BBS654
Data Mining

Pinar Duygulu

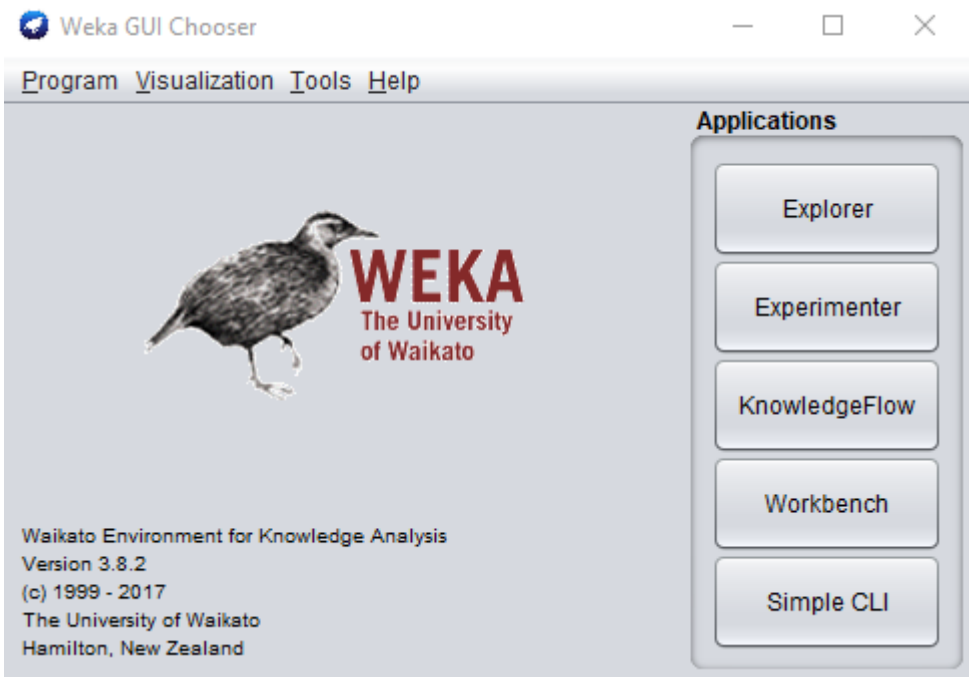
Slides are adapted from
https://www.cs.waikato.ac.nz/ml/weka/

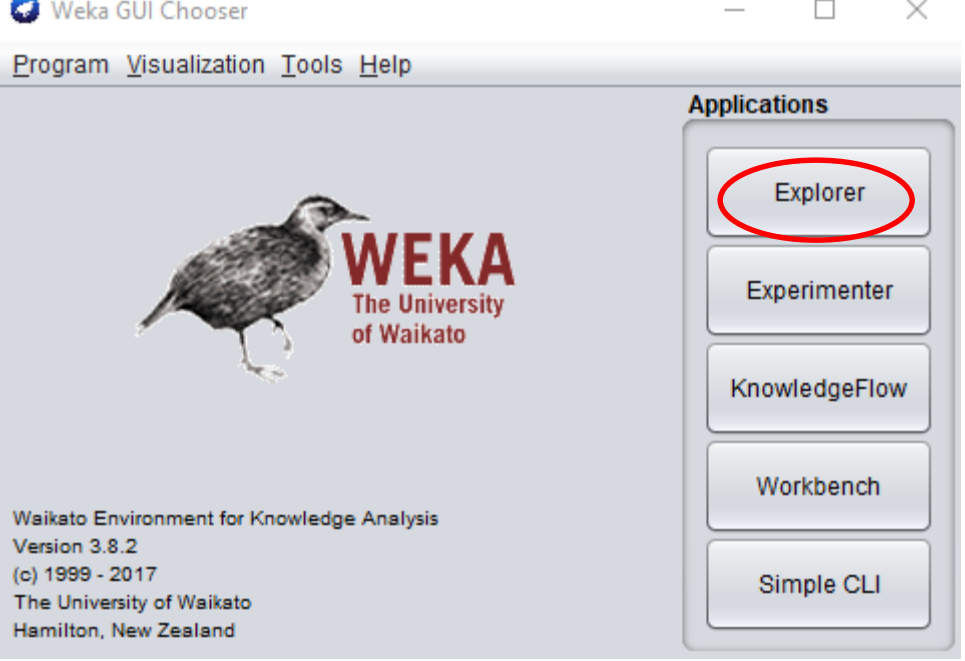
WEKA

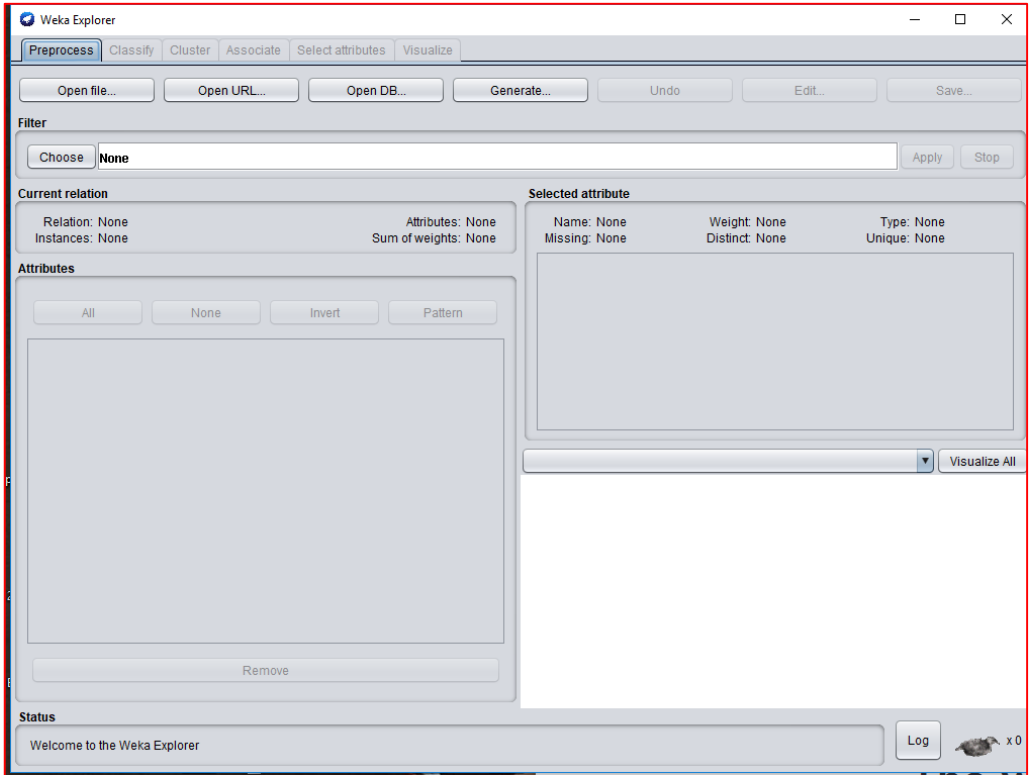
- Data Mining Software in Java
- Weka is a collection of machine learning algorithms for data mining tasks. The algorithms can either be applied directly to a dataset or called from your own Java code. Weka contains tools for data pre-processing, classification, regression, clustering, association rules, and visualization. It is also well-suited for developing new machine learning schemes.

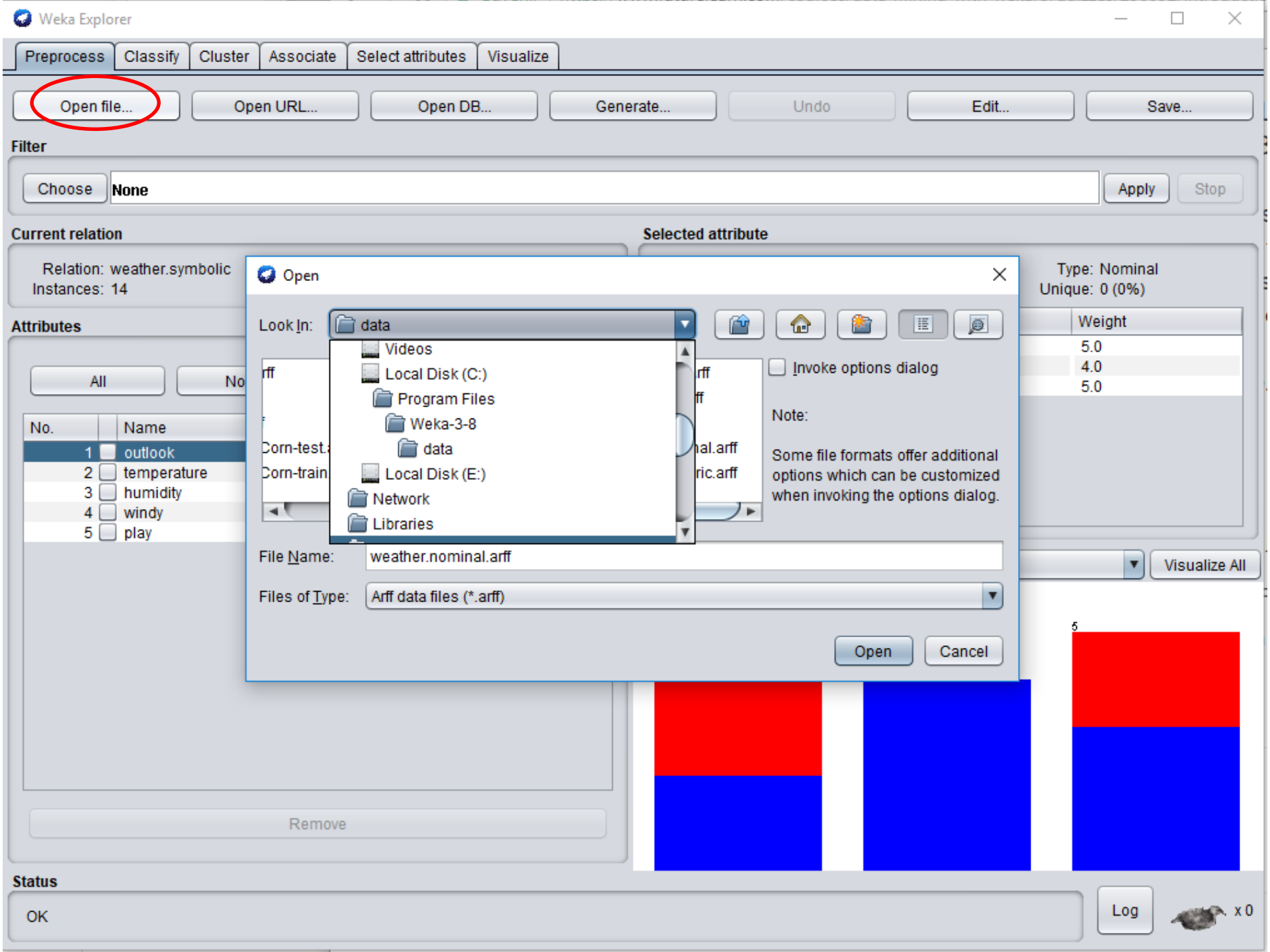
How to install WEKA

- Go to http://www.cs.waikato.ac.nz/ml/weka
- Click the Download button




Weka GUI Chooser
Program Visualization Tools Help
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI
WEKA
The University of Waikato
Waikato Environment for Knowledge Analysis
Version 3.8.2
(c) 1999 - 2017
The University of Waikato
Hamilton, New Zealand


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: None
Instances: None
Attributes: None
Sum of weights: None
Selected attribute
Name: None
Missing: None
Weight: None
Distinct: None
Type: None
Unique: None
Attributes
All
None
Invert
Pattern
Remove
Visualize All
Status
Welcome to the Weka Explorer
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Stop
Current relation
Relation: weather.symbolic
Instances: 14
Selected attribute
Type: Nominal
Unique: 0 (0%)
Attributes
All
No
No.
Name
1 outlook
2 temperature
3 humidity
4 windy
5 play
Weight
5.0
4.0
5.0
Visualize All
Remove
Status
OK
Log
x 0
Open
Look In:
data
Videos
Local Disk (C:)
Program Files
Weka-3-8
data
Local Disk (E:)
Network
Libraries
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
weather.nominal.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel

Viewer

Relation: weather.symbolic

No.	1: outlook	2: temperature	3: humidity	4: windy	5: **play**
	Nominal	Nominal	Nominal	Nominal	Nominal
1	sunny	hot	high	FALSE	no
2	sunny	hot	high	TRUE	no
3	overcast	hot	high	FALSE	yes
4	rainy	mild	high	FALSE	yes
5	rainy	cool	normal	FALSE	yes
6	rainy	cool	normal	TRUE	no
7	overcast	cool	normal	TRUE	yes
8	sunny	mild	high	FALSE	no
9	sunny	cool	normal	FALSE	yes
10	rainy	mild	normal	FALSE	yes
11	sunny	mild	normal	TRUE	yes
12	overcast	mild	high	TRUE	yes
13	overcast	hot	normal	FALSE	yes
14	rainy	mild	high	TRUE	no

Weka Explorer

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

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

Filter

Choose | None | Apply | Stop

Current relation

Relation: iris
Instances: 150
Attributes: 5
Sum of weights: 150

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	sepallength
2	☐	sepalwidth
3	☐	petallength
4	☐	petalwidth
5	☐	class

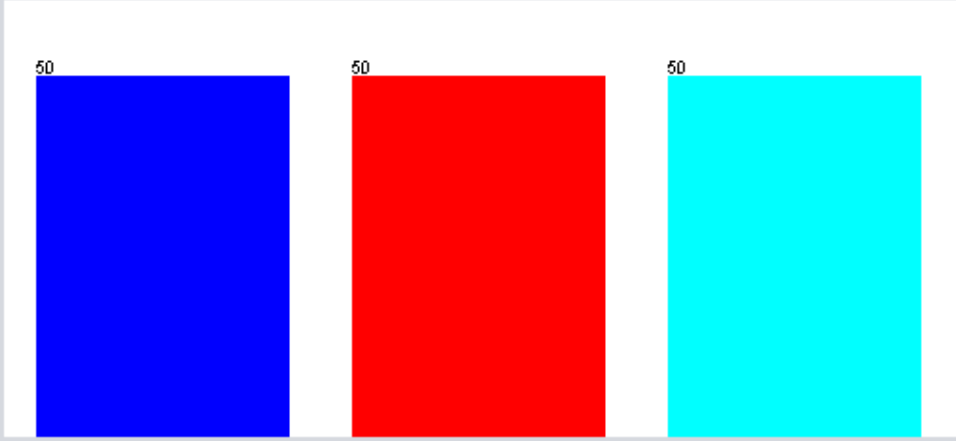
Remove

Selected attribute

Name: class
Missing: 0 (0%)
Distinct: 3
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	Iris-setosa	50	50.0
2	Iris-versicolor	50	50.0
3	Iris-virginica	50	50.0

Class: class (Nom) | Visualize All



Status

OK

Log | x 0

Weka Explorer

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

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

Filter

Choose **None** | Apply | Stop

Current relation

Relation: iris
Instances: 150
Attributes: 5
Sum of weights: 150

Attributes

All | None | Invert | Pattern

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

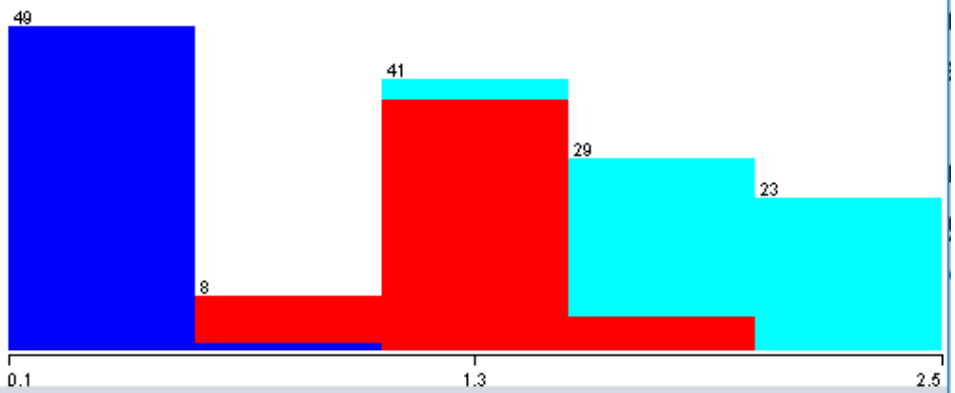
Remove

Selected attribute

Name: petalwidth
Missing: 0 (0%)
Distinct: 22
Type: Numeric
Unique: 2 (1%)

