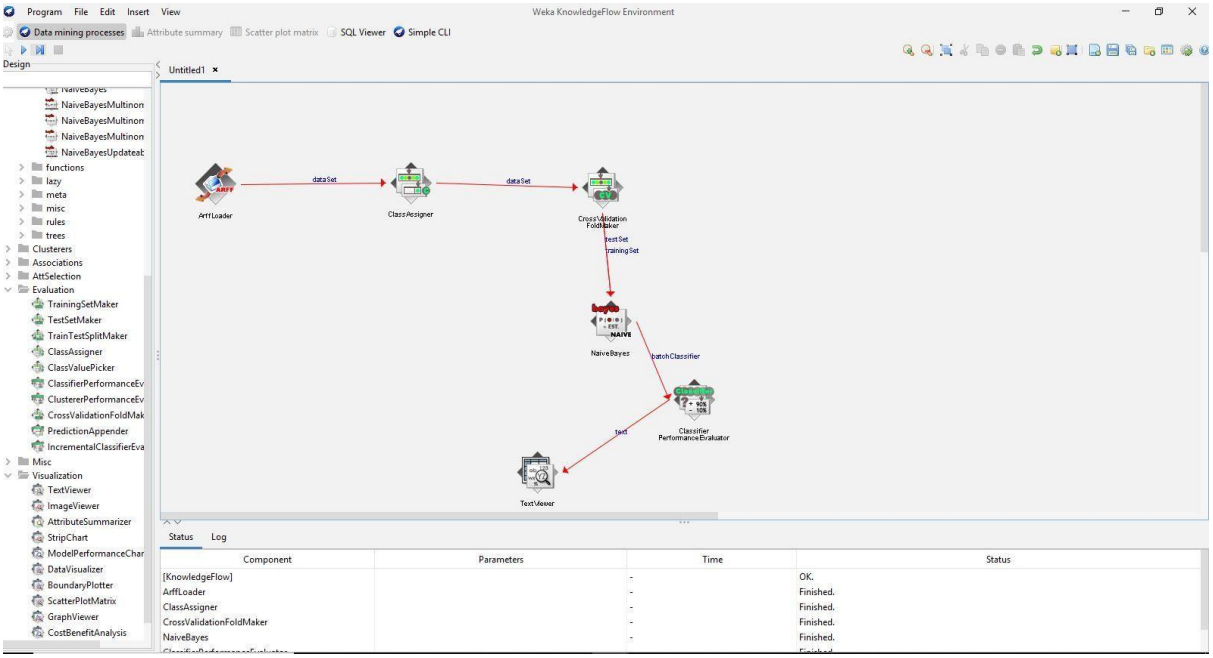
Name: Siddhi Sunil Khade
Division: TE3
Batch: B
Roll.No: 22

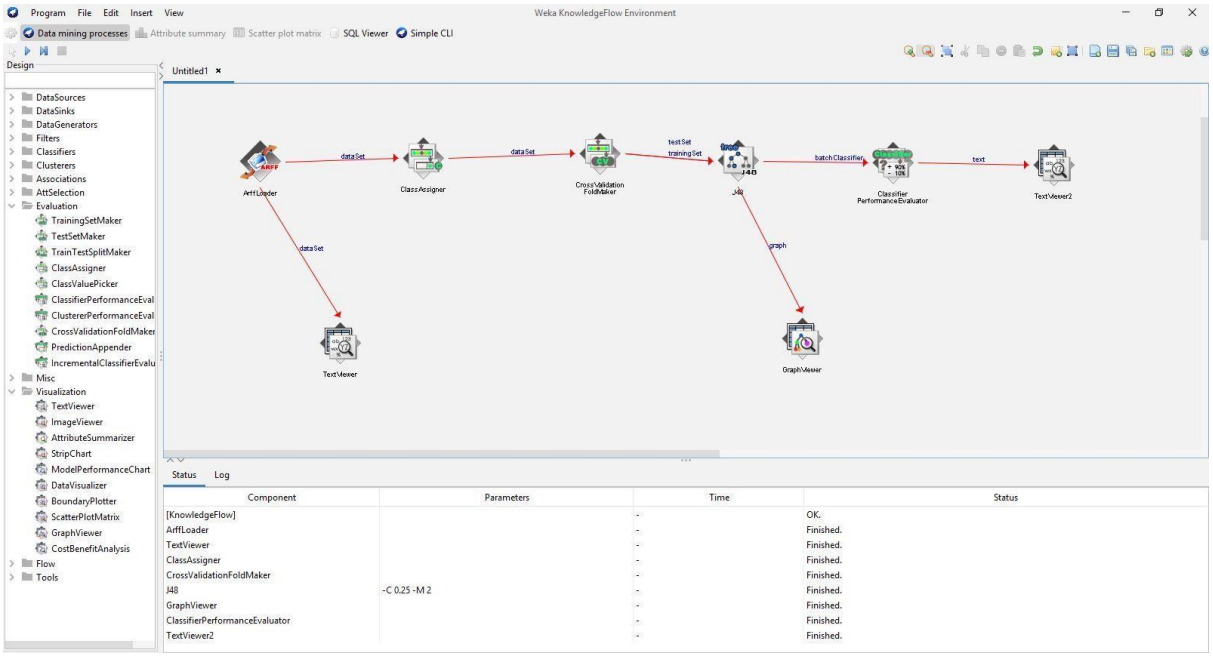
Experiment 6

Aim:

Perform data Pre-processing task and demonstrate Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA).

