Weka Explorer:

1) Naive-bayesian:

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose | NaiveBayes

Test options
- Use training set
- Supplied test set | Set...
- Cross-validation | Folds | 10
- Percentage split | % | 66

More options...

(Nom) Class

Start | Stop

Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes

Classifier output

```
node-caps
  yes                    26.0              32.0
  no                    172.0              52.0
  [total]               198.0              84.0

deg-malig
  1                      60.0              13.0
  2                     103.0              29.0
  3                      41.0              46.0
  [total]               204.0              88.0

breast
  left                  104.0              50.0
  right                  99.0              37.0
  [total]               203.0              87.0

breast-quad
  left_up                72.0              27.0
  left_low               76.0              36.0
  right_up               21.0              14.0
  right_low              19.0               7.0
  central                18.0               5.0
  [total]               206.0              89.0

irradiat
  yes                    38.0              32.0
  no                    165.0              55.0
  [total]               203.0              87.0


Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds
```

Status
OK

Log | x 0

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose NaiveBayes

Test options
- ○ Use training set
- ○ Supplied test set   Set...
- ○ Cross-validation Folds 10
- ○ Percentage split   % 66

More options...

(Nom) Class

Start   Stop

Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes

```
irradiat
        yes              30.0           32.0
        no              165.0           55.0
        [total]         203.0           87.0


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         215               75.1748 %
Incorrectly Classified Instances        71               24.8252 %
Kappa statistic                          0.3693
Mean absolute error                      0.3012
Root mean squared error                  0.4278
Relative absolute error                 72.0082 %
Root relative squared error             93.6095 %
Total Number of Instances              286

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.866    0.518    0.798      0.866   0.831      0.374  0.760     0.879     no-recurrence-events
                 0.482    0.134    0.603      0.482   0.536      0.374  0.760     0.610     recurrence-events
Weighted Avg.    0.752    0.404    0.740      0.752   0.743      0.374  0.760     0.799

=== Confusion Matrix ===

   a   b   <-- classified as
 174  27 |   a = no-recurrence-events
  44  41 |   b = recurrence-events
```
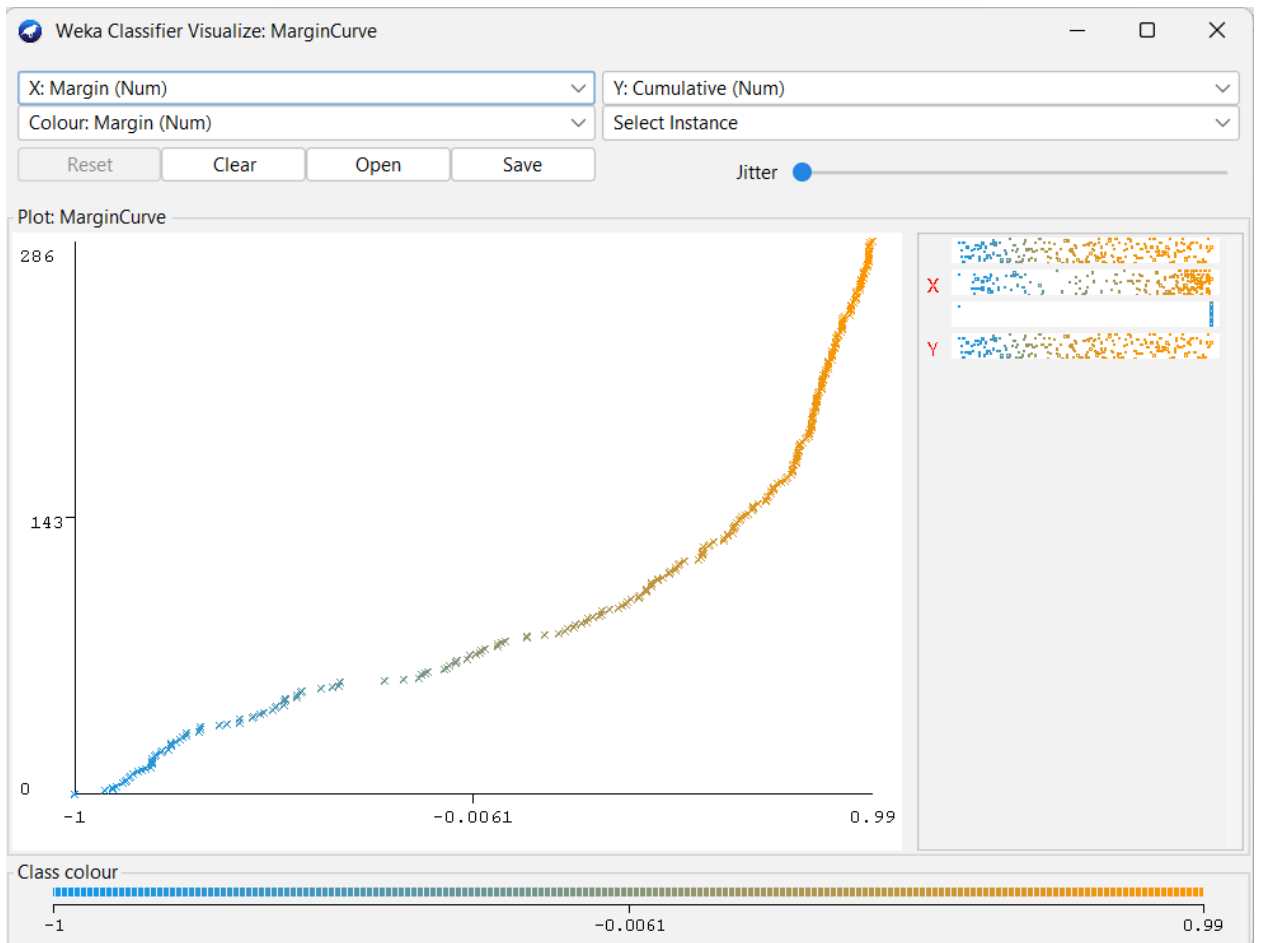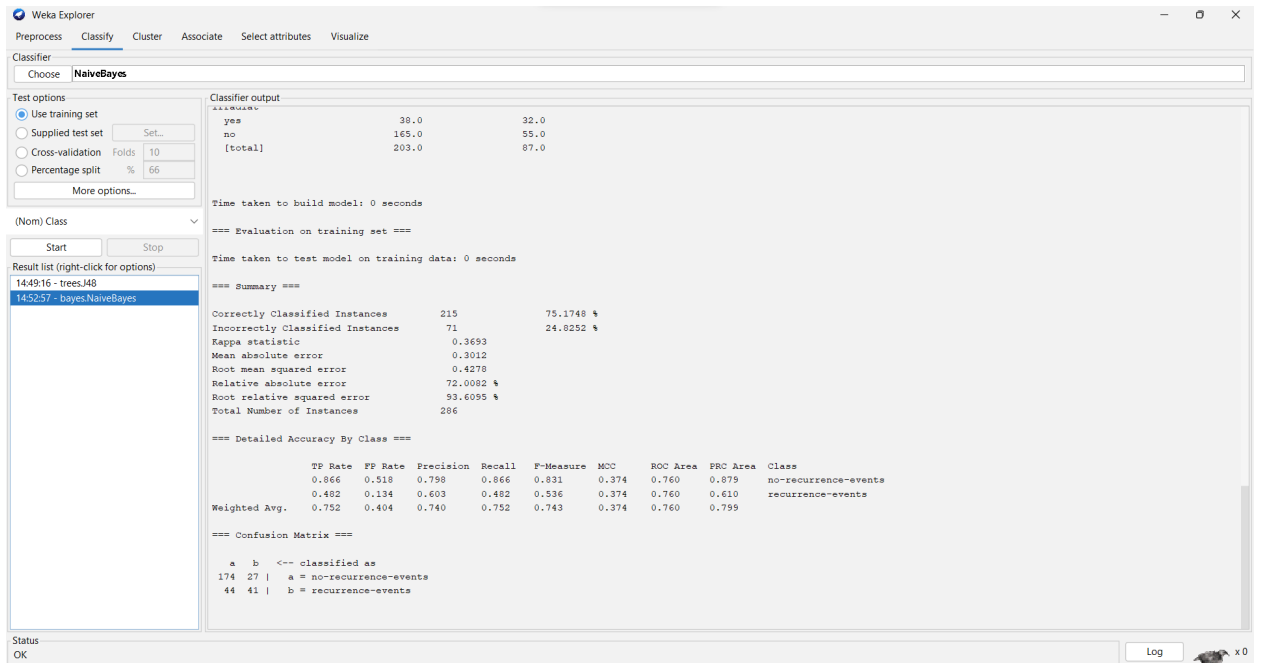
Status
OK   Log   x 0