Statistic	Value
Minimum	0.1
Maximum	2.5
Mean	1.199
StdDev	0.763

Class: class (Nom) | Visualize All



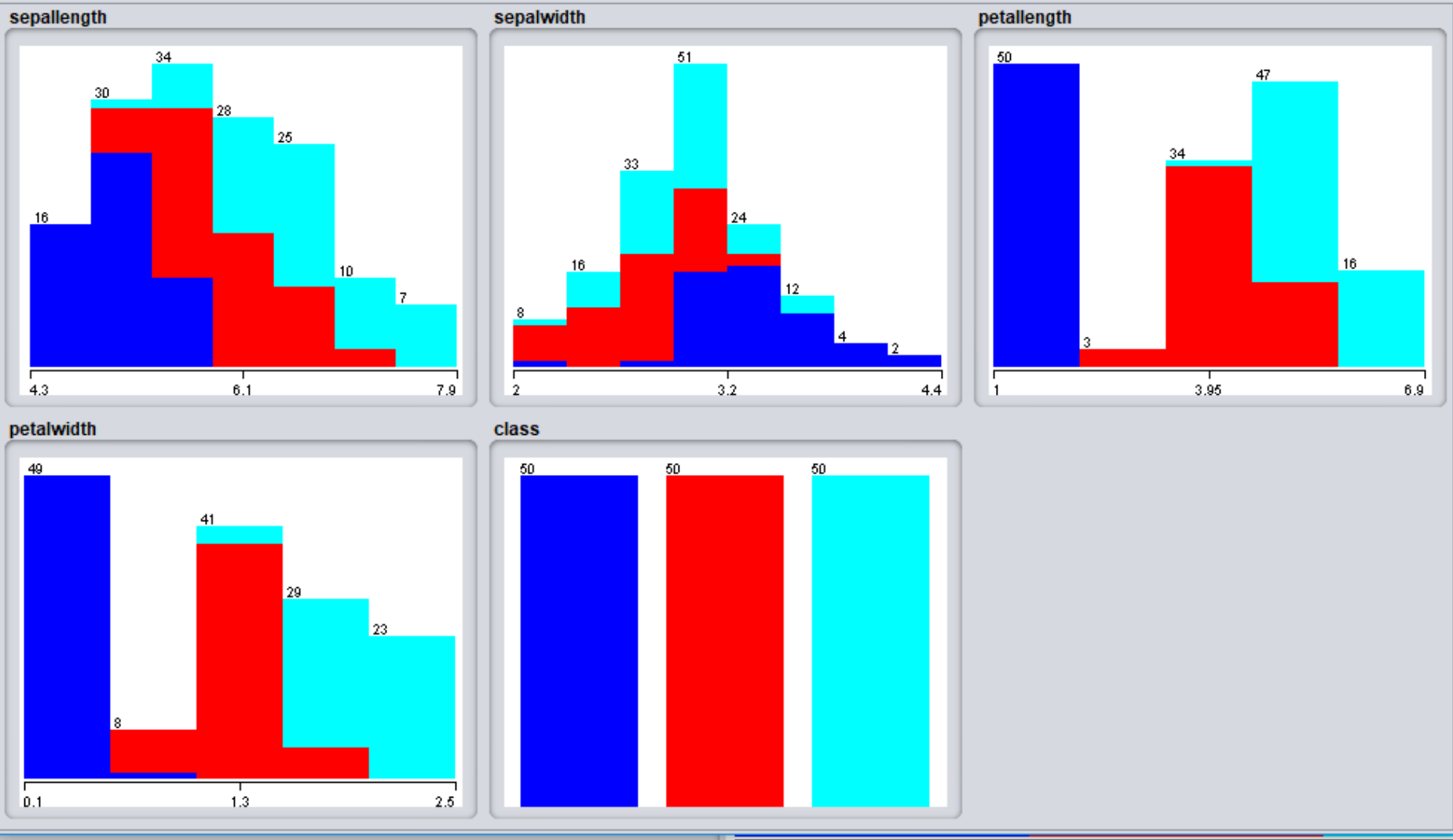
Status

OK | Log | x 0

Weka Explorer

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

All attributes



Status

OK

Log

x 0

```
% 1. Title: Iris Plants Database
%
% 2. Sources:
%      (a) Creator: R.A. Fisher
%      (b) Donor: Michael Marshall (MARSHALL%
PLU@io.arc.nasa.gov)
%      (c) Date: July, 1988
%
% 3. Past Usage:
%    - Publications: too many to mention!!!  Here are a few.
%    1. Fisher,R.A. "The use of multiple measurements in
taxonomic problems"
%       Annual Eugenics, 7, Part II, 179-188 (1936); also in
"Contributions
%       to Mathematical Statistics" (John Wiley, NY, 1950).
%    2. Duda,R.O., & Hart,P.E. (1973) Pattern Classification and
Scene Analysis.
%       (Q327.D83) John Wiley & Sons.  ISBN 0-471-22361-1.  See
page 218.
%    3. Dasarathy, B.V. (1980) "Nosing Around the Neighborhood:
A New System
%       Structure and Classification Rule for Recognition in
Partially Exposed
```

```
%
% 7. Attribute Information:
%    1. sepal length in cm
%    2. sepal width in cm
%    3. petal length in cm
%    4. petal width in cm
%    5. class:
%       -- Iris Setosa
%       -- Iris Versicolour
%       -- Iris Virginica
%
% 8. Missing Attribute Values: None
%
% Summary Statistics:
%                Min  Max   Mean    SD   Class Correlation
%    sepal length: 4.3  7.9   5.84  0.83    0.7826
%     sepal width: 2.0  4.4   3.05  0.43   -0.4194
%    petal length: 1.0  6.9   3.76  1.76    0.9490  (high!)
%     petal width: 0.1  2.5   1.20  0.76    0.9565  (high!)
%
% 9. Class Distribution: 33.3% for each of 3 classes.

@RELATION iris

@ATTRIBUTE sepallength     REAL
@ATTRIBUTE sepalwidth      REAL
@ATTRIBUTE petallength     REAL
@ATTRIBUTE petalwidth REAL
@ATTRIBUTE class      {Iris-setosa,Iris-versicolor,Iris-
virginica}

@DATA
5.1,3.5,1.4,0.2,Iris-setosa
4.9,3.0,1.4,0.2,Iris-setosa
4.7,3.2,1.3,0.2,Iris-setosa
4.6,3.1,1.5,0.2,Iris-setosa
5.0,3.6,1.4,0.2,Iris-setosa
5.4,3.9,1.7,0.4,Iris-setosa
```

Classification

Weka Explorer

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

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) class

Start | Stop

Result list (right-click for options)

14:55:30 - trees.J48

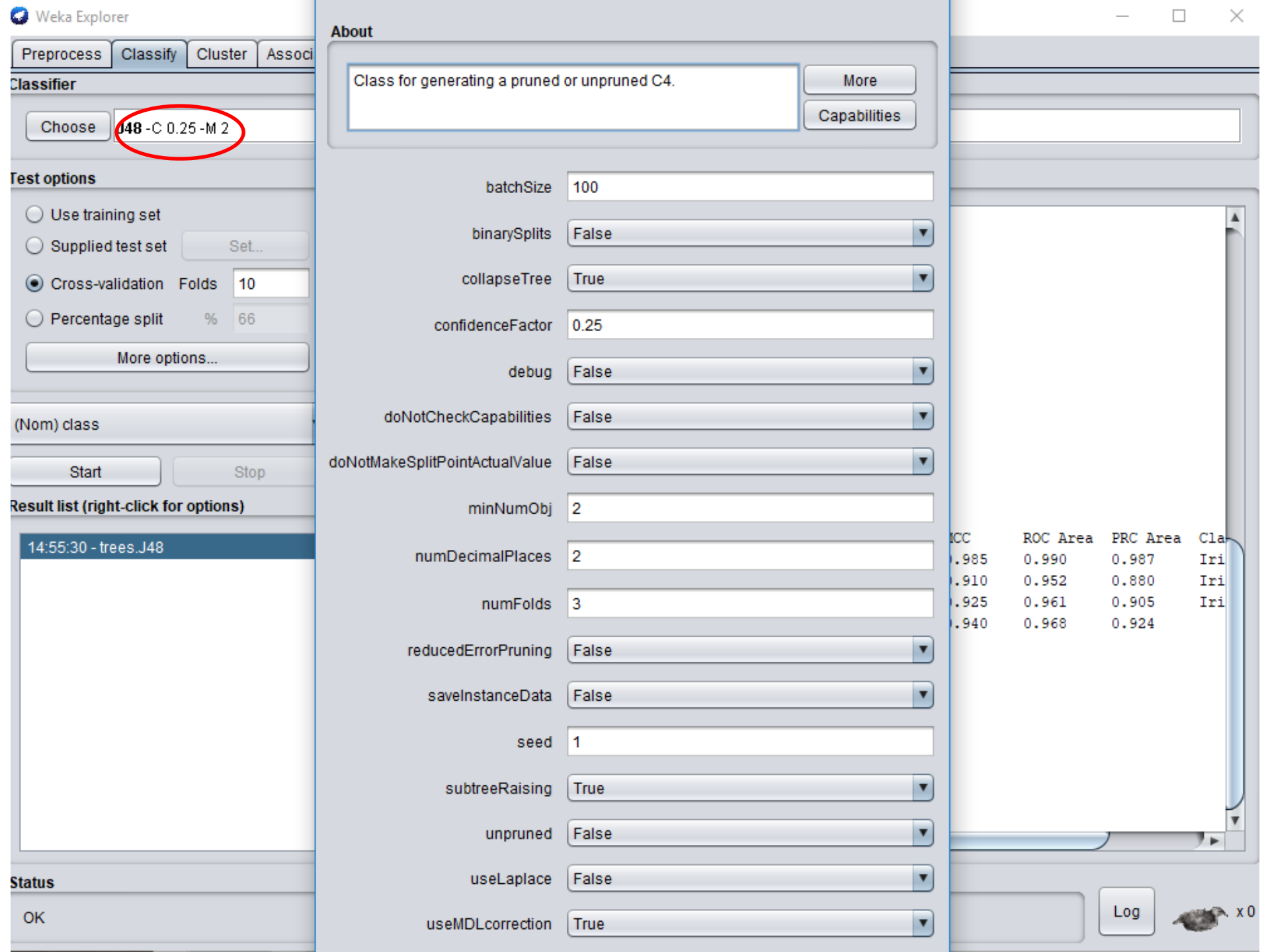
Classifier output

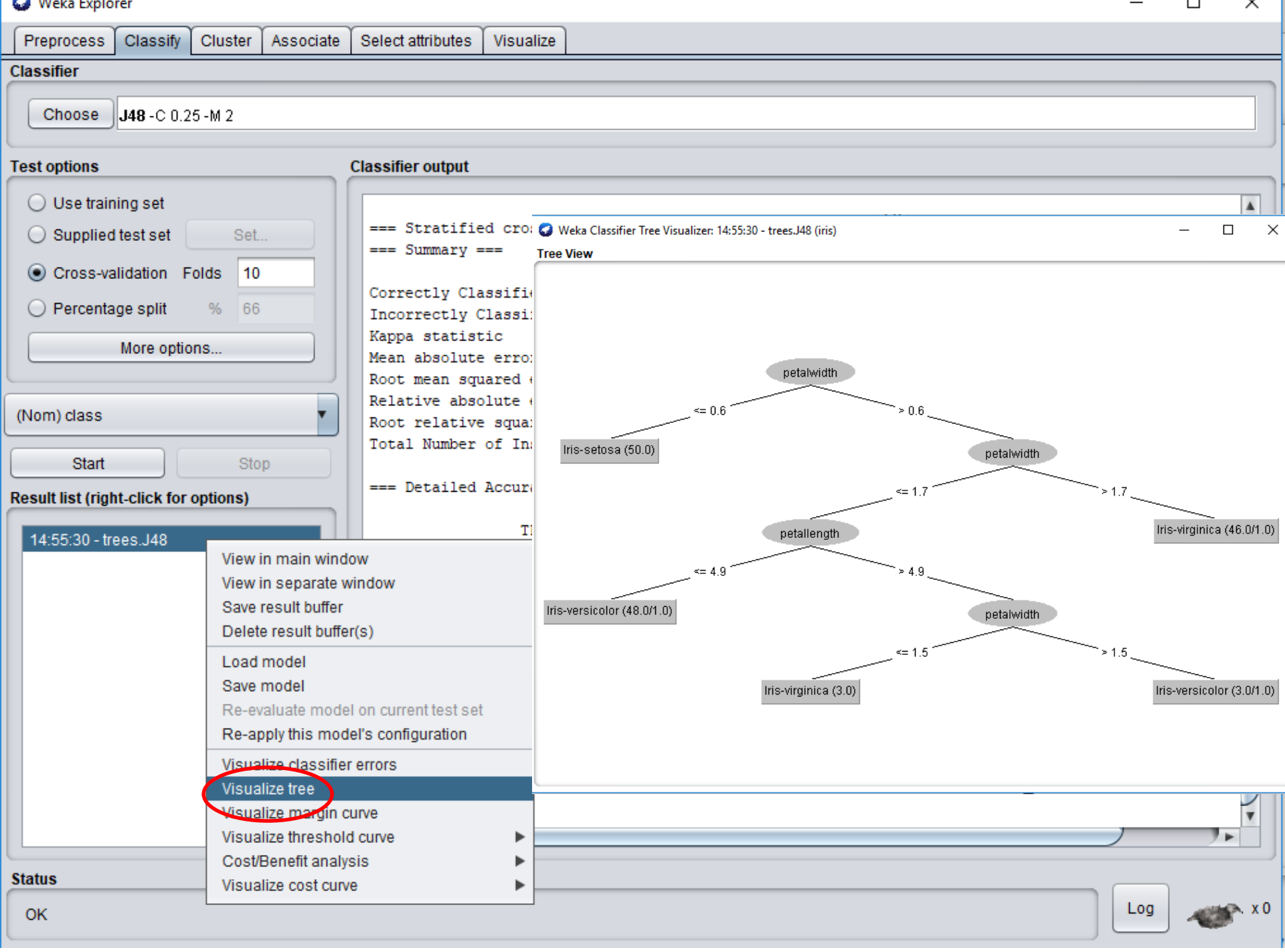
```
=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances         144               96      %
Incorrectly Classified Instances         6                4      %
Kappa statistic                          0.94
Mean absolute error                      0.035
Root mean squared error                  0.1586
Relative absolute error                  7.8705 %
Root relative squared error             33.6353 %
Total Number of Instances              150

=== Detailed Accuracy By Class ===

               TP Rate   FP Rate   Precision   Recall  F-Measure  MCC      ROC Area  PRC Area  Cla
                 0.980    0.000    1.000      0.980    0.990      0.985    0.990     0.987     Iri
                 0.940    0.030    0.940      0.940    0.940      0.910    0.952     0.880     Iri
                 0.960    0.030    0.941      0.960    0.950      0.925    0.961     0.905     Iri
Weighted Avg.    0.960    0.020    0.960      0.960    0.960      0.940    0.968     0.924

=== Confusion Matrix ===

  a  b  c   <-- classified as
 49  1  0 |  a = Iris-setosa
  0 47  3 |  b = Iris-versicolor
  0  2 48 |  c = Iris-virginica
```

Status

OK

Log


Weka Explorer
Preprocess
Classify
Cluster
Associ
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) class
Start
Stop
Result list (right-click for options)
14:55:30 - trees.J48
Status
OK
About
Class for generating a pruned or unpruned C4.
More
Capabilities
batchSize
100
binarySplits
False
collapseTree
True
confidenceFactor
0.25
debug
False
doNotCheckCapabilities
False
doNotMakeSplitPointActualValue
False
minNumObj
2
numDecimalPlaces
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
seed
1
subtreeRaising
True
unpruned
False
useLaplace
False
useMDLcorrection
True
ICC
ROC Area
PRC Area
Cla
.985
0.990
0.987
Iri
.910
0.952
0.880
Iri
.925
0.961
0.905
Iri
.940
0.968
0.924
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
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) class
Start
Stop
Result list (right-click for options)
14:55:30 - trees.J48
Classifier output
=== Stratified cro
=== Summary ===
Correctly Classifi
Incorrectly Classi
Kappa statistic
Mean absolute erro
Root mean squared
Relative absolute
Root relative squa
Total Number of In
=== Detailed Accur
View in main window
View in separate window
Save result buffer
Delete result buffer(s)
Load model
Save model
Re-evaluate model on current test set
Re-apply this model's configuration
Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Cost/Benefit analysis
Visualize cost curve
Weka Classifier Tree Visualizer: 14:55:30 - trees.J48 (iris)
Tree View
petalwidth
<= 0.6
> 0.6
Iris-setosa (50.0)
petalwidth
<= 1.7
> 1.7
petallength
Iris-virginica (46.0/1.0)
<= 4.9
> 4.9
Iris-versicolor (48.0/1.0)
petalwidth
<= 1.5
> 1.5
Iris-virginica (3.0)
Iris-versicolor (3.0/1.0)
Status
OK
Log
x 0

Weka Explorer

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

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

Filter

Choose | None | Apply | Stop

Current relation

Relation: weather.symbolic
Instances: 14
Attributes: 5
Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☑	humidity
4	☐	windy
5	☐	play

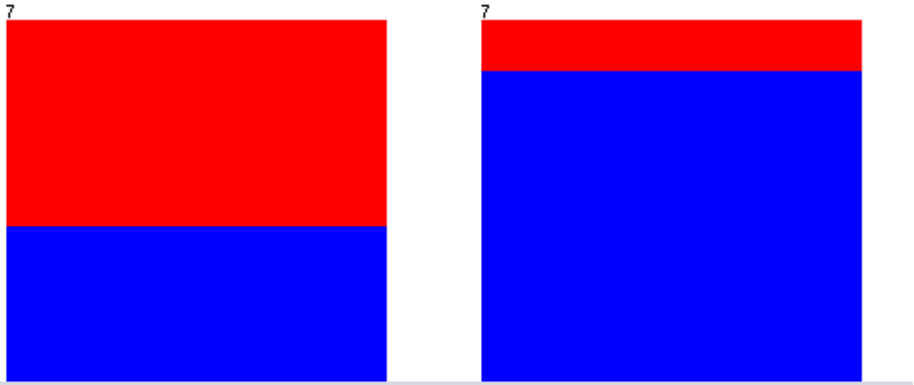
Remove

Selected attribute

Name: humidity
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)