Classification:





Text Viewer

Result list

15:36:20.560 - NaiveBayes

Text

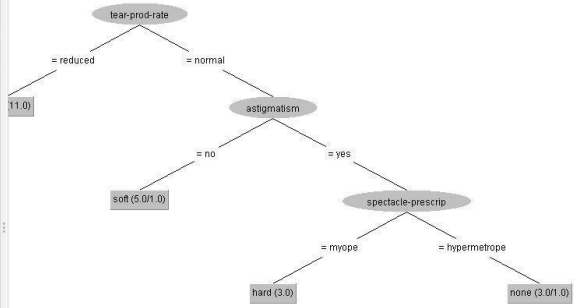
```
=== Evaluation result ===

Scheme: NaiveBayes
Relation: contact-lenses

=== Summary ===

Correctly Classified Instances          17               70.8333 %
Incorrectly Classified Instances         7               29.1667 %
Kappa statistic                          0.4381
Mean absolute error                      0.2545
Root mean squared error                  0.3326
Relative absolute error                 67.3578 %
Root relative squared error             76.1544 %
Total Number of Instances               24

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.800    0.053    0.800      0.800    0.800      0.747    0.947     0.710     soft
               0.250    0.100    0.333      0.250    0.286      0.169    0.925     0.692     hard
               0.800    0.444    0.750      0.800    0.774      0.365    0.830     0.930     none
Weighted Avg.  0.708    0.305    0.691      0.708    0.698      0.412    0.870     0.844

=== Confusion Matrix ===

  a  b  c   <-- classified as
  4  0  1 |  a = soft
  0  1  3 |  b = hard
  1  2 12 |  c = none
```