Weka Classifier Visualize: MarginCurve

X: Margin (Num)   Y: Cumulative (Num)

Colour: Margin (Num)   Select Instance

Reset   Clear   Open   Save   Jitter

Plot: MarginCurve

Class colour
-1   -0.0061   0.99

2) Classification for tree and and j48:


```
=== 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
Test mode:    evaluate on training data

=== 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: 0.01 seconds

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



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

Number of Leaves  :     4

Size of the tree :      6


Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===

Correctly Classified Instances         217               75.8741 %
Incorrectly Classified Instances        69               24.1259 %
Kappa statistic                          0.2899
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
```

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose | DecisionStump

Test options
- Use training set
- Supplied test set | Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) Class

Start | Stop

Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
15:02:07 - trees.DecisionStump

Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.trees.DecisionStump
Relation:     breast-cancer
Instances:    286
Attributes:   10
              age
              menopause
              tumor-size
              inv-nodes
              node-caps
              deg-malig
              breast
              breast-quad
              irradiat
              Class
Test mode:    evaluate on training data

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

Decision Stump

Classifications

deg-malig = 3 : recurrence-events
deg-malig != 3 : no-recurrence-events
deg-malig is missing : no-recurrence-events

Class distributions

deg-malig = 3
no-recurrence-events    recurrence-events
0.47058823529411764     0.5294117647058824
deg-malig != 3
no-recurrence-events    recurrence-events
0.8009950248756219      0.19900497512437812
deg-malig is missing
```

Status
OK

Log

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Classifier

Choose | DecisionStump

Test options
- Use training set
- Supplied test set | Set...
- Cross-validation Folds 10
- Percentage split % 66

More options...

(Nom) Class

Start | Stop

Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
15:02:07 - trees.DecisionStump

Classifier output

```
no-recurrence-events    recurrence-events
0.8009950248756219      0.19900497512437812
deg-malig is missing
no-recurrence-events    recurrence-events
0.7027972027972028      0.2972027972027972


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         206               72.028 %
Incorrectly Classified Instances        80               27.972 %
Kappa statistic                          0.3304
Mean absolute error                      0.3721
Root mean squared error                  0.4314
Relative absolute error                 88.9615 %
Root relative squared error             94.3834 %
Total Number of Instances              286

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall  F-Measure  MCC    ROC Area  PRC Area  Class
                 0.801    0.471    0.801      0.801   0.801      0.330  0.665     0.781     no-recurrence-events
                 0.529    0.199    0.529      0.529   0.529      0.330  0.665     0.420     recurrence-events
Weighted Avg.    0.720    0.390    0.720      0.720   0.720      0.330  0.665     0.674

=== Confusion Matrix ===

   a   b   <-- classified as
 161  40 |   a = no-recurrence-events
  40  45 |   b = recurrence-events
```

Status
OK

Log

3) K-means i.e Clustering

Weka Explorer (top window)

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Clusterer

Choose | SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode
- ( ) Use training set
- ( ) Supplied test set   Set...
- ( ) Percentage split   % 66
- ( ) Classes to clusters evaluation
- (Nom) Class
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)
14:59:28 - SimpleKMeans

Clusterer output