No.	Label	Count	Weight
1	high	7	7.0
2	normal	7	7.0

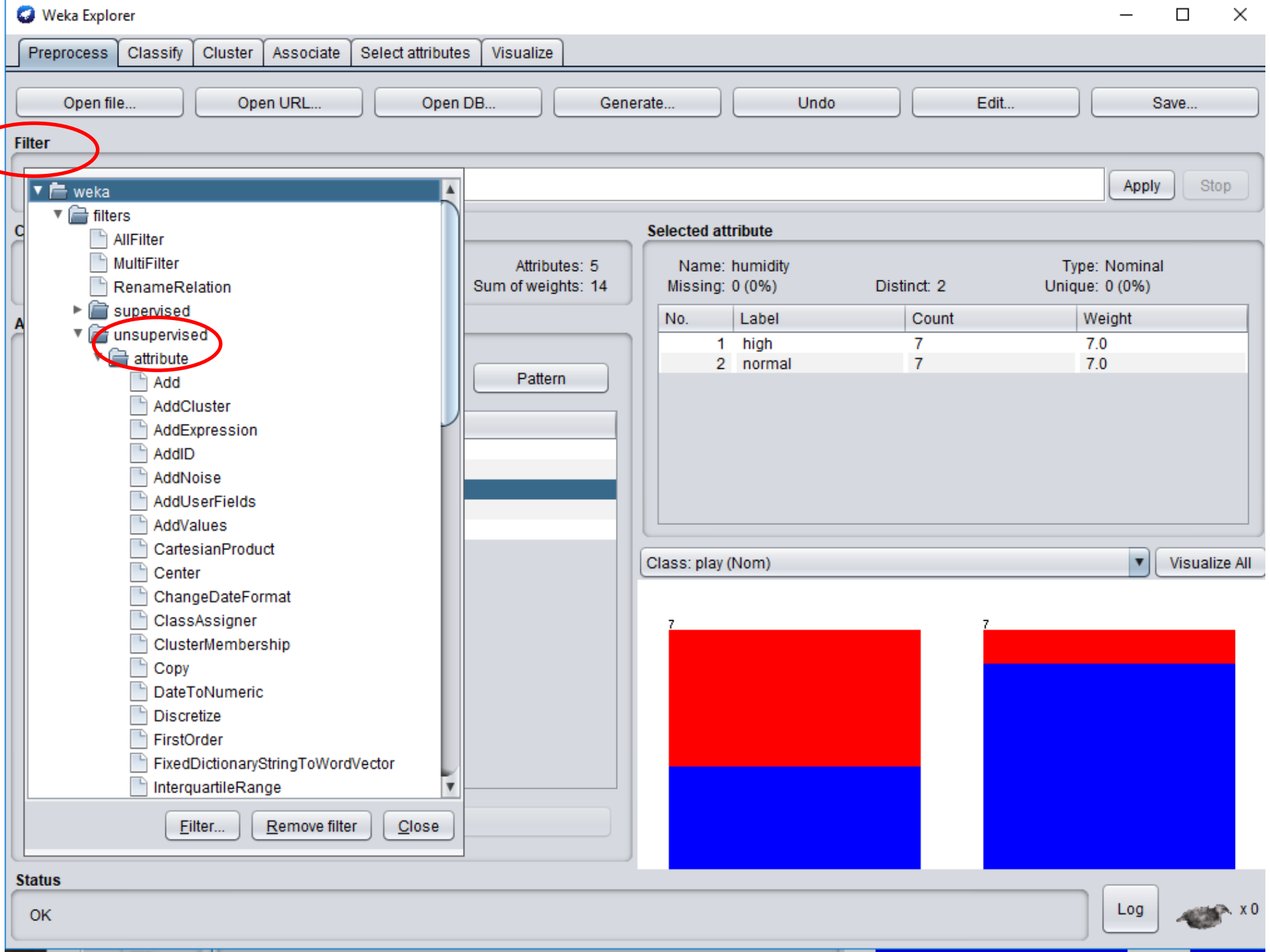
Class: play (Nom) | Visualize All



Status

OK

Log | x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
weka
filters
AllFilter
MultiFilter
RenameRelation
supervised
unsupervised
attribute
Add
AddCluster
AddExpression
AddID
AddNoise
AddUserFields
AddValues
CartesianProduct
Center
ChangeDateFormat
ClassAssigner
ClusterMembership
Copy
DateToNumeric
Discretize
FirstOrder
FixedDictionaryStringToWordVector
InterquartileRange
Filter...
Remove filter
Close
Apply
Stop
Attributes: 5
Sum of weights: 14
Pattern
Selected attribute
Name: humidity
Missing: 0 (0%)
Distinct: 2
Type: Nominal
Unique: 0 (0%)
No.
Label
Count
Weight
1
high
7
7.0
2
normal
7
7.0
Class: play (Nom)
Visualize All
7
7
Status
OK
Log
x 0

Weka Explorer

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

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

Filter

Choose **Remove** | Apply | Stop

Current relation

Relation: weather.symbolic-weka.filters.unsupervised.att... | Attributes: 5
Instances: 14 | Sum of weights: 14

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	outlook
2	☐	temperature
3	☐	humidity
4	☐	windy
5	☐	play

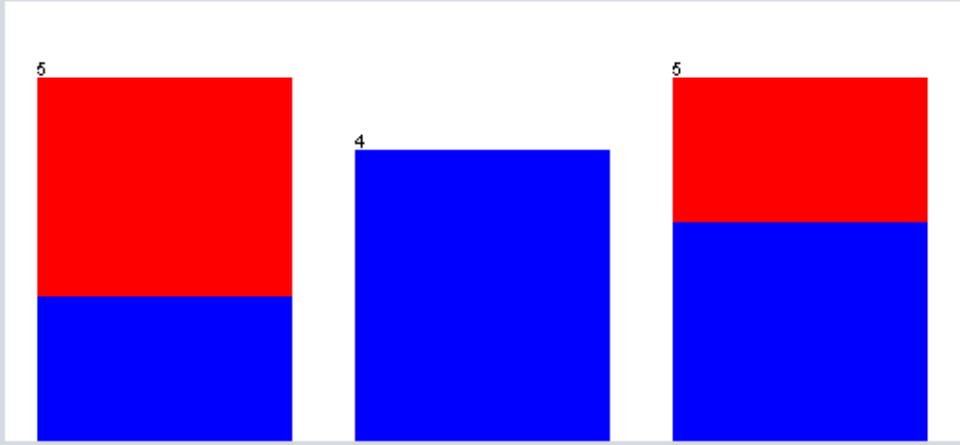
Remove

Selected attribute

Name: outlook | Type: Nominal
Missing: 0 (0%) | Distinct: 3 | Unique: 0 (0%)

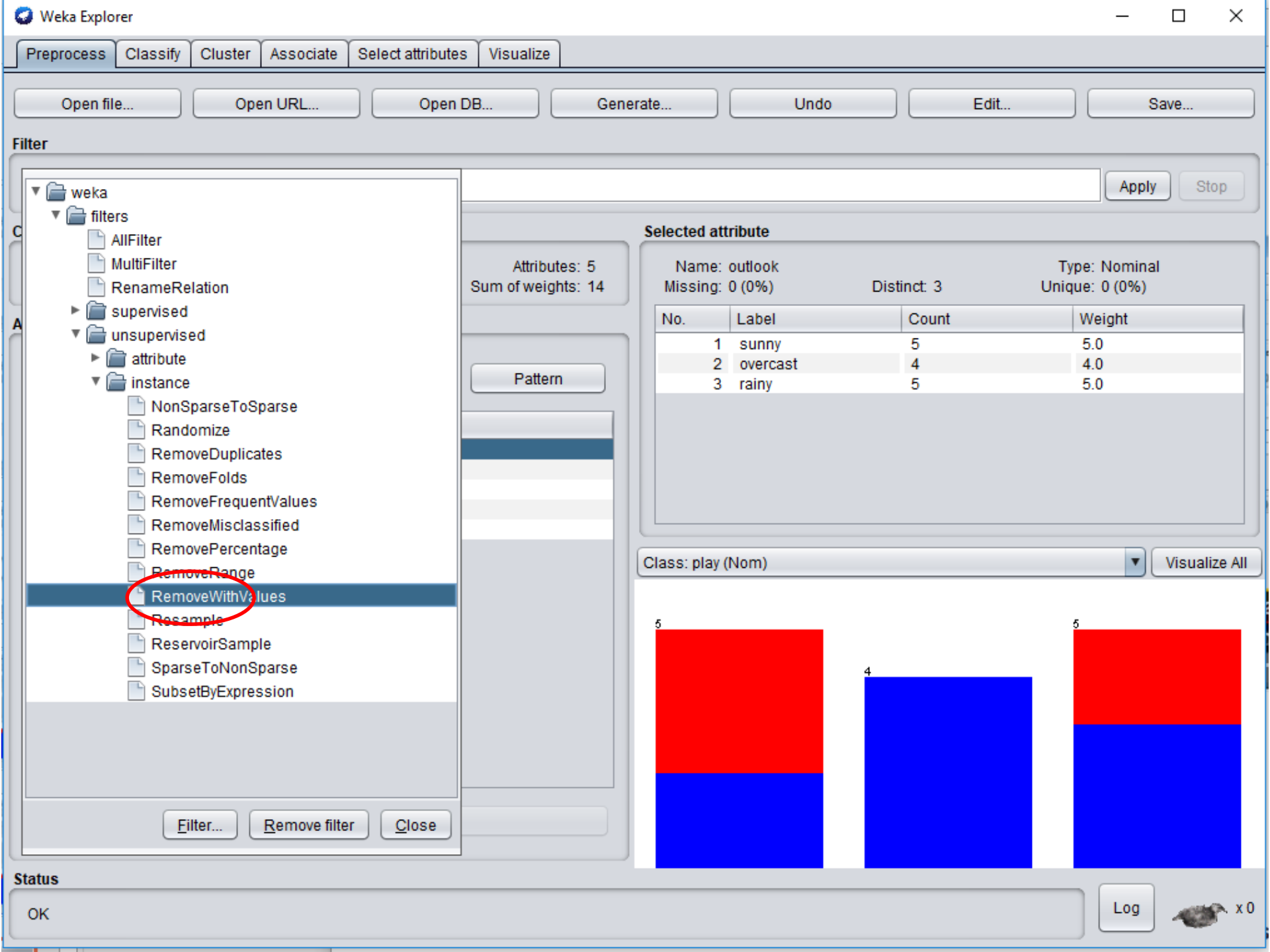
No.	Label	Count	Weight
1	sunny	5	5.0
2	overcast	4	4.0
3	rainy	5	5.0

