Machine Learning with WEKA

Bernhard Pfahringer

(based on material by Eibe Frank, Mark Hall, and Peter Reutemann)

Department of Computer Science, University of Waikato, New Zealand

- WEKA: A Machine Learning Toolkit
- The Explorer
  - Classification and Regression
  - Clustering
  - Association Rules
  - Attribute Selection
  - Data Visualization
- The Experimenter
- The Knowledge Flow GUI
- Other Utilities
- Conclusions

WEKA: the bird

The *Weka* or *woodhen* (Gallirallus australis) is an endemic bird of New Zealand. (Source: *WikiPedia*)

WEKA: the software

- Machine learning/data mining software written in Java (distributed under the GNU Public License)
- Used for research, education, and applications
- Complements “Data Mining” by Witten & Frank
- Main features:
  - Comprehensive set of data pre-processing tools, learning algorithms and evaluation methods
  - Graphical user interfaces (incl. data visualization)
  - Environment for comparing learning algorithms

History

- Project funded by the NZ government since 1993

FRST App Number: 93-WKT-23-719

7. **PROGRAMME GOAL** (State the overall goal of the programme in a maximum of 5 lines).

The programme aims to build a state-of-the-art facility for developing techniques of machine learning and investigating their application in key areas of the New Zealand economy. Specifically we will create a workbench for machine learning, determine the factors that contribute towards its successful application in the agricultural industries, and develop new methods of machine learning and ways of assessing their effectiveness.

- Develop state-of-the art workbench of data mining tools
- Explore fielded applications
- Develop new fundamental methods

History (2)

- Late 1992 - funding was applied for by Ian Witten
- 1993 - development of the interface and infrastructure
  - WEKA acronym coined by Geoff Holmes
  - WEKA's file format "ARFF" was created by Andrew Donkin
    ARFF was rumored to stand for **A**ndrew's **R**idiculous **F**ile **F**ormat
- Sometime in 1994 - first internal release of WEKA
  - TCL/TK user interface + learning algorithms written mostly in C
  - Very much beta software
  - Changes for the b1 release included (among others):

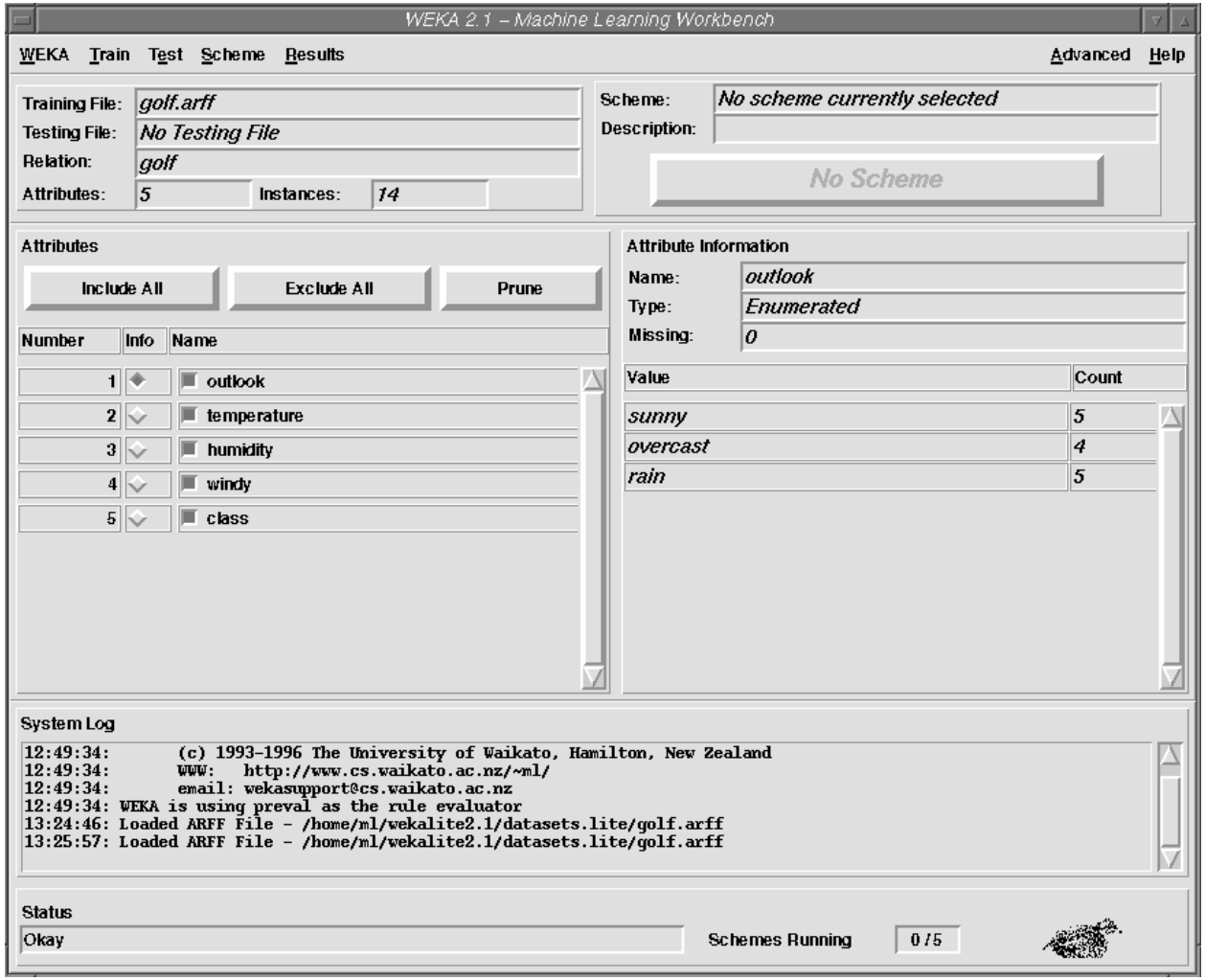
    `"Ambiguous and Unsupported menu commands removed."`