```
=== Run information ===

Scheme:       weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanD:
Relation:     breast-cancer
Instances:    286
Attributes:   10
              age
              menopause
              tumor-size
              inv-nodes
              node-caps
              deg-malig
              breast
              breast-quad
              irradiat
              Class
Test mode:    evaluate on training data


=== Clustering model (full training set) ===


kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:
                                                  Cluster#
```

Status: OK   Log   x 0

Weka Explorer (bottom window)

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Clusterer

Choose | SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Cluster mode
- ( ) Use training set
- ( ) Supplied test set   Set...
- ( ) Percentage split   % 66
- ( ) Classes to clusters evaluation
- (Nom) Class
- [x] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)
14:59:28 - SimpleKMeans

Clusterer output

```
Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:
                                             Cluster#
Attribute                  Full Data              0                 1
                             (286.0)          (225.0)           (61.0)
=============================================================================
age                           50-59            50-59             40-49
menopause                   premeno          premeno           premeno
tumor-size                    30-34            25-29             30-34
inv-nodes                       0-2              0-2               0-2
node-caps                        no               no               yes
deg-malig                         2                2                 3
breast                         left             left              left
breast-quad                left_low         left_low          left_low
irradiat                         no               no                no
Class           no-recurrence-events no-recurrence-events  recurrence-events



Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0      225 ( 79%)
1       61 ( 21%)
```
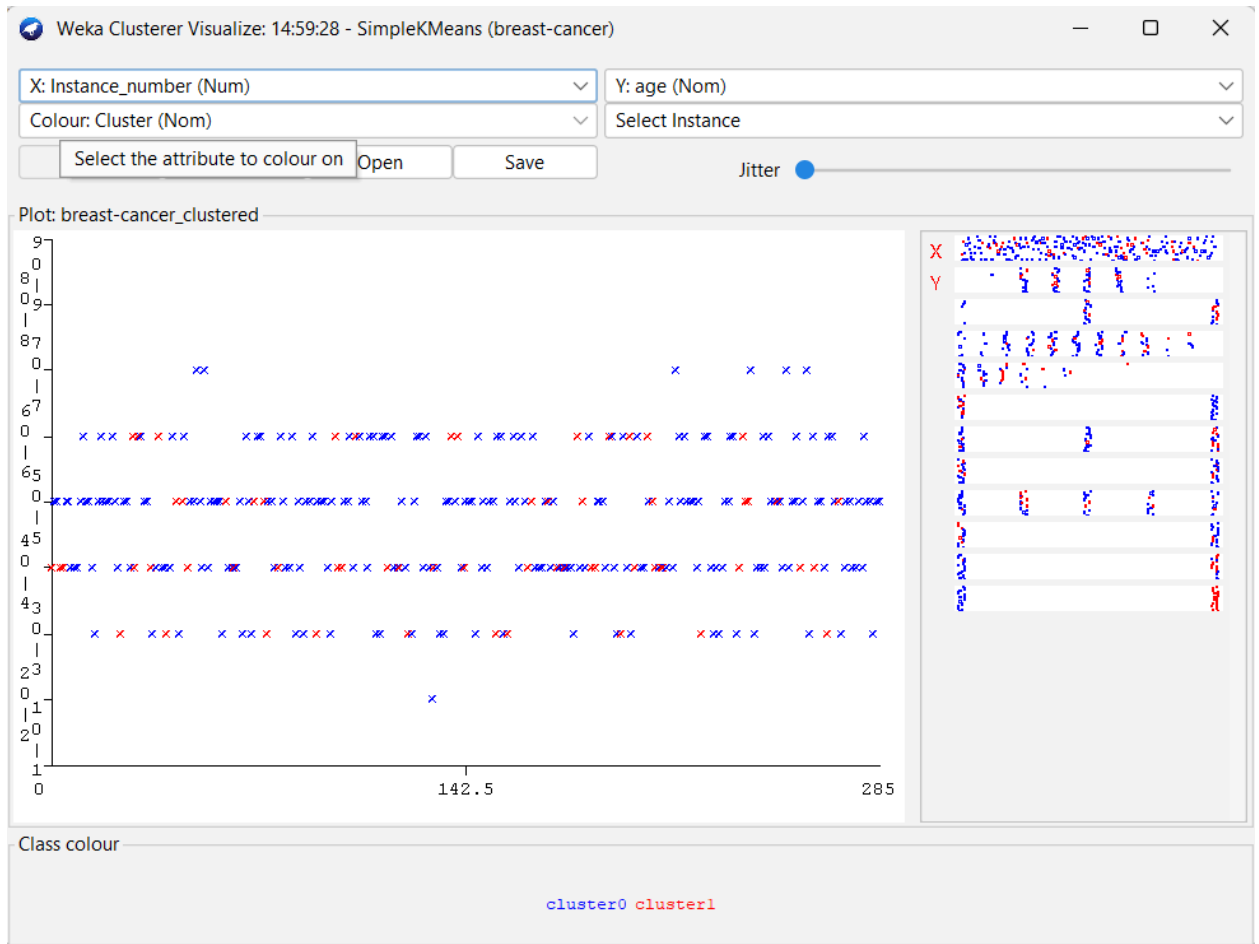
Status: OK   Log   x 0

4) Apriori:

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Associator

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

Start | Stop

Associator output

Result list (right-click for ...)
15:07:28 - Apriori

