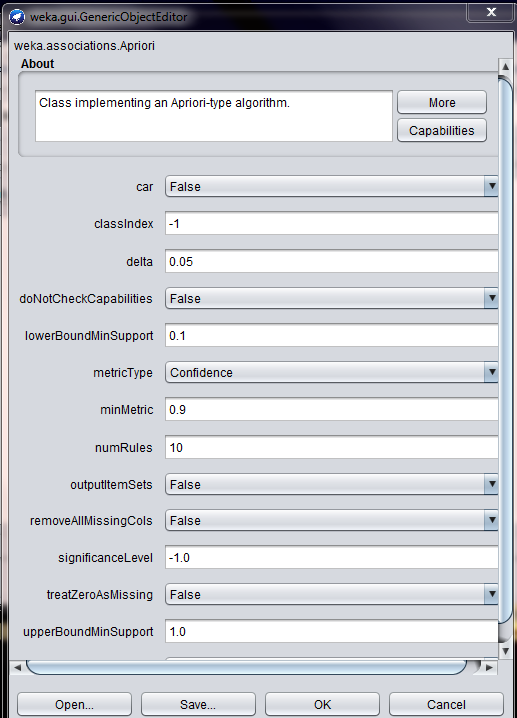
Size of set of large itemsetsL(4): 633

Size of set of large itemsetsL(5): 105

Size of set of large itemsetsL(6): 1

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)

****

**Figure: APriori Configuration for dataset supermarket.arff**