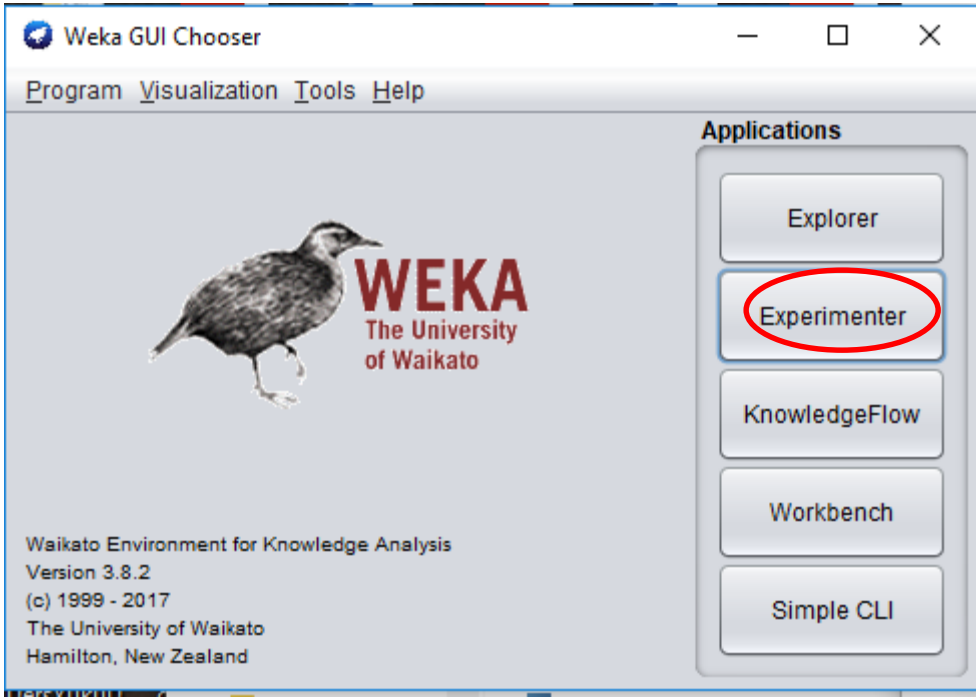
Class: play (Nom) | Visualize All

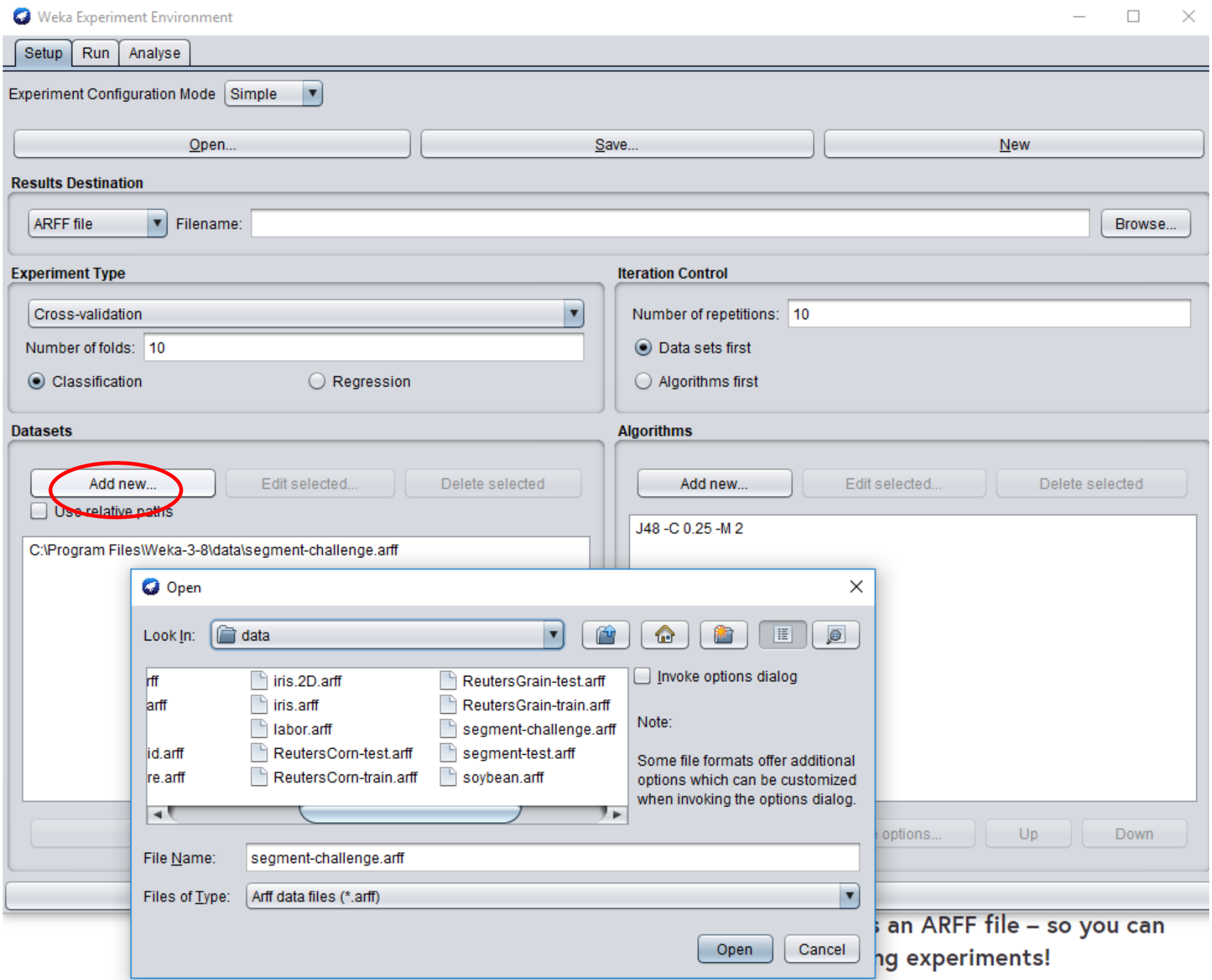


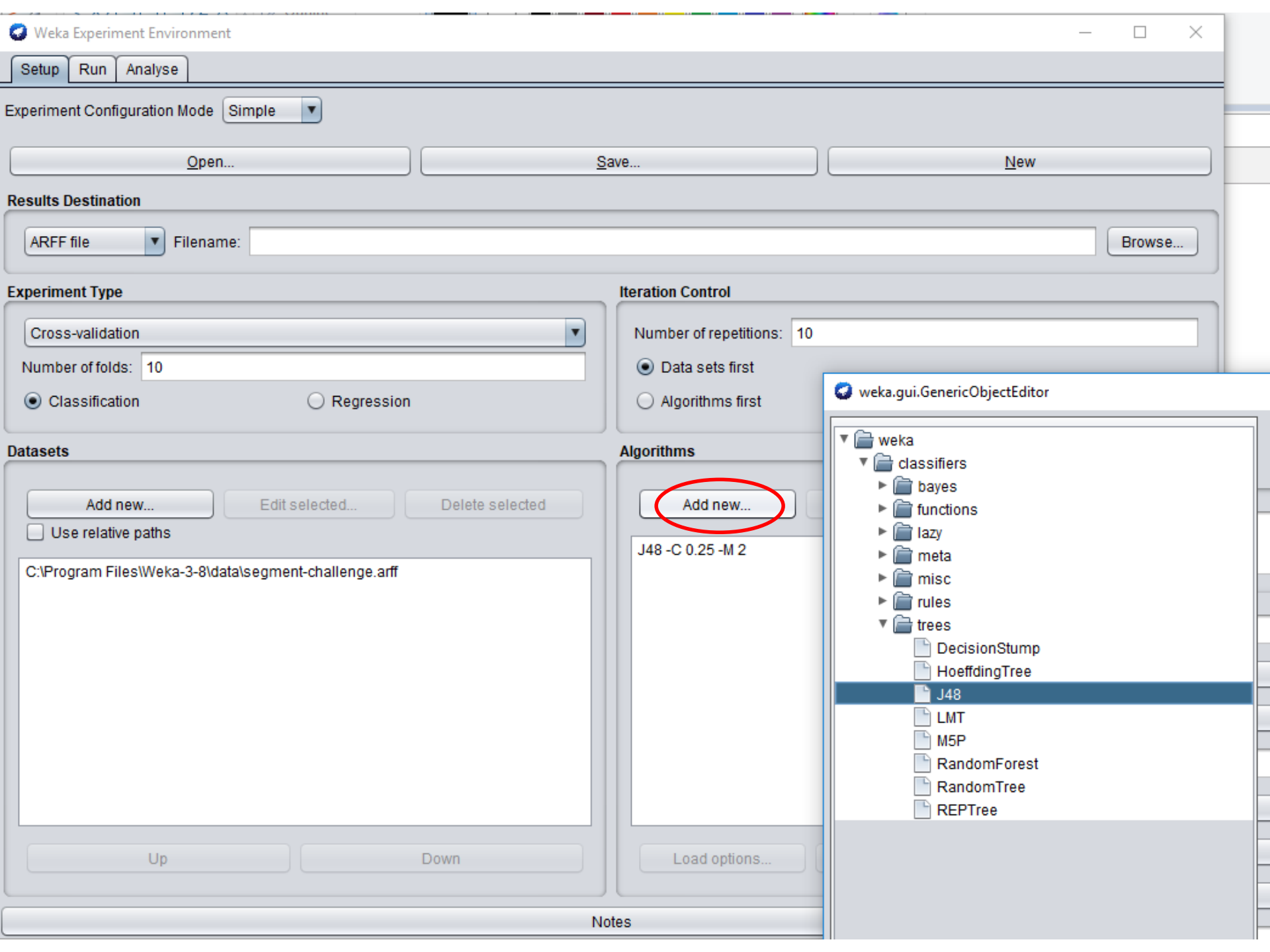
Status

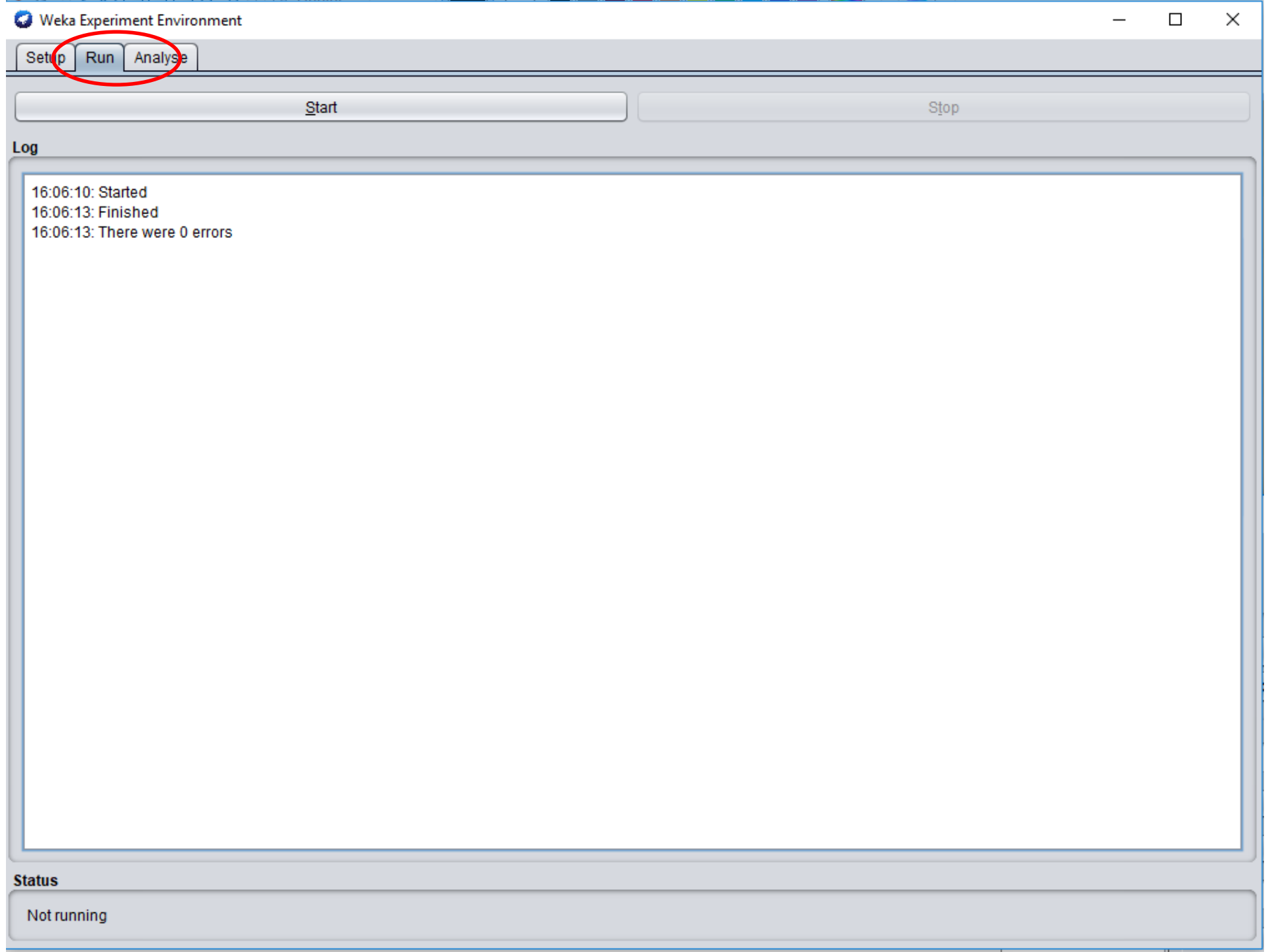
OK | Log | x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Apply
Stop
weka
filters
AllFilter
MultiFilter
RenameRelation
supervised
unsupervised
attribute
instance
NonSparseToSparse
Randomize
RemoveDuplicates
RemoveFolds
RemoveFrequentValues
RemoveMisclassified
RemovePercentage
RemoveRange
RemoveWithValues
Resample
ReservoirSample
SparseToNonSparse
SubsetByExpression
Filter...
Remove filter
Close
Attributes: 5
Sum of weights: 14
Pattern
Selected attribute
Name: outlook
Type: Nominal
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)
No.
Label
Count
Weight
1
sunny
5
5.0
2
overcast
4
4.0
3
rainy
5
5.0
Class: play (Nom)
Visualize All
5
4
5
Status
OK
Log
x 0


Weka GUI Chooser
Program Visualization Tools Help
Applications
Explorer
Experimenter
KnowledgeFlow
Workbench
Simple CLI
WEKA
The University of Waikato
Waikato Environment for Knowledge Analysis
Version 3.8.2
(c) 1999 - 2017
The University of Waikato
Hamilton, New Zealand


Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode
Simple
Open...
Save...
New
Results Destination
ARFF file
Filename:
Browse...
Experiment Type
Cross-validation
Number of folds:
10
Classification
Regression
Iteration Control
Number of repetitions:
10
Data sets first
Algorithms first
Datasets
Add new...
Edit selected...
Delete selected
Use relative paths
C:\Program Files\Weka-3-8\data\segment-challenge.arff
Algorithms
Add new...
Edit selected...
Delete selected
J48 -C 0.25 -M 2
Open
Look In:
data
rff
arff
id.arff
re.arff
iris.2D.arff
iris.arff
labor.arff
ReutersCorn-test.arff
ReutersCorn-train.arff
ReutersGrain-test.arff
ReutersGrain-train.arff
segment-challenge.arff
segment-test.arff
soybean.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
File Name:
segment-challenge.arff
Files of Type:
Arff data files (*.arff)
Open
Cancel
options...
Up
Down
an ARFF file – so you can
ng experiments!


Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode
Simple
Open...
Save...
New
Results Destination
ARFF file
Filename:
Browse...
Experiment Type
Cross-validation
Number of folds:
10
Classification
Regression
Iteration Control
Number of repetitions:
10
Data sets first
Algorithms first
Datasets
Add new...
Edit selected...
Delete selected
Use relative paths
C:\Program Files\Weka-3-8\data\segment-challenge.arff
Up
Down
Algorithms
Add new...
J48 -C 0.25 -M 2
Load options...
Notes
weka.gui.GenericObjectEditor
weka
classifiers
bayes
functions
lazy
meta
misc
rules
trees
DecisionStump
HoeffdingTree
J48
LMT
M5P
RandomForest
RandomTree
REPTree


Weka Experiment Environment
Setup
Run
Analyse
Start
Stop
Log
16:06:10: Started
16:06:13: Finished
16:06:13: There were 0 errors
Status
Not running

Weka Experiment Environment

Setup | Run | Analyse

Source

Got 200 results

File... | Database... | Experiment

Actions

Perform test | Save output | Open Explorer...

Configure test

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows | Cols | Swap

Comparison field: Percent_correct

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: ☐

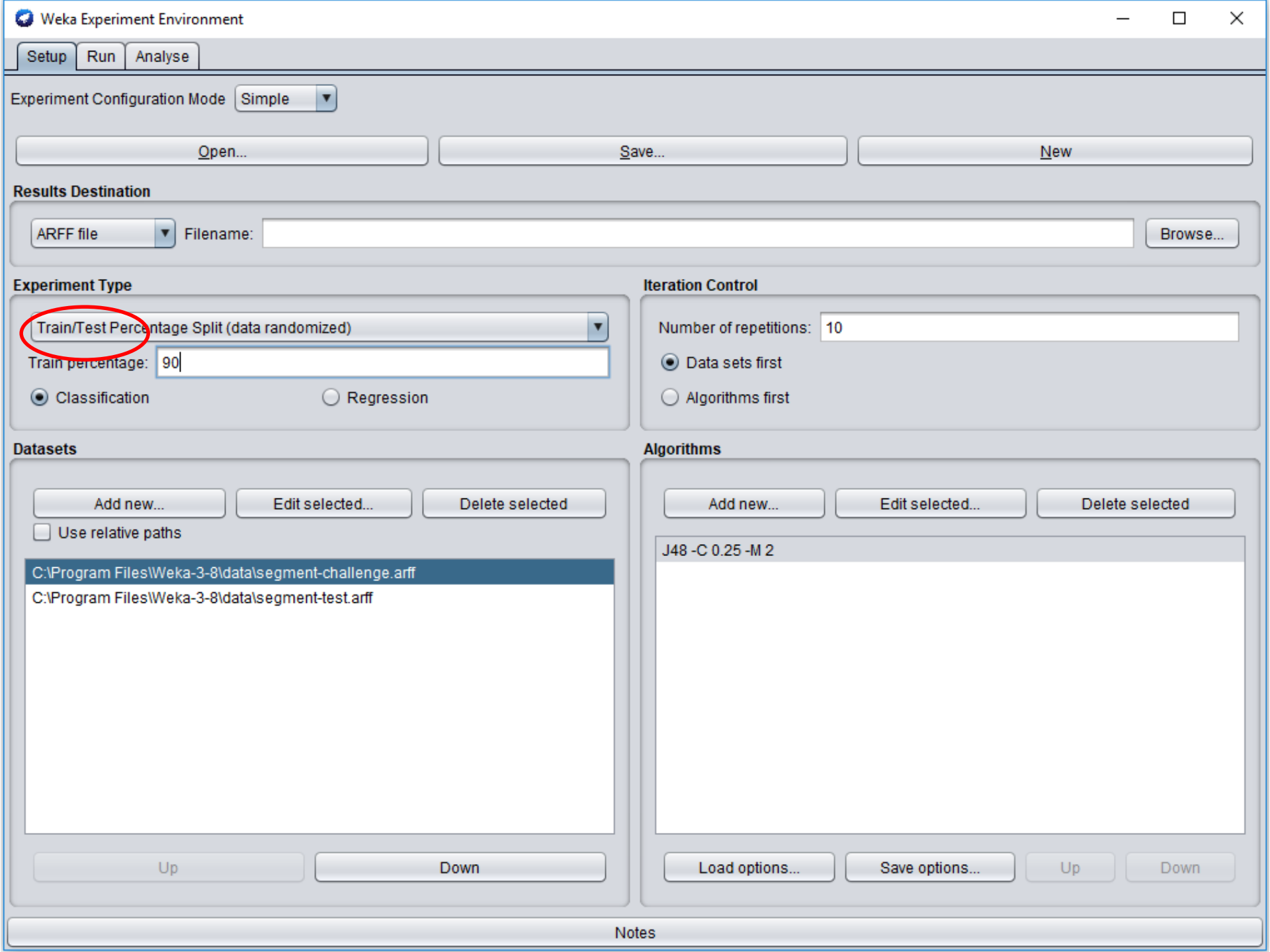
Output Format: Select

Result list

16:09:41 - Available resultsets
16:09:47 - region-centroid-col -
16:17:52 - region-centroid-col -
16:18:14 - region-centroid-col -
16:18:25 - Available resultsets
16:18:28 - Percent_correct - trees.J48 '-C 0.25 -M 2' -21773

Test output

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 0.05 -resul
Analysing:  Percent_correct
Datasets:   1
Resultsets: 1
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       5/24/18 4:18 PM


Dataset                   (1) trees.J48
----------------------------------------
segment                  (200)   94.72 |
----------------------------------------
(v/ /*)                                |


Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644444
```


Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode
Simple
Open...
Save...
New
Results Destination
ARFF file
Filename:
Browse...
Experiment Type
Train/Test Percentage Split (data randomized)
Train percentage:
90
Classification
Regression
Iteration Control
Number of repetitions:
10
Data sets first
Algorithms first
Datasets
Add new...
Edit selected...
Delete selected
Use relative paths
C:\Program Files\Weka-3-8\data\segment-challenge.arff
C:\Program Files\Weka-3-8\data\segment-test.arff
Up
Down
Algorithms
Add new...
Edit selected...
Delete selected
J48 -C 0.25 -M 2
Load options...
Save options...
Up
Down
Notes

Clustering

Weka Explorer

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

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

Filter

Choose | None | Apply | Stop

Current relation

Relation: iris
Instances: 150
Attributes: 5
Sum of weights: 150

Attributes

All | None | Invert | Pattern

No.	Name
1	sepallength
2	sepalwidth
3	petallength
4	petalwidth
5	class

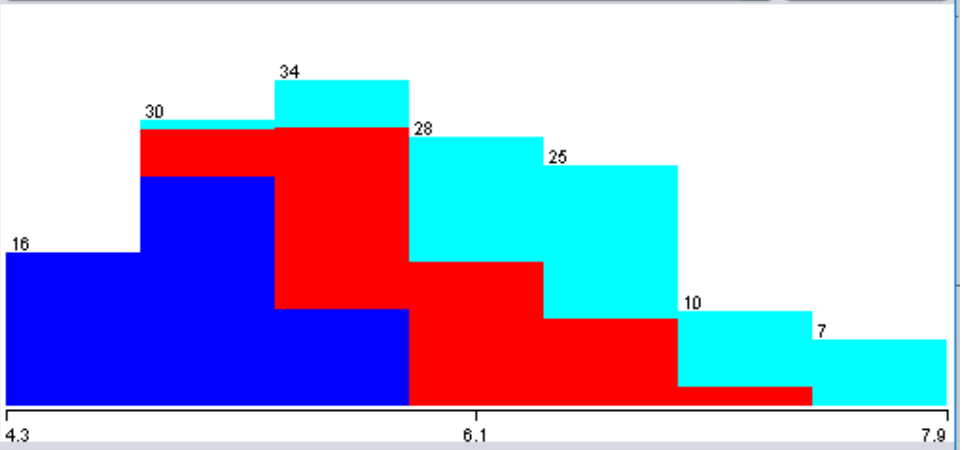
Remove

Selected attribute

Name: sepallength
Missing: 0 (0%)
Distinct: 35
Type: Numeric
Unique: 9 (6%)

Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

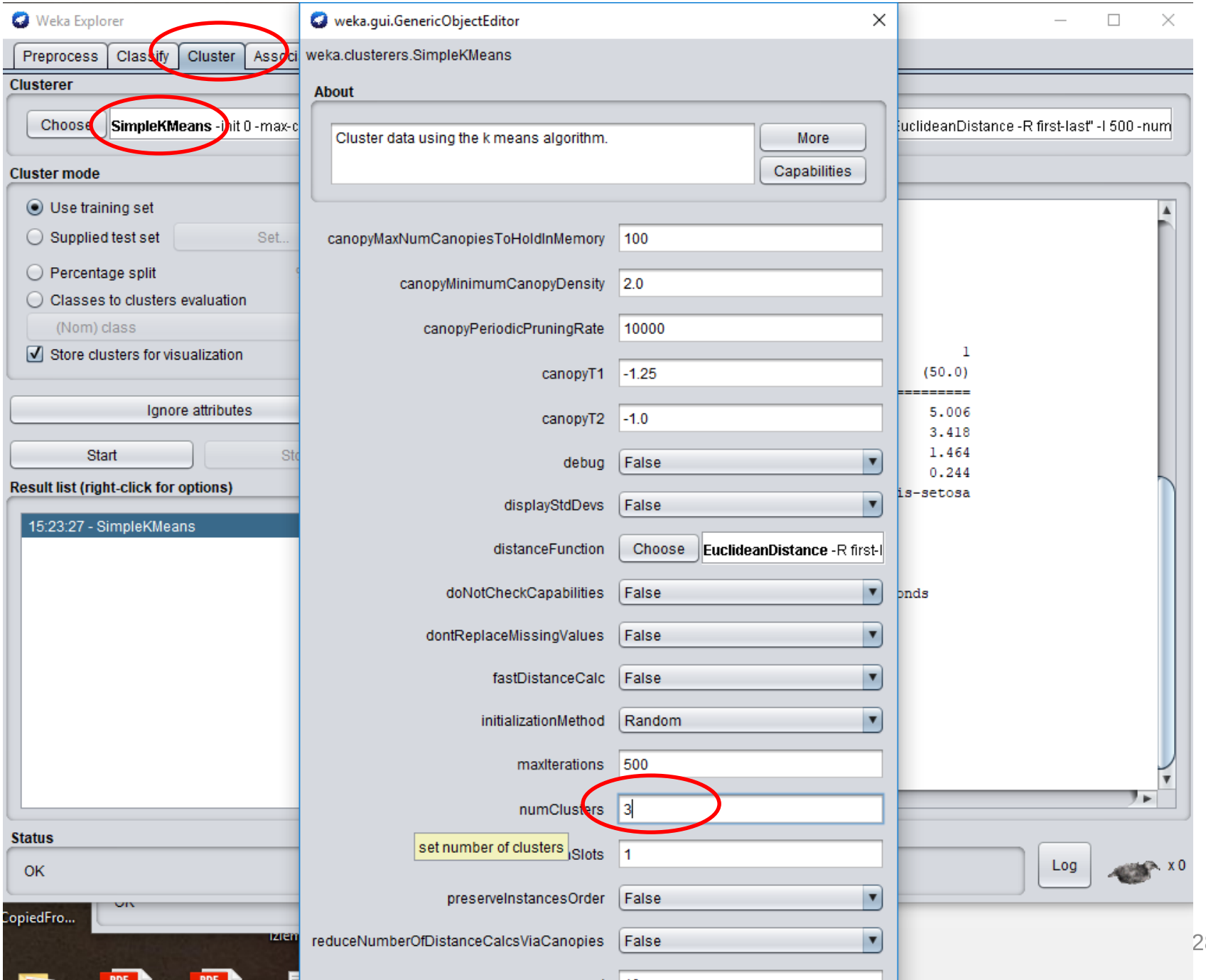
Class: class (Nom) | Visualize All



Status

OK

Log | x 0


Weka Explorer
Preprocess
Classify
Cluster
Clusterer
Choose
SimpleKMeans -init 0 -max-c
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Result list (right-click for options)
15:23:27 - SimpleKMeans
Status
OK
weka.gui.GenericObjectEditor
weka.clusterers.SimpleKMeans
About
Cluster data using the k means algorithm.
More
Capabilities
canopyMaxNumCanopiesToHoldInMemory 100
canopyMinimumCanopyDensity 2.0
canopyPeriodicPruningRate 10000
canopyT1 -1.25
canopyT2 -1.0
debug False
displayStdDevs False
distanceFunction Choose EuclideanDistance -R first-l
doNotCheckCapabilities False
dontReplaceMissingValues False
fastDistanceCalc False
initializationMethod Random
maxIterations 500
numClusters 3
set number of clusters
Slots 1
preserveInstancesOrder False
reduceNumberOfDistanceCalcsViaCanopies False
uclideanDistance -R first-last" -I 500 -num
1
(50.0)
5.006
3.418
1.464
0.244
is-setosa
onds
Log
x 0

Weka Explorer

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 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num

Cluster mode

- (•) Use training set
- ( ) Supplied test set — Set...
- ( ) Percentage split — % 66
- ( ) Classes to clusters evaluation

(Nom) class

[✓] Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans

Clusterer output

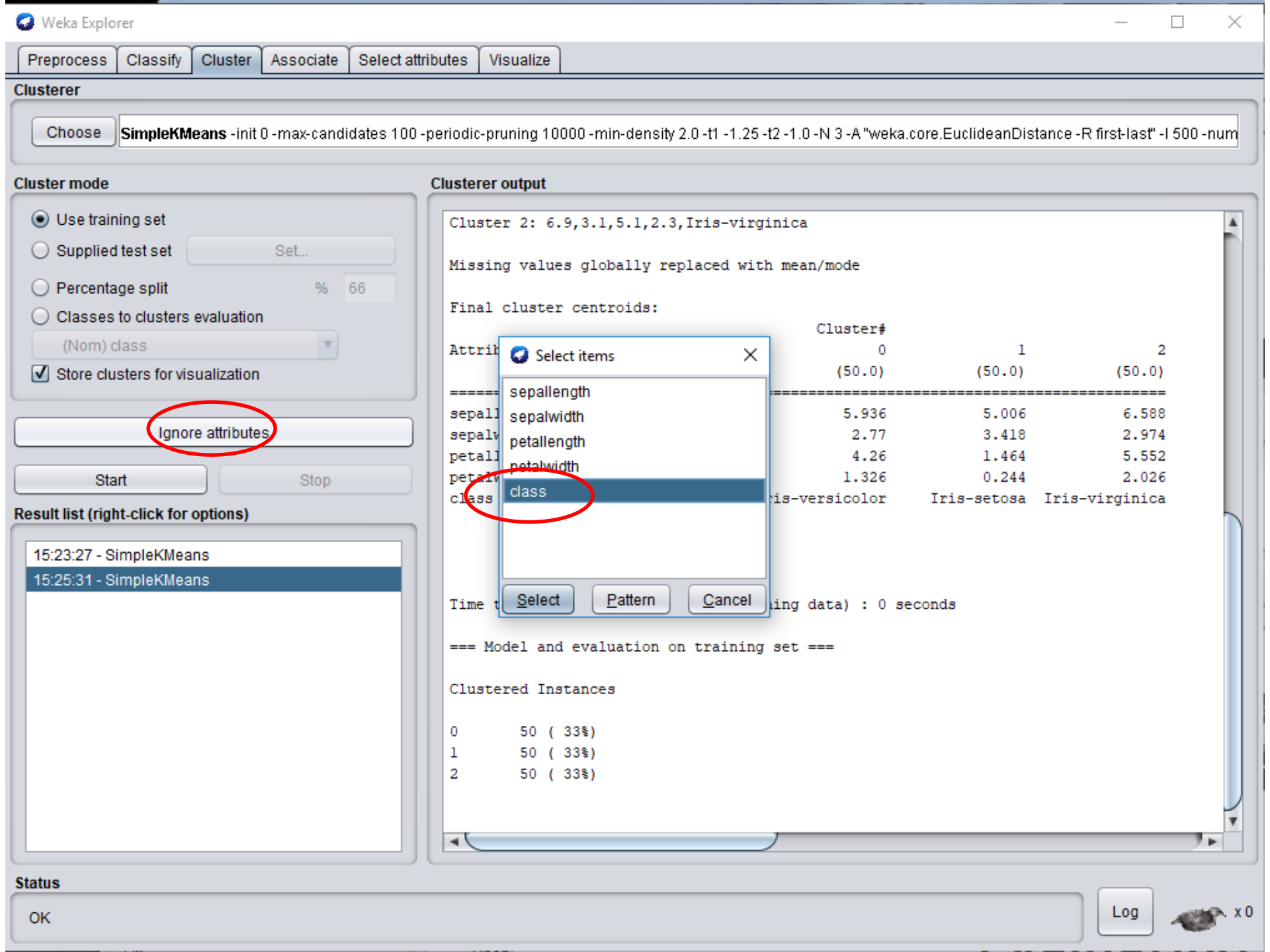
```
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica

Missing values globally replaced with mean/mode

Final cluster centroids:
                                        Cluster#
Attribute               Full Data             0             1             2
                          (150.0)       (50.0)        (50.0)        (50.0)
================================================================================
sepallength                5.8433         5.936         5.006         6.588
sepalwidth                  3.054          2.77         3.418         2.974
petallength                3.7587          4.26         1.464         5.552
petalwidth                 1.1987         1.326         0.244         2.026
class                 Iris-setosa Iris-versicolor     Iris-setosa  Iris-virginica



Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

0       50 ( 33%)
1       50 ( 33%)
2       50 ( 33%)
```

Status

OK

Log

x 0


Weka Explorer
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 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans
Clusterer output
Cluster 2: 6.9,3.1,5.1,2.3,Iris-virginica
Missing values globally replaced with mean/mode
Final cluster centroids:
Cluster#
Attrib
0
1
2
(50.0)
(50.0)
(50.0)
sepall
sepalw
petall
petalw
class
5.936
5.006
6.588
2.77
3.418
2.974
4.26
1.464
5.552
1.326
0.244
2.026
is-versicolor
Iris-setosa
Iris-virginica
Select items
sepallength
sepalwidth
petallength
petalwidth
class
Select
Pattern
Cancel
Time t
ing data) : 0 seconds
=== Model and evaluation on training set ===
Clustered Instances
0 50 (33%)
1 50 (33%)
2 50 (33%)
Status
OK
Log
x 0

Weka Explorer

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 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num

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)

15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans
15:31:48 - SimpleKMeans

Clusterer output

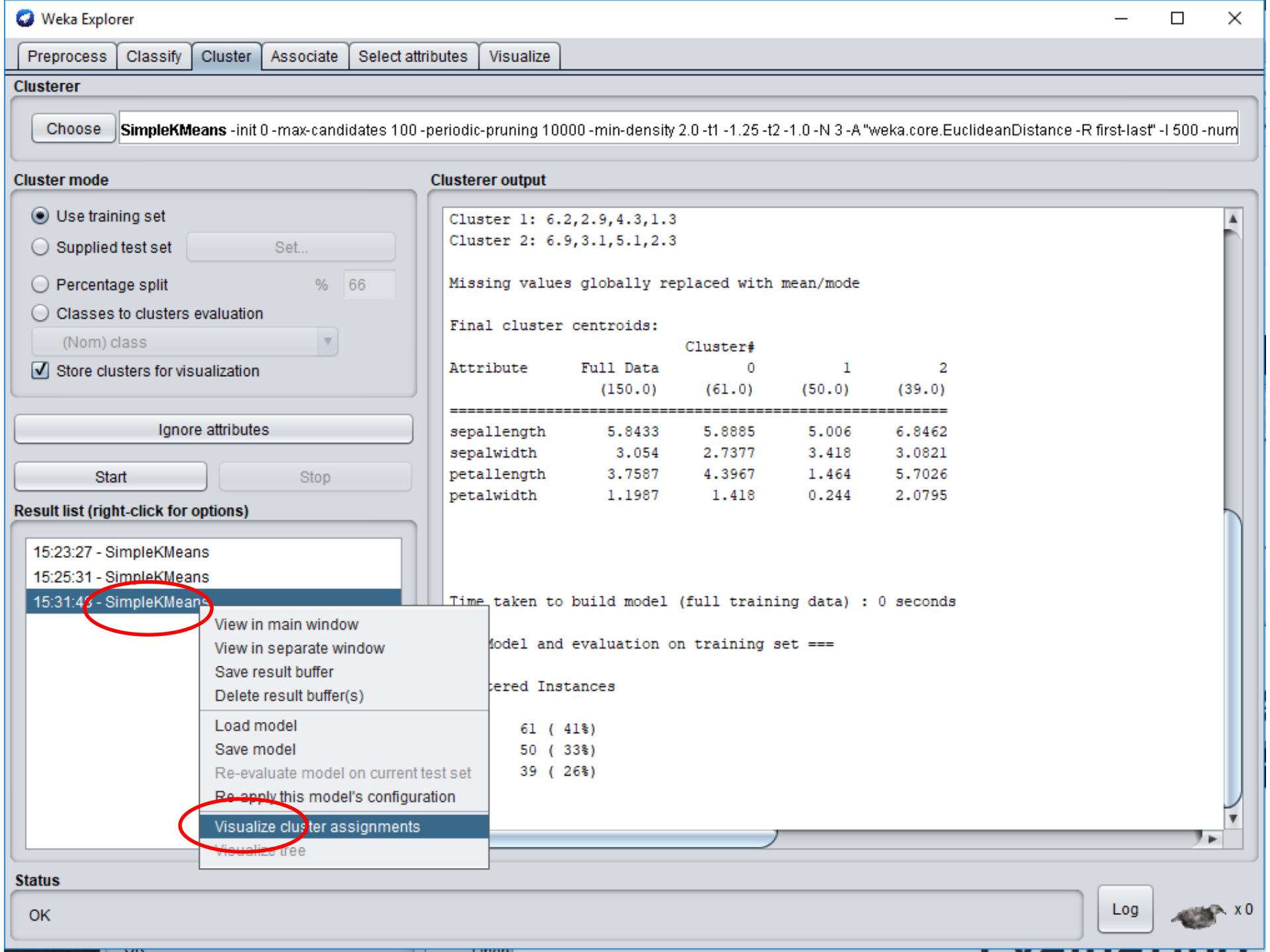
```
Cluster 1: 6.2,2.9,4.3,1.3
Cluster 2: 6.9,3.1,5.1,2.3

Missing values globally replaced with mean/mode

Final cluster centroids:
                           Cluster#
Attribute      Full Data          0          1          2
                 (150.0)     (61.0)     (50.0)     (39.0)
=============================================================
sepallength       5.8433     5.8885      5.006     6.8462
sepalwidth         3.054     2.7377      3.418     3.0821
petallength       3.7587     4.3967      1.464     5.7026
petalwidth        1.1987      1.418      0.244     2.0795