```
=== 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:     breast-cancer
Instances:    286
Attributes:   10
              age
              menopause
              tumor-size
              inv-nodes
              node-caps
              deg-malig
              breast
              breast-quad
              irradiat
              Class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.5 (143 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10

Generated sets of large itemsets:

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

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:
```

Status
OK

Log | x 0

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Associator

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

Start | Stop

Associator output

Result list (right-click for ...)
15:07:28 - Apriori

```
              breast
              breast-quad
              irradiat
              Class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.5 (143 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 10

Generated sets of large itemsets:

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

Size of set of large itemsets L(2): 6

Size of set of large itemsets L(3): 4

Size of set of large itemsets L(4): 1

Best rules found:

 1. inv-nodes=0-2 irradiat=no Class=no-recurrence-events 147 ==> node-caps=no 145    <conf:(0.99)> lift:(1.27) lev:(0.11) [30] conv:(10.97)
 2. inv-nodes=0-2 irradiat=no 183 ==> node-caps=no 177    <conf:(0.97)> lift:(1.25) lev:(0.12) [34] conv:(5.85)
 3. node-caps=no irradiat=no Class=no-recurrence-events 151 ==> inv-nodes=0-2 145    <conf:(0.96)> lift:(1.29) lev:(0.11) [32] conv:(5.51)
 4. inv-nodes=0-2 Class=no-recurrence-events 167 ==> node-caps=no 160    <conf:(0.96)> lift:(1.23) lev:(0.11) [30] conv:(4.67)
 5. inv-nodes=0-2 213 ==> node-caps=no 201    <conf:(0.94)> lift:(1.22) lev:(0.12) [35] conv:(3.67)
 6. node-caps=no irradiat=no 188 ==> inv-nodes=0-2 177    <conf:(0.94)> lift:(1.26) lev:(0.13) [36] conv:(4)
 7. node-caps=no Class=no-recurrence-events 171 ==> inv-nodes=0-2 160    <conf:(0.94)> lift:(1.26) lev:(0.11) [32] conv:(3.64)
 8. irradiat=no Class=no-recurrence-events 164 ==> node-caps=no 151    <conf:(0.92)> lift:(1.19) lev:(0.08) [23] conv:(2.62)
 9. inv-nodes=0-2 node-caps=no Class=no-recurrence-events 160 ==> irradiat=no 145    <conf:(0.91)> lift:(1.19) lev:(0.08) [23] conv:(2.38)
10. node-caps=no 222 ==> inv-nodes=0-2 201    <conf:(0.91)> lift:(1.22) lev:(0.12) [35] conv:(2.58)
```

Status
OK

Log | x 0

Weka

Knowledge Base:

1) naive-bayesian

2) Classification

=== Evaluation result ===

Scheme: J48
Options: -C 0.25 -M 2
Relation: breast-cancer

=== Summary ===

Correctly Classified Instances	216	75.5245 %
Incorrectly Classified Instances	70	24.4755 %
Kappa statistic	0.2826	
Mean absolute error	0.3676	
Root mean squared error	0.4324	
Relative absolute error	87.8635 %	
Root relative squared error	94.6093 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.960	0.729	0.757	0.960	0.846	0.339	0.584	0.736	no-recurrence-events
	0.271	0.040	0.742	0.271	0.397	0.339	0.584	0.436	recurrence-events
Weighted Avg.	0.755	0.524	0.752	0.755	0.713	0.339	0.584	0.647	

=== Confusion Matrix ===