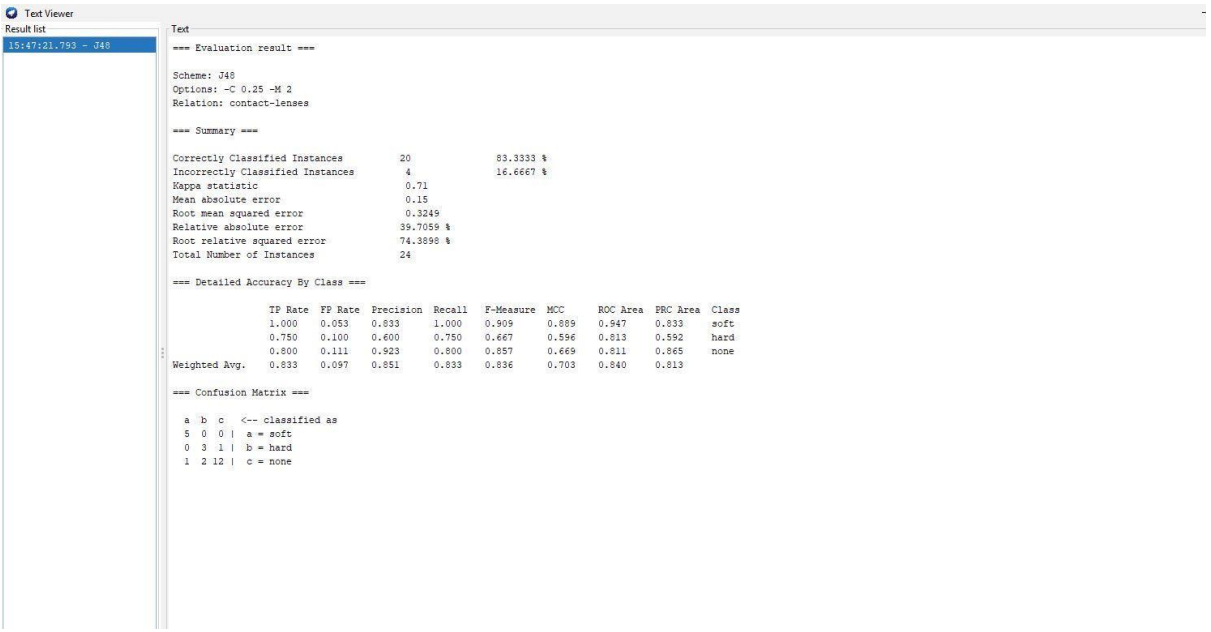
Graph Viewer

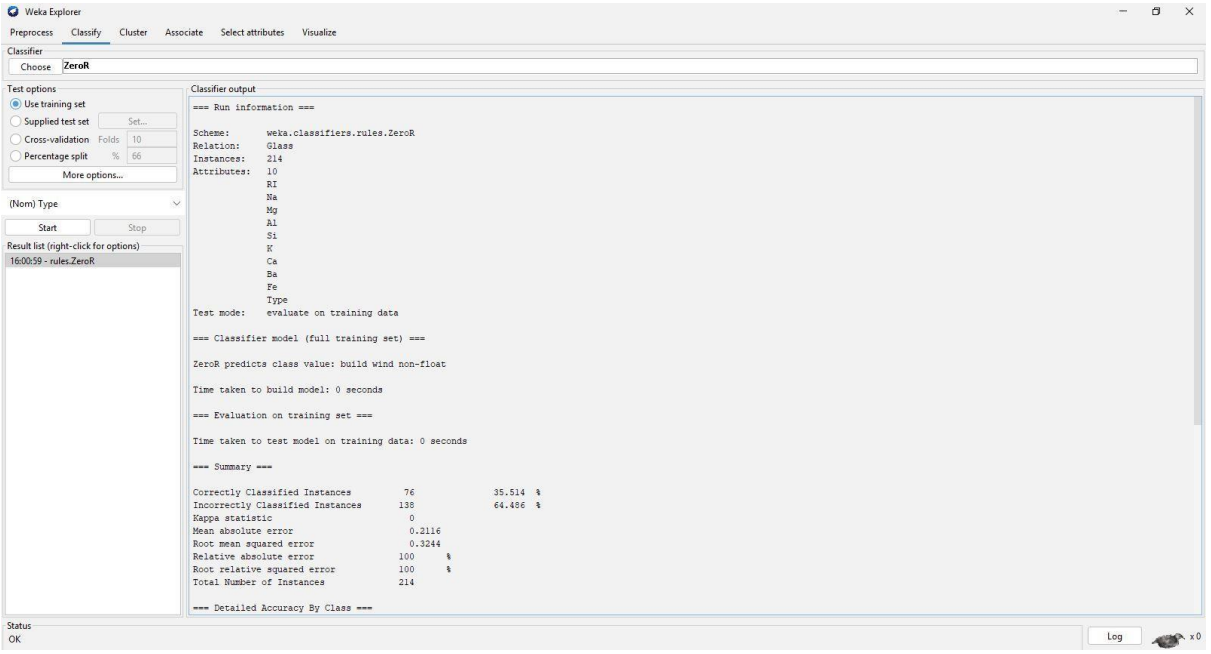
Result list

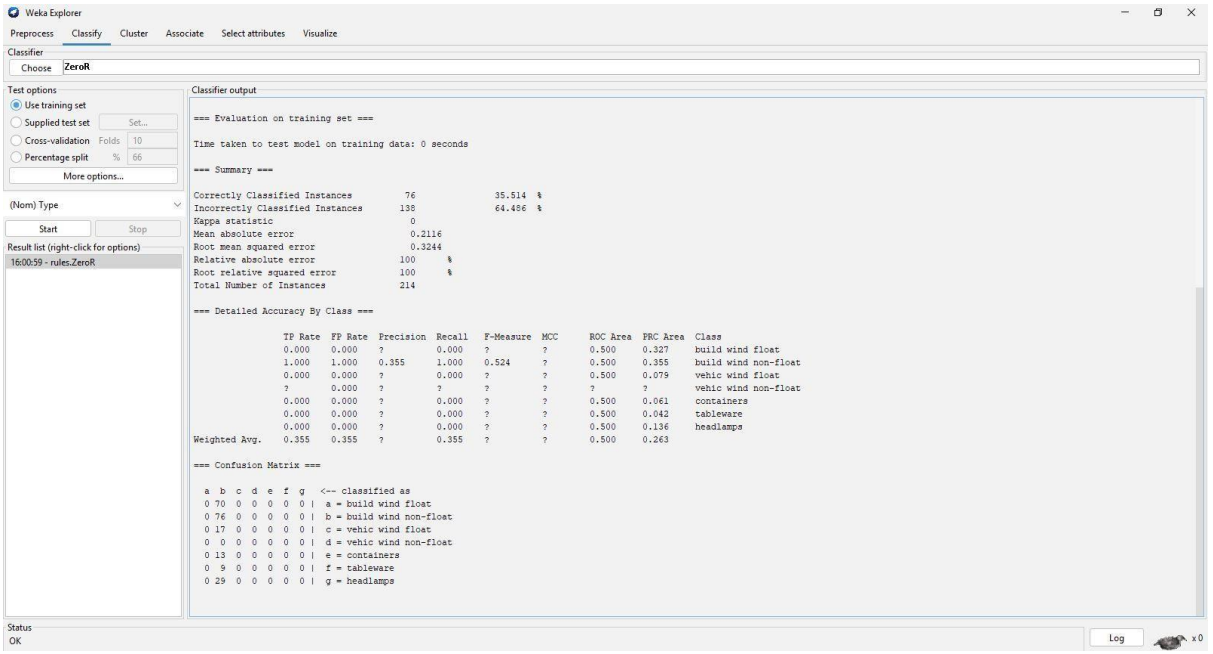
- Set 2 (contact-lenses) J48
- Set 5 (contact-lenses) J48
- Set 1 (contact-lenses) J48
- Set 4 (contact-lenses) J48
- Set 3 (contact-lenses) J48
- Set 6 (contact-lenses) J48
- Set 7 (contact-lenses) J48
- Set 9 (contact-lenses) J48
- Set 8 (contact-lenses) J48
- Set 10 (contact-lenses) J48