Time taken to build model (full training data) : 0 seconds

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

Clustered Instances

0       61 ( 41%)
1       50 ( 33%)
2       39 ( 26%)
```

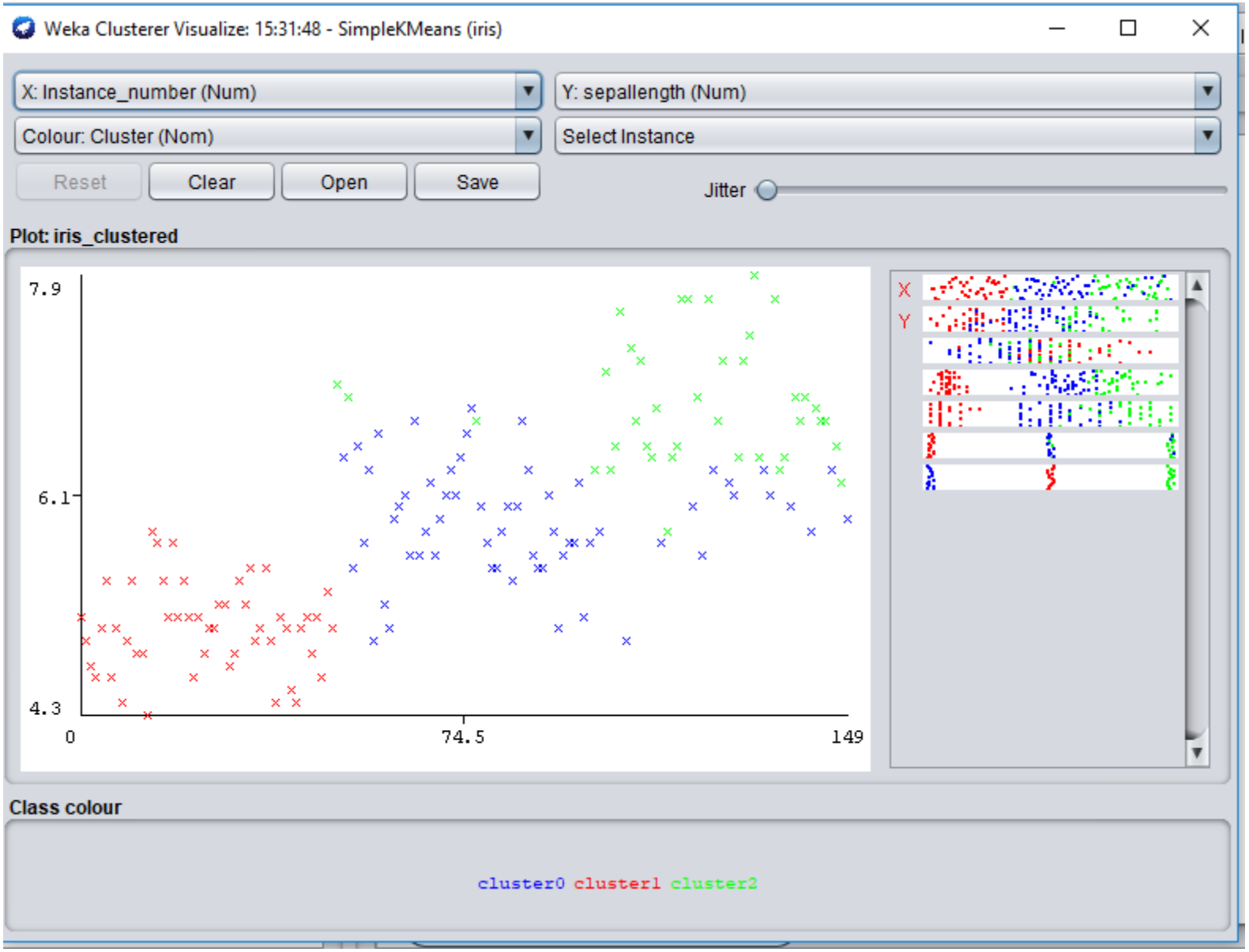
Status

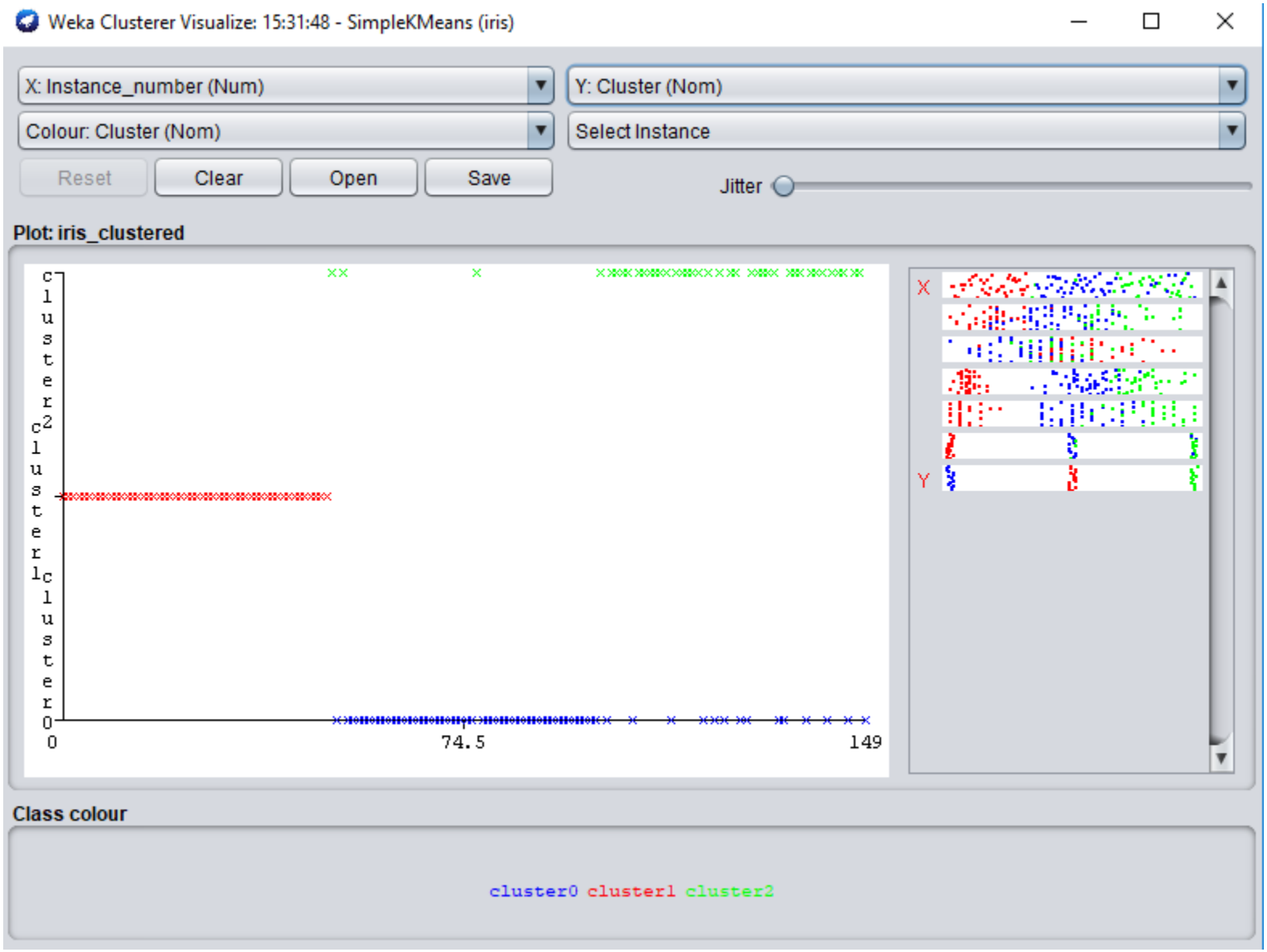
OK

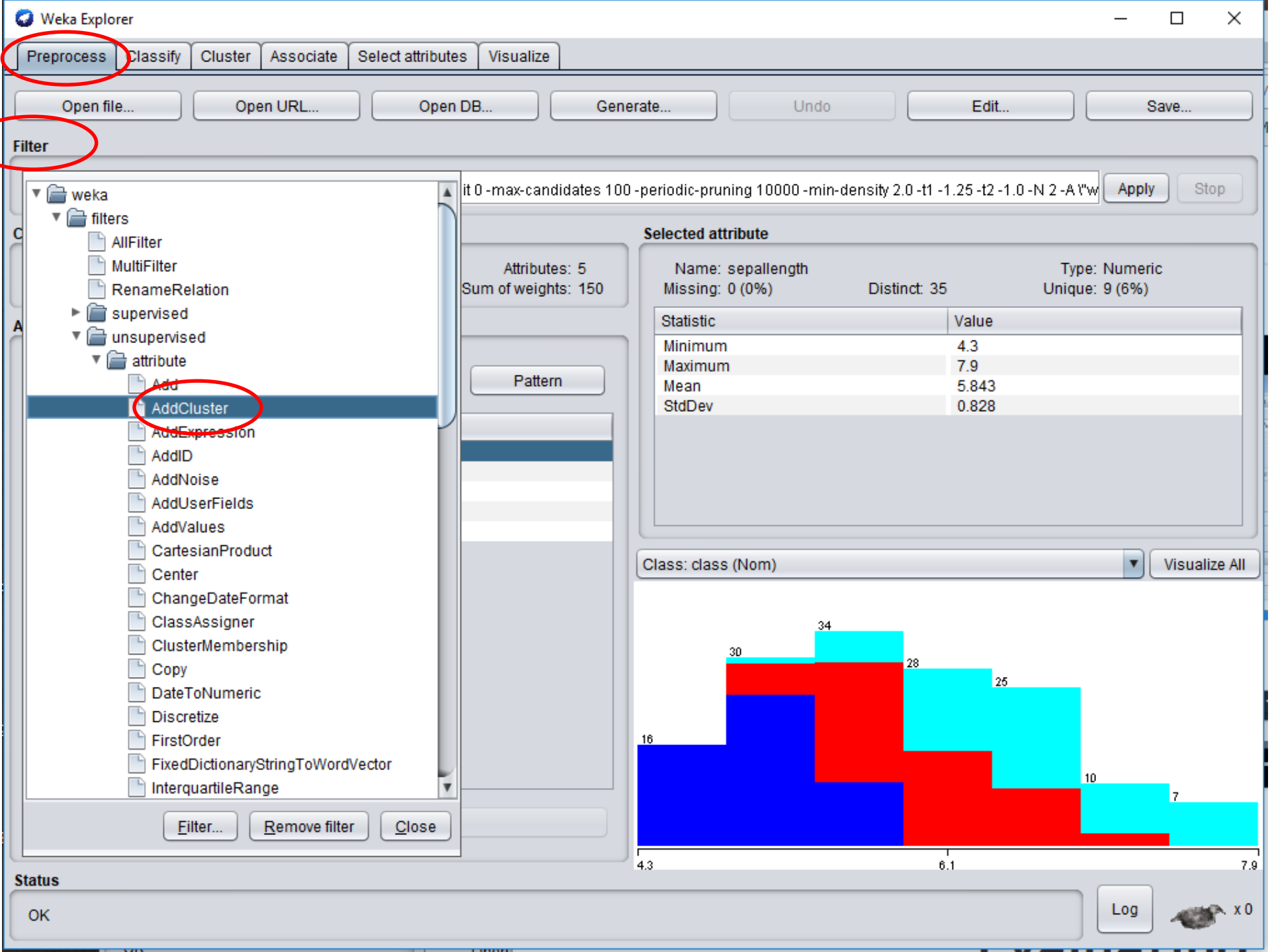
Log

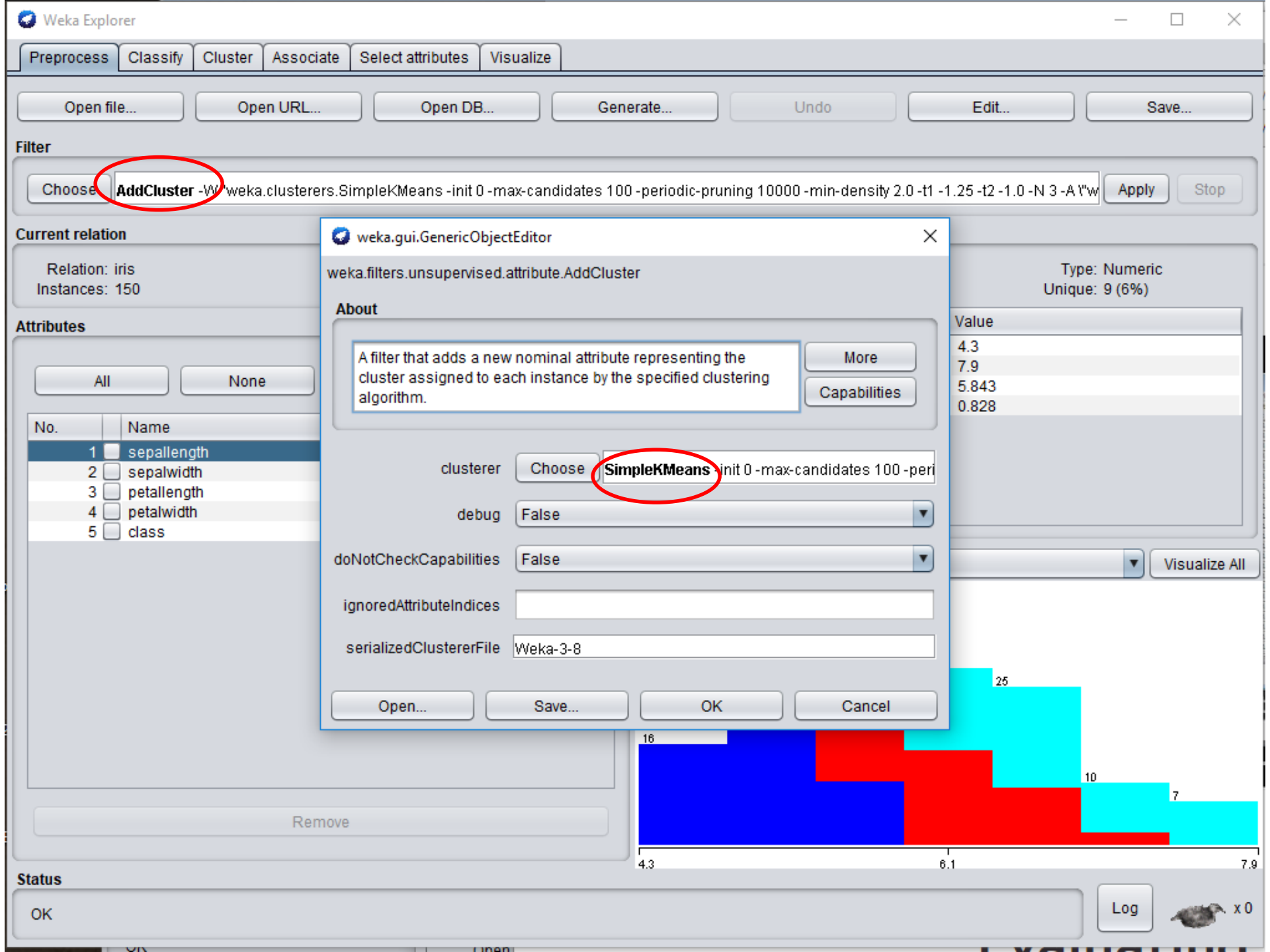
x 0

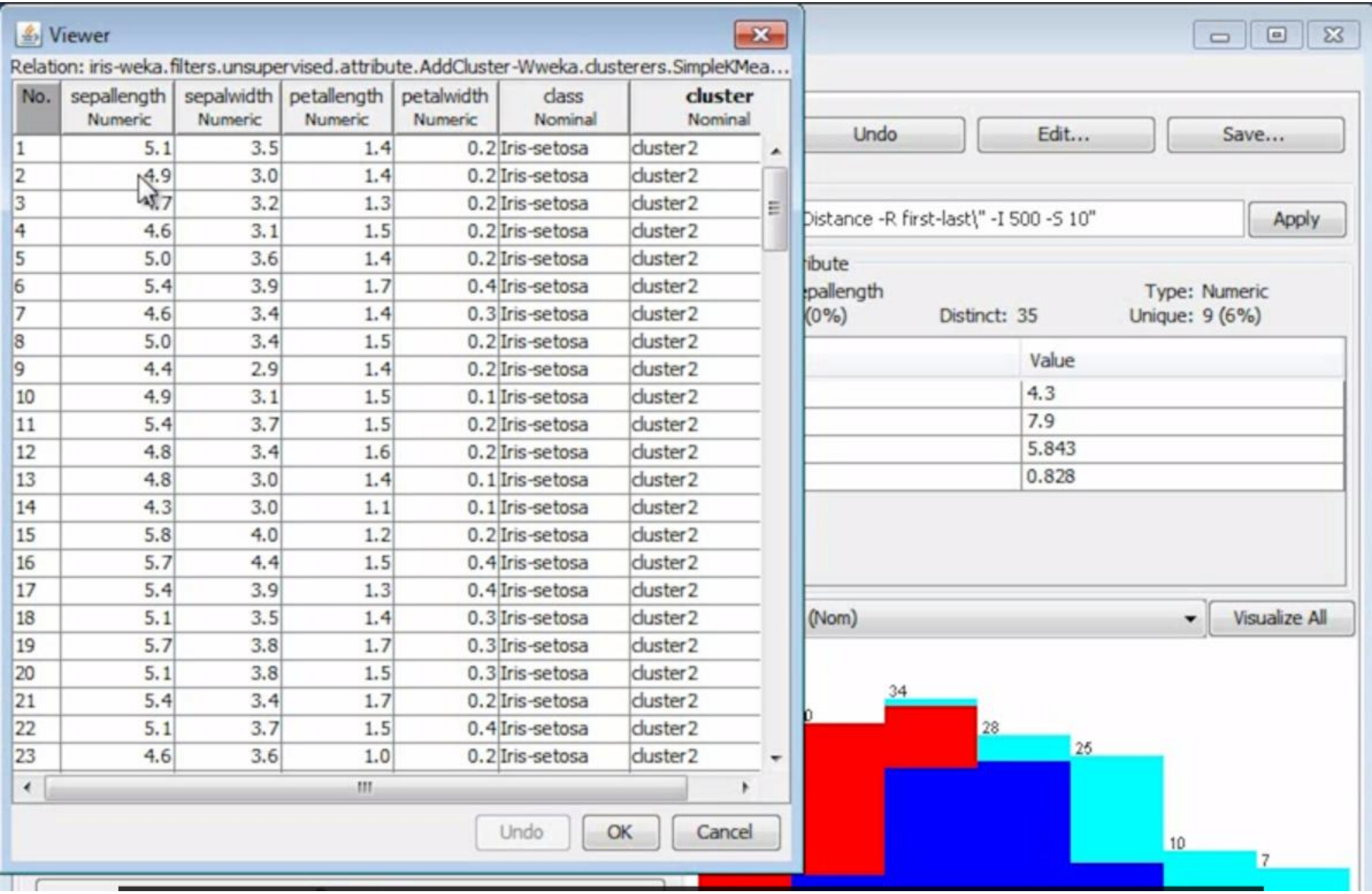

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans
15:31:43 - SimpleKMeans
View in main window
View in separate window
Save result buffer
Delete result buffer(s)
Load model
Save model
Re-evaluate model on current test set
Re-apply this model's configuration
Visualize cluster assignments
Visualize tree
Clusterer output
Cluster 1: 6.2,2.9,4.3,1.3
Cluster 2: 6.9,3.1,5.1,2.3
Missing values globally replaced with mean/mode
Final cluster centroids:
Cluster#
Attribute Full Data 0 1 2
(150.0) (61.0) (50.0) (39.0)
sepallength 5.8433 5.8885 5.006 6.8462
sepalwidth 3.054 2.7377 3.418 3.0821
petallength 3.7587 4.3967 1.464 5.7026
petalwidth 1.1987 1.418 0.244 2.0795
Time taken to build model (full training data) : 0 seconds
odel and evaluation on training set ===
tered Instances
61 (41%)
50 (33%)
39 (26%)
Status
OK
Log
x 0


Weka Clusterer Visualize: 15:31:48 - SimpleKMeans (iris)
X: Instance_number (Num)
Y: sepallength (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_clustered
7.9
6.1
4.3
0
74.5
149
X
Y
Class colour
cluster0 cluster1 cluster2


Weka Clusterer Visualize: 15:31:48 - SimpleKMeans (iris)
X: Instance_number (Num)
Y: Cluster (Nom)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_clustered
cluster2
cluster1
cluster0
0
74.5
149
X
Y
Class colour
cluster0 cluster1 cluster2


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
weka
filters
AllFilter
MultiFilter
RenameRelation
supervised
unsupervised
attribute
Add
AddCluster
AddExpression
AddID
AddNoise
AddUserFields
AddValues
CartesianProduct
Center
ChangeDateFormat
ClassAssigner
ClusterMembership
Copy
DateToNumeric
Discretize
FirstOrder
FixedDictionaryStringToWordVector
InterquartileRange
Filter...
Remove filter
Close
it 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A \"w
Apply
Stop
Attributes: 5
Sum of weights: 150
Pattern
Selected attribute
Name: sepallength
Type: Numeric
Missing: 0 (0%)
Distinct: 35
Unique: 9 (6%)
Statistic
Value
Minimum
4.3
Maximum
7.9
Mean
5.843
StdDev
0.828
Class: class (Nom)
Visualize All
34
30
28
25
16
10
7
4.3
6.1
7.9
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
AddCluster -W "weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 3 -A \"w
Apply
Stop
Current relation
Relation: iris
Instances: 150
Attributes
All
None
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Type: Numeric
Unique: 9 (6%)
Value
4.3
7.9
5.843
0.828
Visualize All
16
25
10
7
4.3
6.1
7.9
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.AddCluster
About
A filter that adds a new nominal attribute representing the cluster assigned to each instance by the specified clustering algorithm.
More
Capabilities
clusterer
Choose
SimpleKMeans -init 0 -max-candidates 100 -peri
debug
False
doNotCheckCapabilities
False
ignoredAttributeIndices
serializedClustererFile
Weka-3-8
Open...
Save...
OK
Cancel


Viewer
Relation: iris-weka.filters.unsupervised.attribute.AddCluster-Wweka.clusterers.SimpleKMea...
No. | sepallength Numeric | sepalwidth Numeric | petallength Numeric | petalwidth Numeric | class Nominal | cluster Nominal
1 | 5.1 | 3.5 | 1.4 | 0.2 | Iris-setosa | cluster2
2 | 4.9 | 3.0 | 1.4 | 0.2 | Iris-setosa | cluster2
3 | 4.7 | 3.2 | 1.3 | 0.2 | Iris-setosa | cluster2
4 | 4.6 | 3.1 | 1.5 | 0.2 | Iris-setosa | cluster2
5 | 5.0 | 3.6 | 1.4 | 0.2 | Iris-setosa | cluster2
6 | 5.4 | 3.9 | 1.7 | 0.4 | Iris-setosa | cluster2
7 | 4.6 | 3.4 | 1.4 | 0.3 | Iris-setosa | cluster2
8 | 5.0 | 3.4 | 1.5 | 0.2 | Iris-setosa | cluster2
9 | 4.4 | 2.9 | 1.4 | 0.2 | Iris-setosa | cluster2
10 | 4.9 | 3.1 | 1.5 | 0.1 | Iris-setosa | cluster2
11 | 5.4 | 3.7 | 1.5 | 0.2 | Iris-setosa | cluster2
12 | 4.8 | 3.4 | 1.6 | 0.2 | Iris-setosa | cluster2
13 | 4.8 | 3.0 | 1.4 | 0.1 | Iris-setosa | cluster2
14 | 4.3 | 3.0 | 1.1 | 0.1 | Iris-setosa | cluster2
15 | 5.8 | 4.0 | 1.2 | 0.2 | Iris-setosa | cluster2
16 | 5.7 | 4.4 | 1.5 | 0.4 | Iris-setosa | cluster2
17 | 5.4 | 3.9 | 1.3 | 0.4 | Iris-setosa | cluster2
18 | 5.1 | 3.5 | 1.4 | 0.3 | Iris-setosa | cluster2
19 | 5.7 | 3.8 | 1.7 | 0.3 | Iris-setosa | cluster2
20 | 5.1 | 3.8 | 1.5 | 0.3 | Iris-setosa | cluster2
21 | 5.4 | 3.4 | 1.7 | 0.2 | Iris-setosa | cluster2
22 | 5.1 | 3.7 | 1.5 | 0.4 | Iris-setosa | cluster2
23 | 4.6 | 3.6 | 1.0 | 0.2 | Iris-setosa | cluster2
Undo
OK
Cancel
Undo
Edit...
Save...
Distance -R first-last\" -I 500 -S 10"
Apply
Type: Numeric
Distinct: 35
Unique: 9 (6%)
Value
4.3
7.9
5.843
0.828
(Nom)
Visualize All
34
28
25
10
7

Weka Explorer

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 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num

Cluster mode

- Use training set
- Supplied test set — Set...
- Percentage split — % 66
- Classes to clusters evaluation
  - (Nom) class
- Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans
15:31:48 - SimpleKMeans
15:42:23 - SimpleKMeans

Clusterer output