```
   a   b   <-- classified as
 193   8 |   a = no-recurrence-events
  62  23 |   b = recurrence-events
```

Graph Viewer

Tree View

Result list
- Set 3 (breast-cancer) J48
- Set 4 (breast-cancer) J48
- Set 6 (breast-cancer) J48
- Set 2 (breast-cancer) J48
- Set 5 (breast-cancer) J48
- Set 1 (breast-cancer) J48
- Set 7 (breast-cancer) J48
- Set 9 (breast-cancer) J48
- Set 8 (breast-cancer) J48
- Set 10 (breast-cancer) J48

Tree:
- node-caps
  - = yes → deg-malig
    - = 1 → recurrence-events (0.58/0.19)
    - = 2 → no-recurrence-events (24.19/7.0)
    - = 3 → recurrence-events (25.39/5.39)
  - = no → no-recurrence-events (207.83/49.61)

Close   Clear results
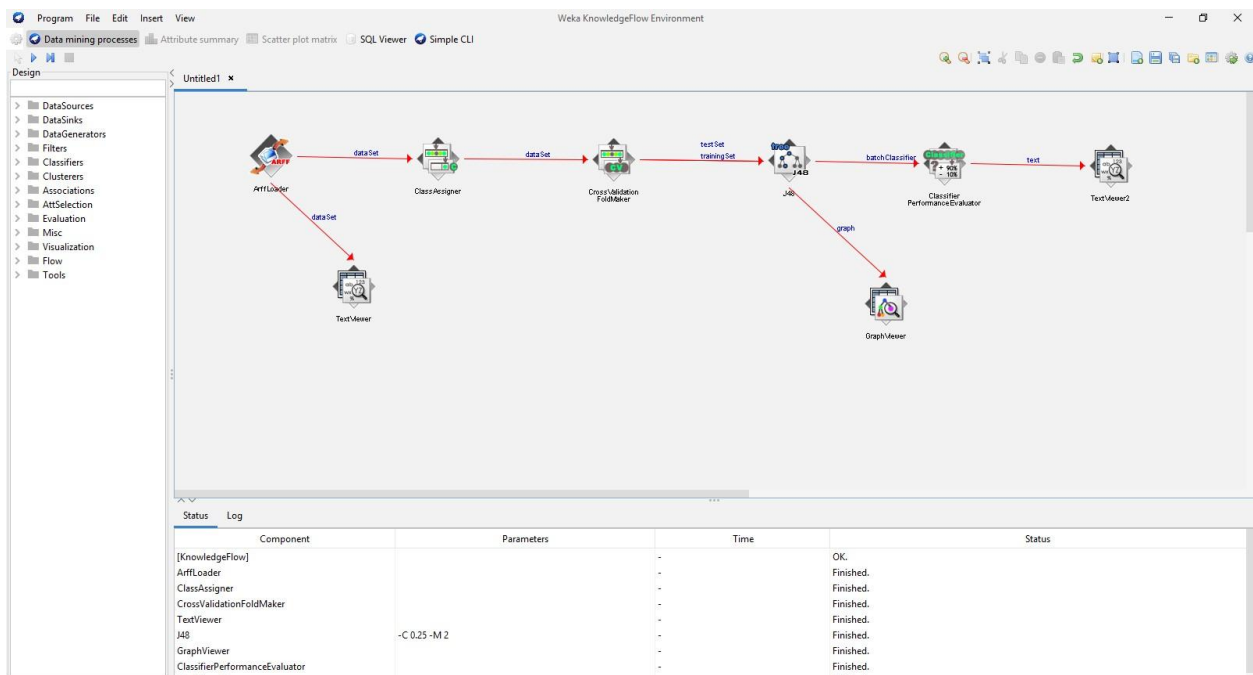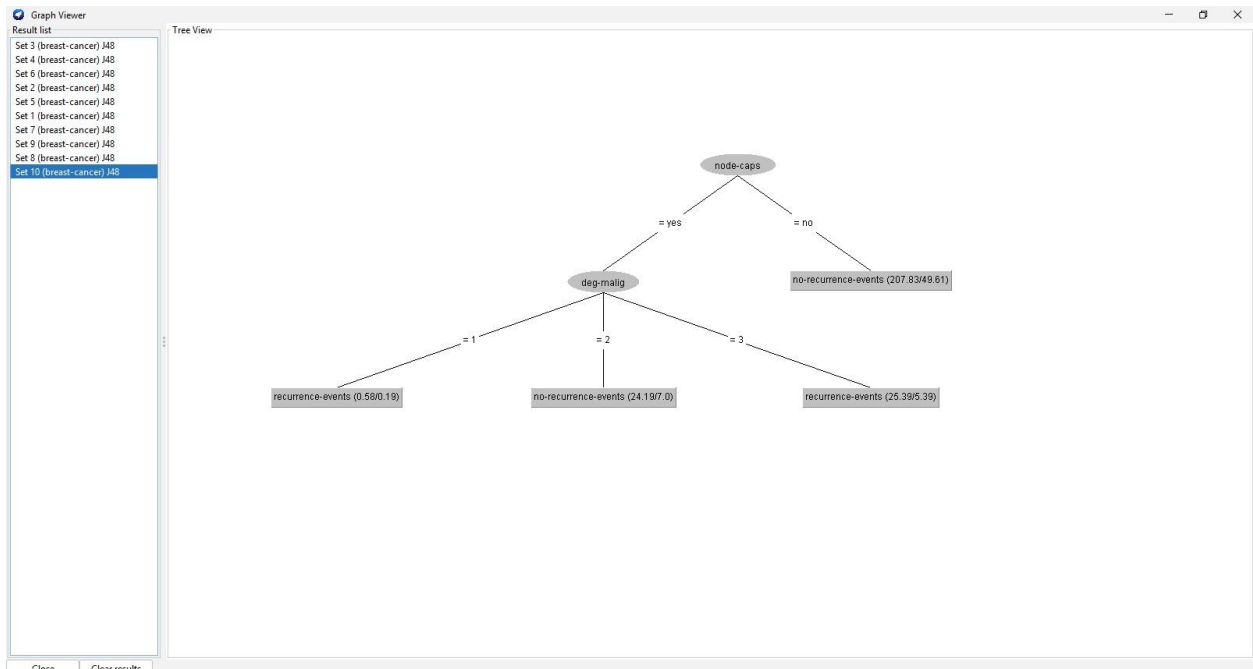
Program   File   Edit   Insert   View   Weka KnowledgeFlow Environment

Data mining processes   Attribute summary   Scatter plot matrix   SQL Viewer   Simple CLI

Design

- DataSources
- DataSinks
- DataGenerators
- Filters
- Classifiers
- Clusterers
- Associations
- AttSelection
- Evaluation
- Misc
- Visualization
- Flow
- Tools

Untitled1

ArffLoader → dataSet → ClassAssigner → dataSet → CrossValidation FoldMaker → testSet / trainingSet → J48 → batchClassifier → Classifier PerformanceEvaluator → text → TextViewer2

ArffLoader → dataSet → TextViewer

J48 → graph → GraphViewer

Status   Log

Component	Parameters	Time	Status
[KnowledgeFlow]		-	OK.
ArffLoader		-	Finished.