Tree View

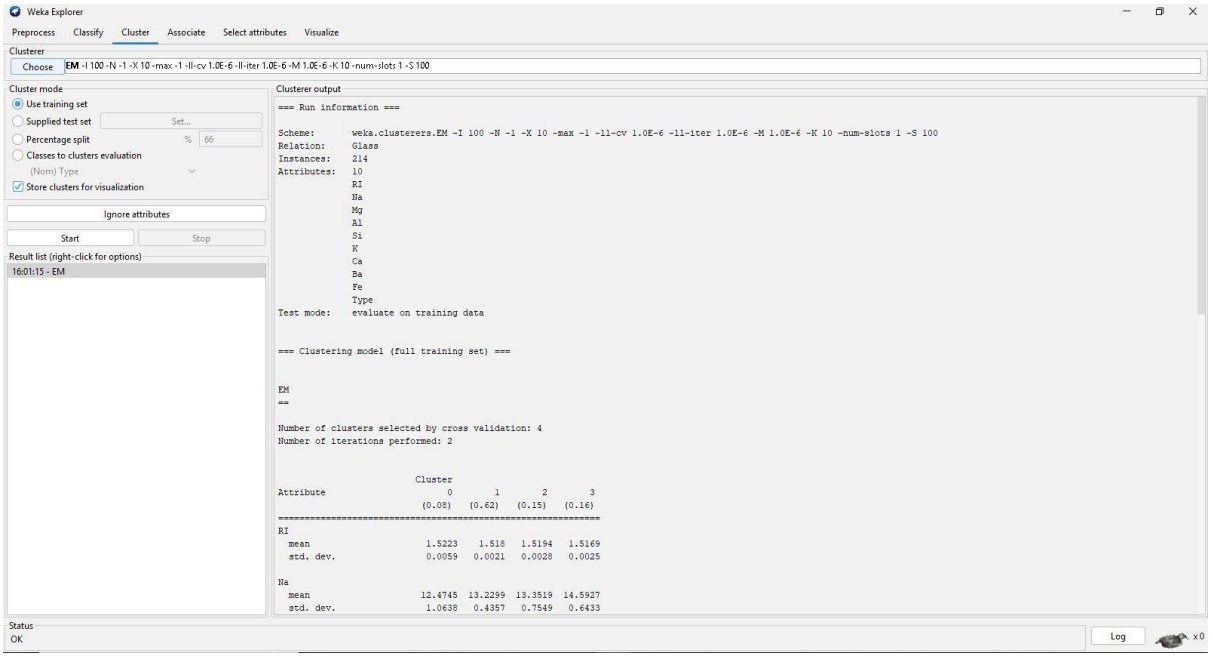


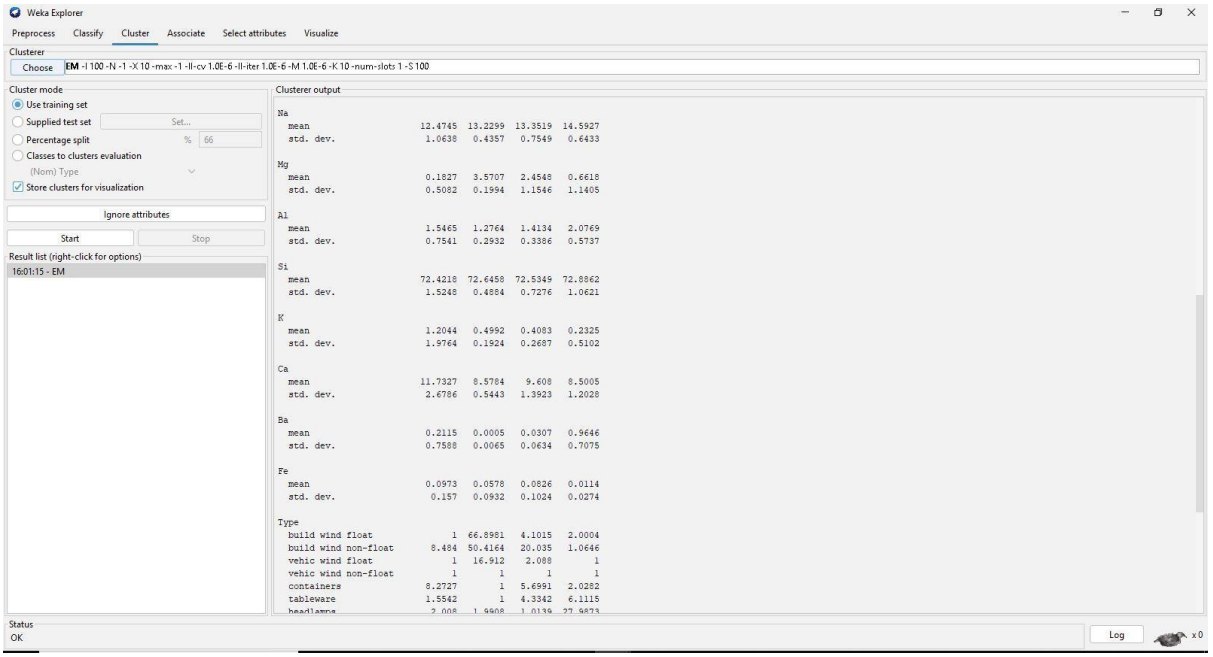

Text Viewer
Result list
15:47:21.793 - J48
Text
=== Evaluation result ===
Scheme: J48
Options: -C 0.25 -M 2
Relation: contact-lenses
=== Summary ===
Correctly Classified Instances 20 83.3333 %
Incorrectly Classified Instances 4 16.6667 %
Kappa statistic 0.71
Mean absolute error 0.15
Root mean squared error 0.3249
Relative absolute error 39.7059 %
Root relative squared error 74.3898 %
Total Number of Instances 24
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
1.000 0.053 0.833 1.000 0.909 0.889 0.947 0.833 soft
0.750 0.100 0.600 0.750 0.667 0.596 0.813 0.592 hard
0.800 0.111 0.923 0.800 0.857 0.669 0.811 0.865 none
Weighted Avg. 0.833 0.097 0.851 0.833 0.836 0.703 0.840 0.813
=== Confusion Matrix ===
a b c <-- classified as
5 0 0 | a = soft
0 3 1 | b = hard
1 2 12 | c = none