```
Cluster 1: 6.2,2.9,4.3,1.3
Cluster 2: 6.9,3.1,5.1,2.3

Missing values globally replaced with mean/mode

Final cluster centroids:
                           Cluster#
Attribute      Full Data          0          1          2
                 (150.0)     (61.0)     (50.0)     (39.0)
=========================================================
sepallength       5.8433     5.8885      5.006     6.8462
sepalwidth         3.054     2.7377      3.418     3.0821
petallength       3.7587     4.3967      1.464     5.7026
petalwidth        1.1987      1.418      0.244     2.0795



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

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

Clustered Instances

0       61 ( 41%)
1       50 ( 33%)
2       39 ( 26%)
```

Status

weka.core.converters.CSVLoader failed to load weka_iris_cluster

Log | x 0

Weka Explorer

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 3 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num

Cluster mode

- ○ Use training set
- ○ Supplied test set — Set...
- ○ Percentage split — % 66
- ◉ Classes to clusters evaluation

(Nom) class

☑ Store clusters for visualization

Ignore attributes

Start | Stop

Result list (right-click for options)

15:23:27 - SimpleKMeans
15:25:31 - SimpleKMeans
15:31:48 - SimpleKMeans
15:42:23 - SimpleKMeans
15:49:59 - SimpleKMeans

Clusterer output

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

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

Clustered Instances

0       61 ( 41%)
1       50 ( 33%)
2       39 ( 26%)


Class attribute: class
Classes to Clusters:

  0  1  2  <-- assigned to cluster
  0 50  0 | Iris-setosa
 47  0  3 | Iris-versicolor
 14  0 36 | Iris-virginica

Cluster 0 <-- Iris-versicolor
Cluster 1 <-- Iris-setosa
Cluster 2 <-- Iris-virginica

Incorrectly clustered instances :       17.0     11.3333 %
```

Status

OK

Log

x 0

Association

Association rules

- With association rules, there is no "class" attribute
- Rules can predict any attribute, or combination of attributes
- Need a different kind of algorithm: "Apriori"

Here are some association rules for the weather data:

1. outlook = overcast ==> play = yes
2. temperature = cool ==> humidity = normal
3. humidity = normal & windy = false ==> play = yes
4. outlook = sunny & play = no ==> humidity = high
5. outlook = sunny & humidity = high ==> play = no
6. outlook = rainy & play = yes ==> windy = false
7. outlook = rainy & windy = false ==> play = yes
8. temperature = cool & play = yes ==> humidity = normal
9. outlook = sunny & temperature = hot ==> humidity = high

Outlook	Temp	Humidity	Windy	Play
sunny	hot	high	false	no
sunny	hot	high	true	no
overcast	hot	high	false	yes
rainy	mild	high	false	yes
rainy	cool	normal	false	yes
rainy	cool	normal	true	no
overcast	cool	normal	true	yes
sunny	mild	high	false	no
sunny	cool	normal	false	yes
rainy	mild	normal	false	yes
sunny	mild	normal	true	yes
overcast	mild	high	true	yes
overcast	hot	normal	false	yes

			support	confidence
1. outlook = overcast	==>	play = yes	4	100%
2. temperature = cool	==>	humidity = normal	4	100%
3. humidity = normal & windy = false	==>	play = yes	4	100%
4. outlook = sunny & play = no	==>	humidity = high	3	100%
5. outlook = sunny & humidity = high	==>	play = no	3	100%
6. outlook = rainy & play = yes	==>	windy = false	3	100%
7. outlook = rainy & windy = false	==>	play = yes	3	100%
8. temperature = cool & play = yes	==>	humidity = normal	3	100%
9. outlook = sunny & temperature = hot	==>	humidity = high	2	100%
10. temperature = hot & play = no	==>	outlook = sunny	2	100%

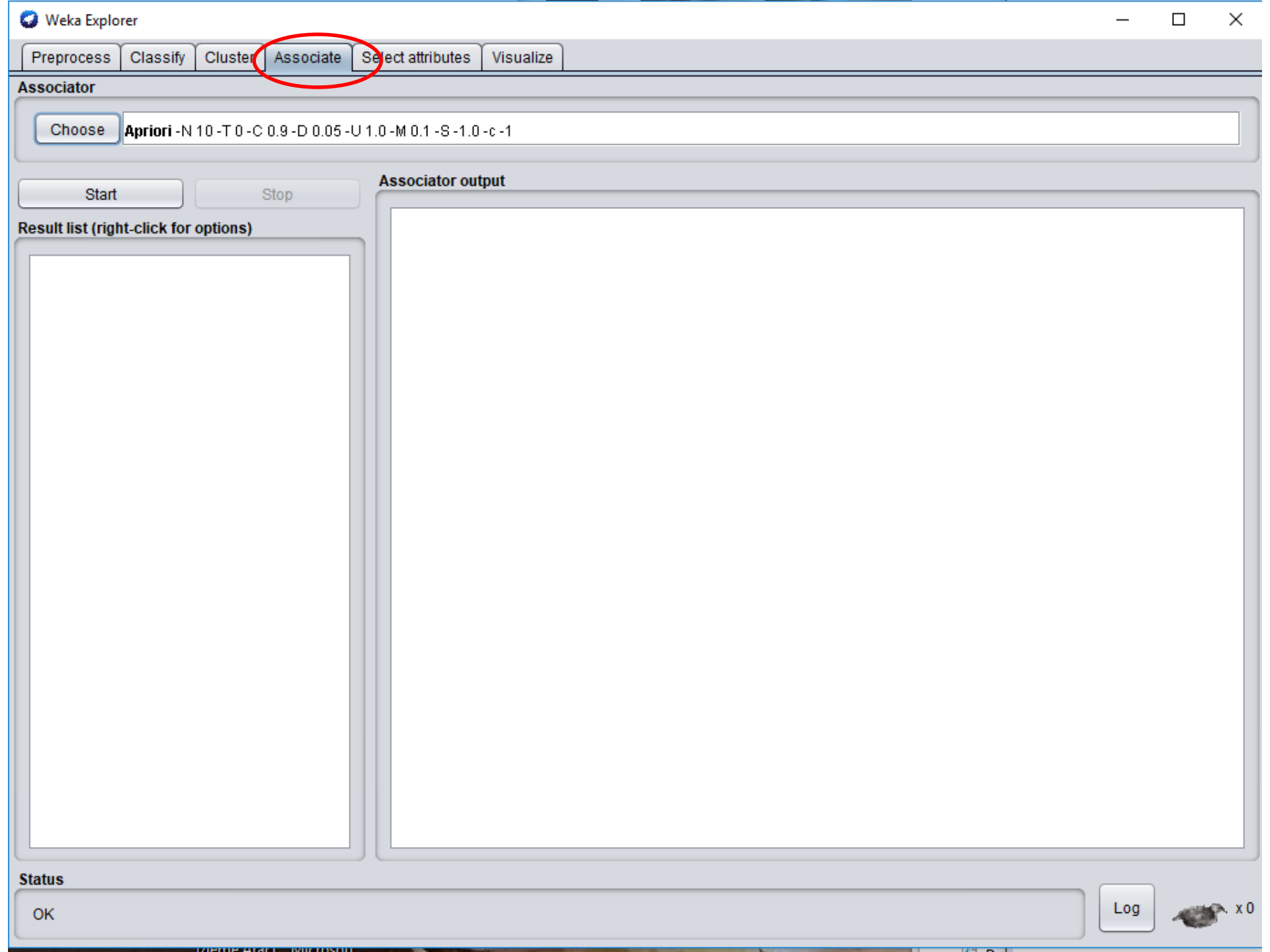
- **Itemset** set of attribute-value pairs, e.g.

  humidity = normal & windy = false & play = yes support = 4

- 7 potential rules from this itemset:

	support	confidence
If humidity = normal & windy = false ==> play = yes	4	4/4
If humidity = normal & play = yes ==> windy = false	4	4/6
If windy = false & play = yes ==> humidity = normal	4	4/6
If humidity = normal ==> windy = false & play = yes	4	4/7
If windy = false ==> humidity = normal & play = yes	4	4/8
If play = yes ==> humidity = normal & windy = false	4	4/9
==> humidity = normal & windy = false & play = yes	4	4/14

- Generate high-support itemsets, get several rules from each
- Strategy: iteratively reduce the minimum support until the required number of rules is found with a given minimum confidence


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

14:27:22 - Apriori

Associator output

```
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Size of set of large itemsets L(2): 47

Size of set of large itemsets L(3): 39

Size of set of large itemsets L(4): 6

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 4. outlook=sunny play=no 3 ==> humidity=high 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 5. outlook=sunny humidity=high 3 ==> play=no 3    <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
 6. outlook=rainy play=yes 3 ==> windy=FALSE 3    <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
 7. outlook=rainy windy=FALSE 3 ==> play=yes 3    <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
 8. temperature=cool play=yes 3 ==> humidity=normal 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 9. outlook=sunny temperature=hot 2 ==> humidity=high 2    <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2    <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)
```

Status

OK

Log

x 0

- Weather data has 336 rules with confidence 100%!
  - *but only 8 have support ≥ 3, only 58 have support ≥ 2*
- Weka: specify minimum confidence level (minMetric, default 90%)
  number of rules sought (numRules, default 10)
- Support is expressed as a proportion of the number of instances
- Weka runs Apriori algorithm several times
  starts at upperBoundMinSupport (usually left at 100%)
  decreases by delta at each iteration (default 5%)
  stops when numRules reached
  … or at lowerBoundMinSupport (default 10%)

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17
Generated sets of large itemsets:
Size of set of large itemsets L(1): 12
Size of set of large itemsets L(2): 47
Size of set of large itemsets L(3): 39
Size of set of large itemsets L(4): 6

Best rules found:
1. outlook = overcast 4 ==> play = yes 4

- 17 cycles of Apriori algorithm:
  - *support = 100%, 95%, 90%, …, 20%, 15%*
  - *14, 13, 13, …, 3, 2 instances*
  - *only 8 rules with conf > 0.9 & support ≥ 3*
- to see itemsets, set outputItemSets
  - *they're based on the final support value, i.e. 2*

12 one-item sets with support ≥ 2

outlook = sunny 5
outlook = overcast 4
...
play = no 5

47 two-item sets with support ≥ 2

outlook = sunny & temperature = hot 2
outlook = sunny & humidity = high 3
...

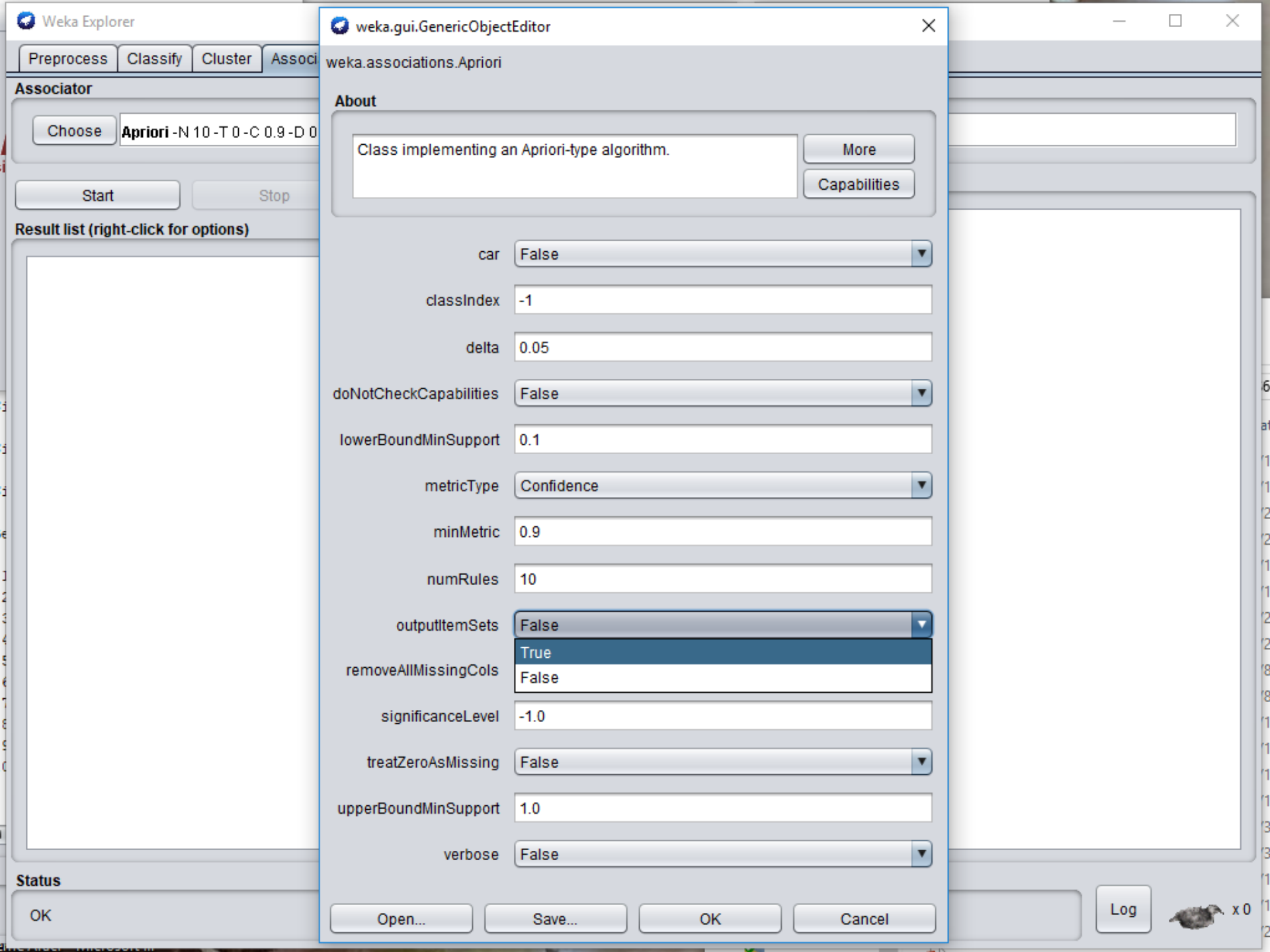
39 three-item sets with support ≥ 2

outlook = sunny & temperature=hot & humidity=high 2
outlook = sunny & humidity = high & play = no 3
outlook = sunny & windy = false & play = no 2
...