ClassAssigner		-	Finished.
CrossValidationFoldMaker		-	Finished.
TextViewer		-	Finished.
J48	-C 0.25 -M 2	-	Finished.
GraphViewer		-	Finished.
ClassifierPerformanceEvaluator		-	Finished.

3) k-means:

4) Apriori:

Data mining processes   Attribute summary   Scatter plot matrix   SQL Viewer   Simple CLI

Design

ARFFLoader
DataGrid
> DataSinks
> DataGenerators
> Filters
> Classifiers
> Clusterers
∨ Associations
   Apriori
   FilteredAssociator
   FPGrowth
> AttSelection
∨ Evaluation
   TrainingSetMaker
   TestSetMaker
   TrainTestSplitMaker
   ClassAssigner
   ClassValuePicker
   ClassifierPerformanceEv
   ClustererPerformanceEv
   CrossValidationFoldMak
   PredictionAppender
   IncrementalClassifierEva
> Misc
∨ Visualization
   TextViewer
   ImageViewer
   AttributeSummarizer
   StripChart
   ModelPerformanceChar
   DataVisualizer
   BoundaryPlotter
   ScatterPlotMatrix
   GraphViewer
   CostBenefitAnalysis
> Flow
> Tools

Untitled1

ArffLoader   dataSet → TrainingSetMaker   trainingSet → Apriori   text → TextViewer

Status   Log

Component	Parameters	Time	Status
[KnowledgeFlow]		-	OK.
ArffLoader		-	Finished.
TrainingSetMaker		-	Finished.
Apriori	-N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1	-	Finished.
TextViewer		-	Finished.

Type here to search   ENG 15:07:50 06-10-2022

Text Viewer

Result list
15:05:08.082 - Model: Apri
15:06:46.715 - Model: Apri

Text

=== Associator model ===