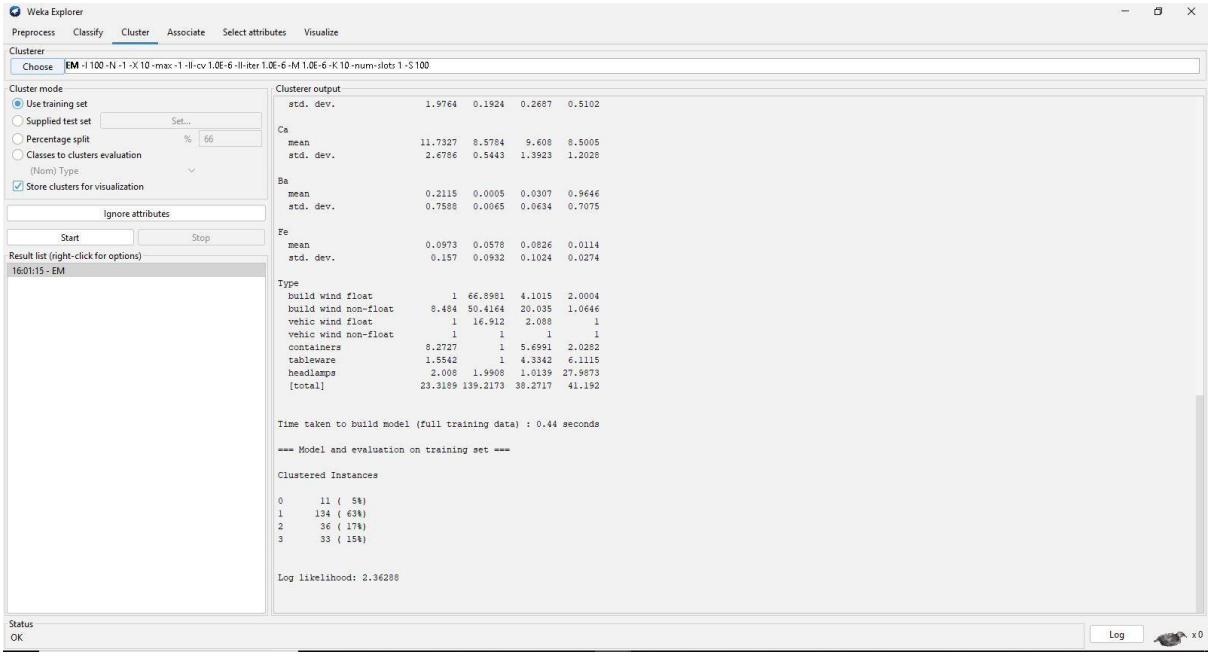

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Type
Start
Stop
Result list (right-click for options)
16:00:59 - rules.ZeroR
Classifier output
=== Run information ===
Scheme: weka.classifiers.rules.ZeroR
Relation: Glass
Instances: 214
Attributes: 10
RI
Na
Mg
Al
Si
K
Ca
Ba
Fe
Type
Test mode: evaluate on training data
=== Classifier model (full training set) ===
ZeroR predicts class value: build wind non-float
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 76 35.514 %
Incorrectly Classified Instances 138 64.486 %
Kappa statistic 0
Mean absolute error 0.2116
Root mean squared error 0.3244
Relative absolute error 100 %
Root relative squared error 100 %
Total Number of Instances 214
=== Detailed Accuracy By Class ===
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) Type
Start
Stop
Result list (right-click for options)
16:00:59 - rules.ZeroR
Classifier output
=== Evaluation on training set ===
Time taken to test model on training data: 0 seconds
=== Summary ===
Correctly Classified Instances 76 35.514 %
Incorrectly Classified Instances 138 64.486 %
Kappa statistic 0
Mean absolute error 0.2116
Root mean squared error 0.3244
Relative absolute error 100 %
Root relative squared error 100 %
Total Number of Instances 214
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
0.000 0.000 ? 0.000 ? ? 0.500 0.327 build wind float
1.000 1.000 0.355 1.000 0.524 ? 0.500 0.355 build wind non-float
0.000 0.000 ? 0.000 ? ? 0.500 0.079 vehic wind float
? 0.000 ? ? ? ? ? ? vehic wind non-float
0.000 0.000 ? 0.000 ? ? 0.500 0.061 containers
0.000 0.000 ? 0.000 ? ? 0.500 0.042 tableware
0.000 0.000 ? 0.000 ? ? 0.500 0.136 headlamps
Weighted Avg. 0.355 0.355 ? 0.355 ? ? 0.500 0.263
=== Confusion Matrix ===
a b c d e f g <-- classified as
0 70 0 0 0 0 0 | a = build wind float
0 76 0 0 0 0 0 | b = build wind non-float
0 17 0 0 0 0 0 | c = vehic wind float
0 0 0 0 0 0 0 | d = vehic wind non-float
0 13 0 0 0 0 0 | e = containers
0 9 0 0 0 0 0 | f = tableware
0 29 0 0 0 0 0 | g = headlamps
Status
OK
Log
x 0

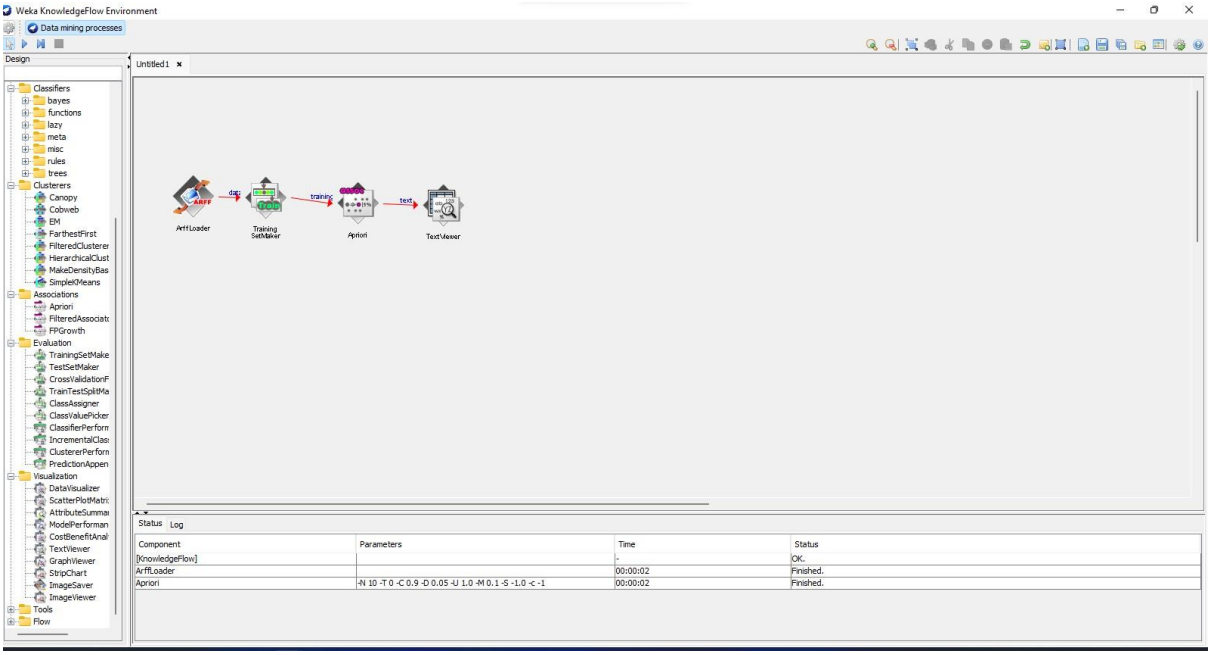
Clustering:

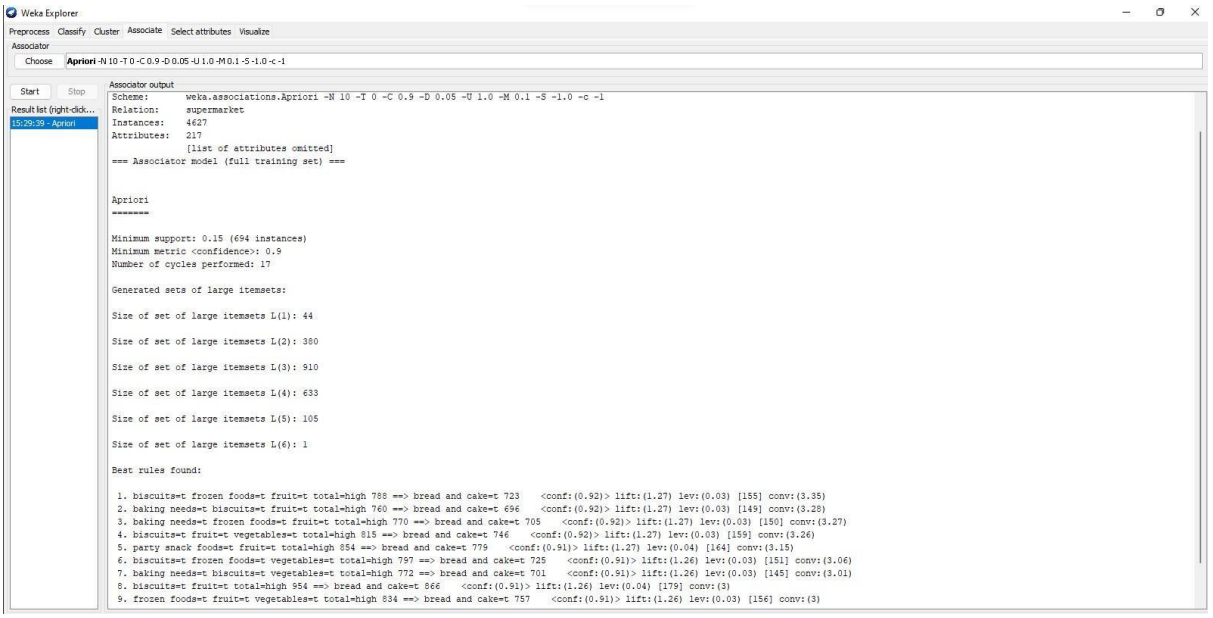

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) Type
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:01:15 - EM
Clusterer output
=== Run information ===
Scheme: weka.clusterers.EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Relation: Glass
Instances: 214
Attributes: 10
RI
Na
Mg
Al
Si
K
Ca
Ba
Fe
Type
Test mode: evaluate on training data
=== Clustering model (full training set) ===
EM
==
Number of clusters selected by cross validation: 4
Number of iterations performed: 2
Cluster
Attribute 0 1 2 3
(0.08) (0.62) (0.15) (0.16)
RI
mean 1.5223 1.518 1.5194 1.5169
std. dev. 0.0059 0.0021 0.0028 0.0025
Na
mean 12.4745 13.2299 13.3519 14.5927
std. dev. 1.0638 0.4357 0.7549 0.6433
Status
OK
Log
x 0