6 four-item sets with support ≥ 2

outlook = sunny & humidity = high & windy = false & play = no 2

Subtitles off
English, selected
简体中文


Weka Explorer
Preprocess
Classify
Cluster
Associ
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
-1
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.1
metricType
Confidence
minMetric
0.9
numRules
10
outputItemSets
False
True
False
removeAllMissingCols
significanceLevel
-1.0
treatZeroAsMissing
False
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

16:45:16 - Apriori
16:45:36 - Apriori

Associator output

```
=== Run information ===

Scheme:       weka.associations.Apriori -I -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Relation:     weather.symbolic
Instances:    14
Attributes:   5
              outlook
              temperature
              humidity
              windy
              play
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.15 (2 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 17

Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Large Itemsets L(1):
outlook=sunny 5
outlook=overcast 4
outlook=rainy 5
temperature=hot 4
```

Status

OK

Log

x 0

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

16:45:16 - Apriori
16:45:36 - Apriori

Associator output

```
Generated sets of large itemsets:

Size of set of large itemsets L(1): 12

Large Itemsets L(1):
outlook=sunny 5
outlook=overcast 4
outlook=rainy 5
temperature=hot 4
temperature=mild 6
temperature=cool 4
humidity=high 7
humidity=normal 7
windy=TRUE 6
windy=FALSE 8
play=yes 9
play=no 5

Size of set of large itemsets L(2): 47

Large Itemsets L(2):
outlook=sunny temperature=hot 2
outlook=sunny temperature=mild 2
outlook=sunny humidity=high 3
outlook=sunny humidity=normal 2
outlook=sunny windy=TRUE 2
outlook=sunny windy=FALSE 3
outlook=sunny play=yes 2
outlook=sunny play=no 3
outlook=overcast temperature=hot 2
```

Status

OK

Log

x 0

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

16:45:16 - Apriori
16:45:36 - Apriori

Associator output

```
windy=FALSE play=yes 6
windy=FALSE play=no 2

Size of set of large itemsets L(3): 39

Large Itemsets L(3):
outlook=sunny temperature=hot humidity=high 2
outlook=sunny temperature=hot play=no 2
outlook=sunny humidity=high windy=FALSE 2
outlook=sunny humidity=high play=no 3
outlook=sunny humidity=normal play=yes 2
outlook=sunny windy=FALSE play=no 2
outlook=overcast temperature=hot windy=FALSE 2
outlook=overcast temperature=hot play=yes 2
outlook=overcast humidity=high play=yes 2
outlook=overcast humidity=normal play=yes 2
outlook=overcast windy=TRUE play=yes 2
outlook=overcast windy=FALSE play=yes 2
outlook=rainy temperature=mild humidity=high 2
outlook=rainy temperature=mild windy=FALSE 2
outlook=rainy temperature=mild play=yes 2
outlook=rainy temperature=cool humidity=normal 2
outlook=rainy humidity=normal windy=FALSE 2
outlook=rainy humidity=normal play=yes 2
outlook=rainy windy=TRUE play=no 2
outlook=rainy windy=FALSE play=yes 3
temperature=hot humidity=high windy=FALSE 2
temperature=hot humidity=high play=no 2
temperature=hot windy=FALSE play=yes 2
temperature=mild humidity=high windy=TRUE 2
```

Status

OK

Log

x 0

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

16:45:16 - Apriori
16:45:36 - Apriori

Associator output

```
humidity=high windy=TRUE play=no 2
humidity=high windy=FALSE play=yes 2
humidity=high windy=FALSE play=no 2
humidity=normal windy=TRUE play=yes 2
humidity=normal windy=FALSE play=yes 4

Size of set of large itemsets L(4): 6

Large Itemsets L(4):
outlook=sunny temperature=hot humidity=high play=no 2
outlook=sunny humidity=high windy=FALSE play=no 2
outlook=overcast temperature=hot windy=FALSE play=yes 2
outlook=rainy temperature=mild windy=FALSE play=yes 2
outlook=rainy humidity=normal windy=FALSE play=yes 2
temperature=cool humidity=normal windy=FALSE play=yes 2

Best rules found:

 1. outlook=overcast 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 2. temperature=cool 4 ==> humidity=normal 4    <conf:(1)> lift:(2) lev:(0.14) [2] conv:(2)
 3. humidity=normal windy=FALSE 4 ==> play=yes 4    <conf:(1)> lift:(1.56) lev:(0.1) [1] conv:(1.43)
 4. outlook=sunny play=no 3 ==> humidity=high 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 5. outlook=sunny humidity=high 3 ==> play=no 3    <conf:(1)> lift:(2.8) lev:(0.14) [1] conv:(1.93)
 6. outlook=rainy play=yes 3 ==> windy=FALSE 3    <conf:(1)> lift:(1.75) lev:(0.09) [1] conv:(1.29)
 7. outlook=rainy windy=FALSE 3 ==> play=yes 3    <conf:(1)> lift:(1.56) lev:(0.08) [1] conv:(1.07)
 8. temperature=cool play=yes 3 ==> humidity=normal 3    <conf:(1)> lift:(2) lev:(0.11) [1] conv:(1.5)
 9. outlook=sunny temperature=hot 2 ==> humidity=high 2    <conf:(1)> lift:(2) lev:(0.07) [1] conv:(1)
10. temperature=hot play=no 2 ==> outlook=sunny 2    <conf:(1)> lift:(2.8) lev:(0.09) [1] conv:(1.29)
```

Status

OK

Log

x 0

Market basket analysis

- Look at supermarket.arff
  - *collected from an actual New Zealand supermarket*
- 4500 instances, 220 attributes; 1M attribute values
- Missing values used to indicate that the basket did not contain that item
- 92% of values are missing
  - *average basket contains 220×8% = 18 items*
- Most popular items: bread-and-cake (3330), vegetables (2961), frozen foods (2717), biscuits (2605)

Weka Explorer

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

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

Filter

Choose | None | Apply

Current relation

Relation: supermarket

Instances: 4627 Attributes: 217

Attributes

All | None | Invert | Pattern

No.	Name
11	department11
12	baby needs
13	bread and cake
14	baking needs
15	coupons
16	juice-sat-cord-ms
17	tea
18	biscuits
19	canned fish-meat
20	canned fruit
21	canned vegetables
22	breakfast food
23	cigs-tobacco pkts

Remove

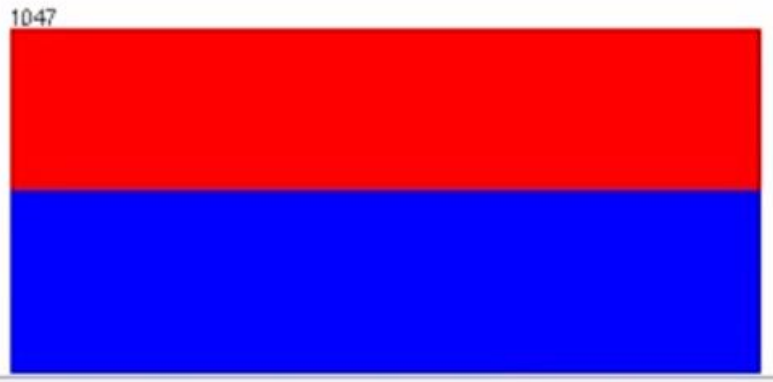
Selected attribute

Name: department1 Type: Nominal

Missing: 3580 (77%) Distinct: 1 Unique: 0 (0%)

No.	Label	Count
1	t	1047

Class: total (Nom) | Visualize All



Status

OK

Log x 0

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-clic...

14:27:22 - Apriori
14:42:39 - Apriori

Associator output

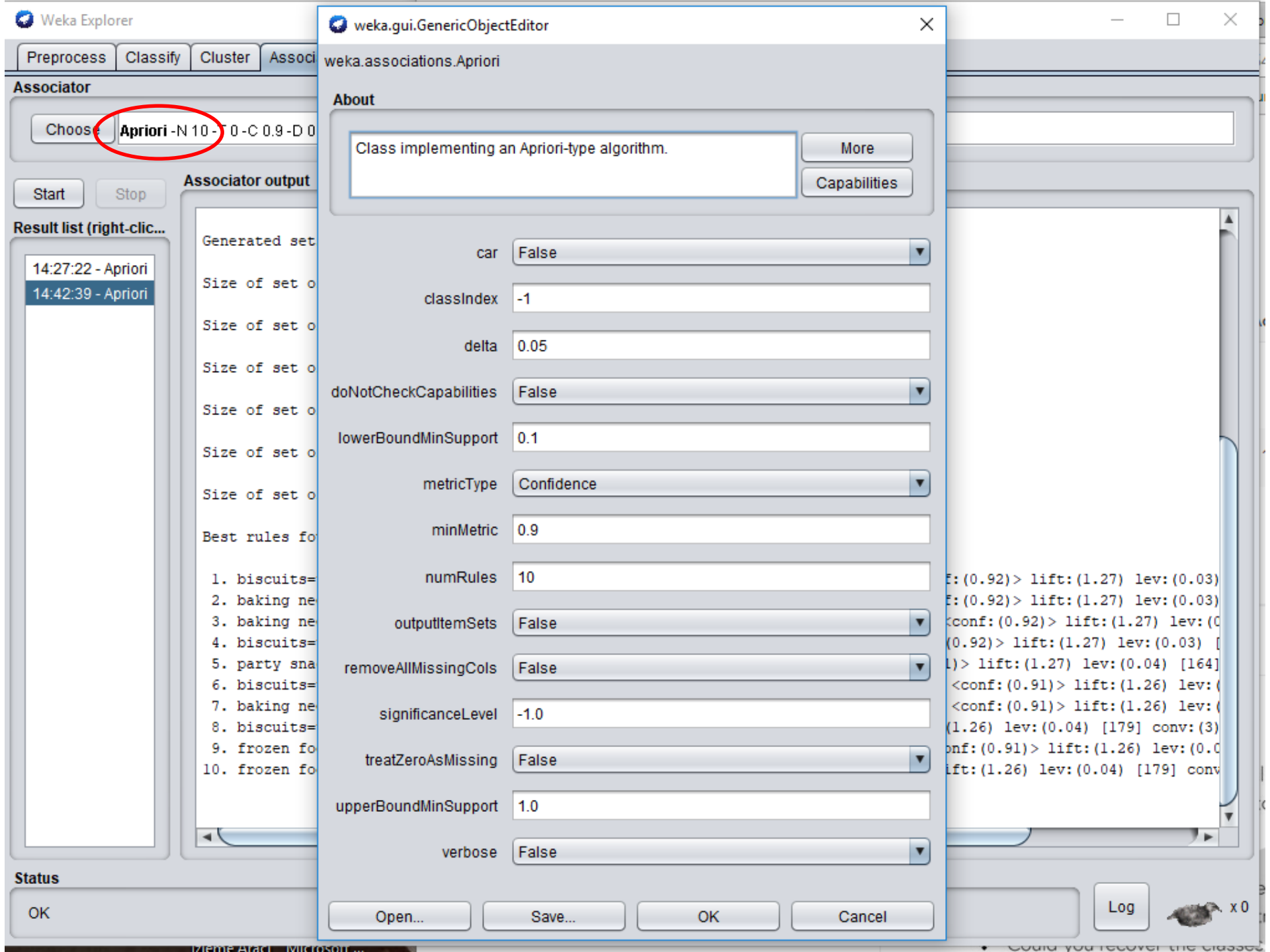
```
Relation:     supermarket
Instances:    4627
Attributes:   217
              [list of attributes omitted]
=== Associator model (full training set) ===


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 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:
```

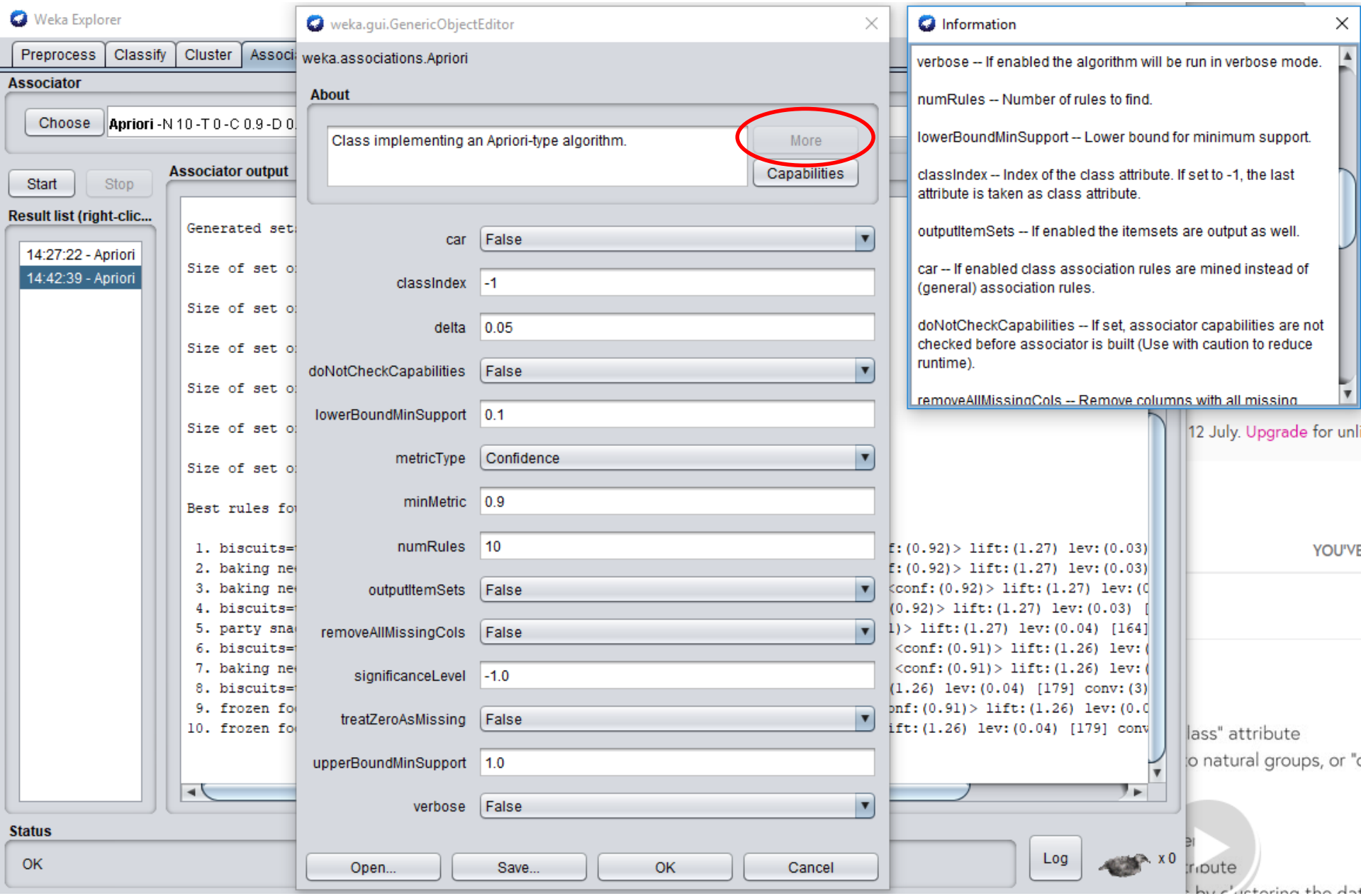
Status

OK

Log

x 0


Weka Explorer
Preprocess
Classify
Cluster
Associ
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0
Start
Stop
Result list (right-clic...
14:27:22 - Apriori
14:42:39 - Apriori
Associator output
Generated set
Size of set o
Size of set o
Size of set o
Size of set o
Size of set o
Size of set o
Best rules fo
1. biscuits=
2. baking ne
3. baking ne
4. biscuits=
5. party sna
6. biscuits=
7. baking ne
8. biscuits=
9. frozen fo
10. frozen fo
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
-1
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.1
metricType
Confidence
minMetric
0.9
numRules
10
outputItemSets
False
removeAllMissingCols
False
significanceLevel
-1.0
treatZeroAsMissing
False
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associ
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.
Start
Stop
Result list (right-clic...
14:27:22 - Apriori
14:42:39 - Apriori
Associator output
Generated set
Size of set o
Size of set o
Size of set o
Size of set o
Size of set o
Size of set o
Best rules fo
1. biscuits=
2. baking ne
3. baking ne
4. biscuits=
5. party sna
6. biscuits=
7. baking ne
8. biscuits=
9. frozen fo
10. frozen fo
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Class implementing an Apriori-type algorithm.
More
Capabilities
car
False
classIndex
-1
delta
0.05
doNotCheckCapabilities
False
lowerBoundMinSupport
0.1
metricType
Confidence
minMetric
0.9
numRules
10
outputItemSets
False
removeAllMissingCols
False
significanceLevel
-1.0
treatZeroAsMissing
False
upperBoundMinSupport
1.0
verbose
False
Open...
Save...
OK
Cancel
Information
verbose -- If enabled the algorithm will be run in verbose mode.
numRules -- Number of rules to find.
lowerBoundMinSupport -- Lower bound for minimum support.
classIndex -- Index of the class attribute. If set to -1, the last attribute is taken as class attribute.
outputItemSets -- If enabled the itemsets are output as well.
car -- If enabled class association rules are mined instead of (general) association rules.
doNotCheckCapabilities -- If set, associator capabilities are not checked before associator is built (Use with caution to reduce runtime).
removeAllMissingCols -- Remove columns with all missing
12 July. Upgrade for unli
YOU'VE
lass" attribute
o natural groups, or "