Scheme: Apriori
Relation: supermarket

Apriori
=======

Minimum support: 0.3 (1388 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 25

Size of set of large itemsets L(2): 69

Size of set of large itemsets L(3): 20

Best rules found:

 1. biscuits=t vegetables=t 1764 ==> bread and cake=t 1487   <conf:(0.84)> lift:(1.17) lev:(0.05) [217] conv:(1.78)
 2. total=high 1679 ==> bread and cake=t 1413   <conf:(0.84)> lift:(1.17) lev:(0.04) [204] conv:(1.76)
 3. biscuits=t milk-cream=t 1767 ==> bread and cake=t 1485   <conf:(0.84)> lift:(1.17) lev:(0.05) [213] conv:(1.75)
 4. biscuits=t fruit=t 1837 ==> bread and cake=t 1541   <conf:(0.84)> lift:(1.17) lev:(0.05) [218] conv:(1.73)
 5. biscuits=t frozen foods=t 1810 ==> bread and cake=t 1510   <conf:(0.83)> lift:(1.16) lev:(0.04) [207] conv:(1.69)
 6. frozen foods=t fruit=t 1861 ==> bread and cake=t 1548   <conf:(0.83)> lift:(1.16) lev:(0.05) [208] conv:(1.66)
 7. frozen foods=t milk-cream=t 1826 ==> bread and cake=t 1516   <conf:(0.83)> lift:(1.15) lev:(0.04) [201] conv:(1.65)
 8. baking needs=t milk-cream=t 1907 ==> bread and cake=t 1580   <conf:(0.83)> lift:(1.15) lev:(0.04) [207] conv:(1.63)
 9. milk-cream=t fruit=t 2038 ==> bread and cake=t 1684   <conf:(0.83)> lift:(1.15) lev:(0.05) [217] conv:(1.61)
10. baking needs=t biscuits=t 1764 ==> bread and cake=t 1456   <conf:(0.83)> lift:(1.15) lev:(0.04) [186] conv:(1.6)

Close   Settings   Clear results