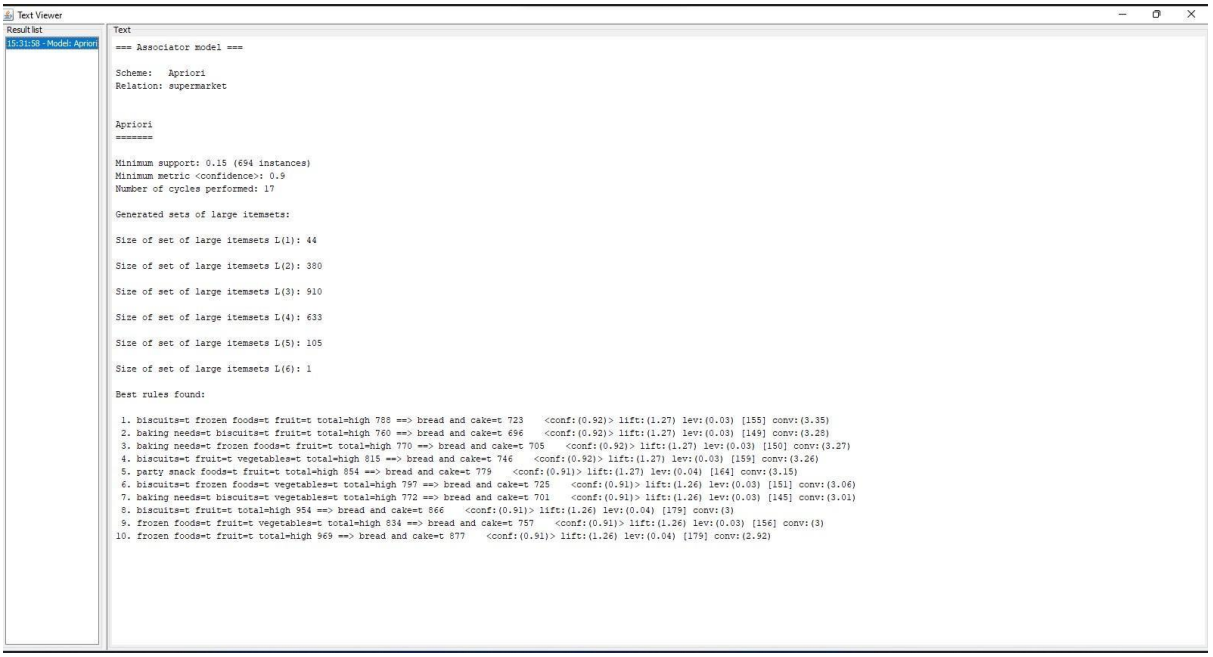

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) Type
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:01:15 - EM
Clusterer output
Na
mean 12.4745 13.2299 13.3519 14.5927
std. dev. 1.0638 0.4357 0.7549 0.6433
Mg
mean 0.1827 3.5707 2.4548 0.6618
std. dev. 0.5082 0.1994 1.1546 1.1405
Al
mean 1.5465 1.2764 1.4134 2.0769
std. dev. 0.7541 0.2932 0.3386 0.5737
Si
mean 72.4218 72.6458 72.5349 72.8862
std. dev. 1.5248 0.4884 0.7276 1.0621
K
mean 1.2044 0.4992 0.4083 0.2325
std. dev. 1.9764 0.1924 0.2687 0.5102
Ca
mean 11.7327 8.5784 9.608 8.5005
std. dev. 2.6786 0.5443 1.3923 1.2028
Ba
mean 0.2115 0.0005 0.0307 0.9646
std. dev. 0.7588 0.0065 0.0634 0.7075
Fe
mean 0.0973 0.0578 0.0826 0.0114
std. dev. 0.157 0.0932 0.1024 0.0274
Type
build wind float 1 66.8981 4.1015 2.0004
build wind non-float 8.484 50.4164 20.035 1.0646
vehic wind float 1 16.912 2.088 1
vehic wind non-float 1 1 1 1
containers 8.2727 1 5.6991 2.0282
tableware 1.5542 1 4.3342 6.1115
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -X 10 -max -1 -ll-cv 1.0E-6 -ll-iter 1.0E-6 -M 1.0E-6 -K 10 -num-slots 1 -S 100
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) Type
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:01:15 - EM
Clusterer output
std. dev. 1.9764 0.1924 0.2687 0.5102
Ca
mean 11.7327 8.5784 9.608 8.5005
std. dev. 2.6786 0.5443 1.3923 1.2028
Ba
mean 0.2115 0.0005 0.0307 0.9646
std. dev. 0.7588 0.0065 0.0634 0.7075
Fe
mean 0.0973 0.0578 0.0826 0.0114
std. dev. 0.157 0.0932 0.1024 0.0274
Type
build wind float 1 66.8981 4.1015 2.0004
build wind non-float 8.484 50.4164 20.035 1.0646
vehic wind float 1 16.912 2.088 1
vehic wind non-float 1 1 1 1
containers 8.2727 1 5.6991 2.0282
tableware 1.5542 1 4.3342 6.1115
headlamps 2.008 1.9908 1.0139 27.9873
[total] 23.3189 139.2173 38.2717 41.192
Time taken to build model (full training data) : 0.44 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 11 (5%)
1 134 (63%)
2 36 (17%)
3 33 (15%)
Log likelihood: 2.36288
Status
OK
Log
x 0

Association:






Text Viewer
Result list
15:31:58 - Model: Apriori
Text
=== Associator model ===
Scheme: Apriori
Relation: supermarket
Apriori
=======
Minimum support: 0.15 (694 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17
Generated sets of large itemsets:
Size of set of large itemsets L(1): 44
Size of set of large itemsets L(2): 380
Size of set of large itemsets L(3): 910
Size of set of large itemsets L(4): 633
Size of set of large itemsets L(5): 105
Size of set of large itemsets L(6): 1
Best rules found:
1. biscuits=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)