    `"Crashing processes handled (in most cases :-)"`
- October 1996 - first public release: WEKA 2.1

History (3)

- July 1997 - WEKA 2.2
  - Schemes: 1R, T2, K*, M5, M5Class, IB1-4, FOIL, PEBLS, support for C5
  - Included a facility (based on Unix makefiles) for configuring and running large scale experiments
- Early 1997 - decision was made to rewrite WEKA in Java
  - Originated from code written by Eibe Frank for his PhD
  - Originally codenamed **JAWS** (**JA**va **W**eka **S**ystem)
- May 1998 - WEKA 2.3
  - Last release of the TCL/TK-based system
- Mid 1999 - WEKA 3 (100% Java) released
  - Version to complement the Data Mining book
  - Development version (including GUI)

The GUI back then…



TCL/TK interface of Weka 2.1

WEKA: versions

- There are several versions of WEKA:
  - WEKA 3.4: “book version” compatible with description in data mining book
  - WEKA 3.5.5: “development version” with lots of improvements
- This talk is based on a nightly snapshot of WEKA 3.5.5 (12-Feb-2007)

WEKA only deals with “flat” files

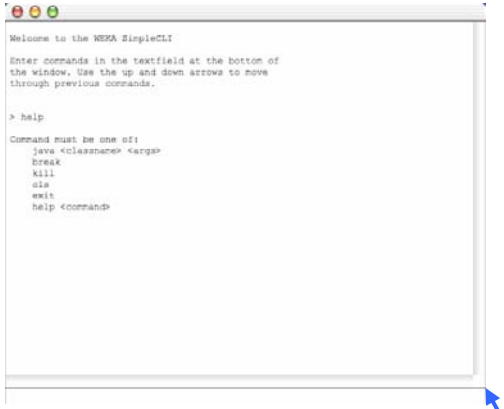
```
@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data
63,male,typ_angina,233,no,not_present
67,male,asympt,286,yes,present
67,male,asympt,229,yes,present
38,female,non_anginal,?,no,not_present
...
```

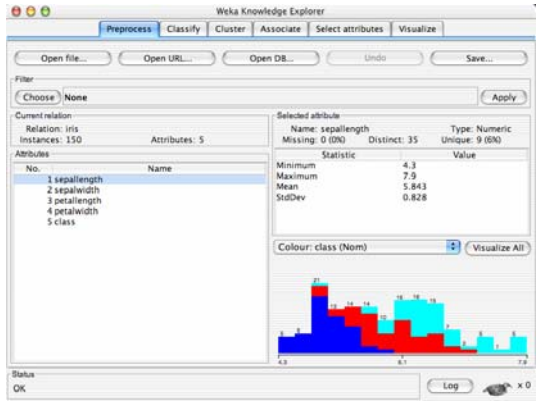
Flat file in ARFF format

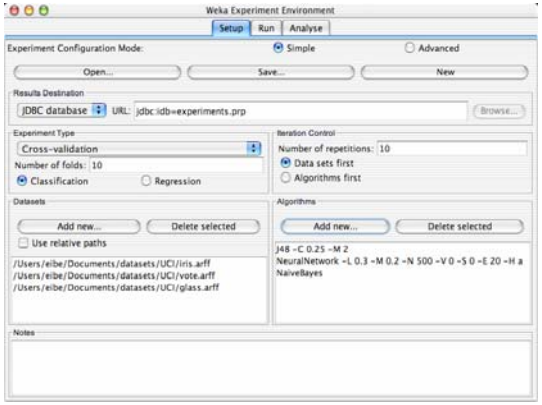
WEKA only deals with “flat” files

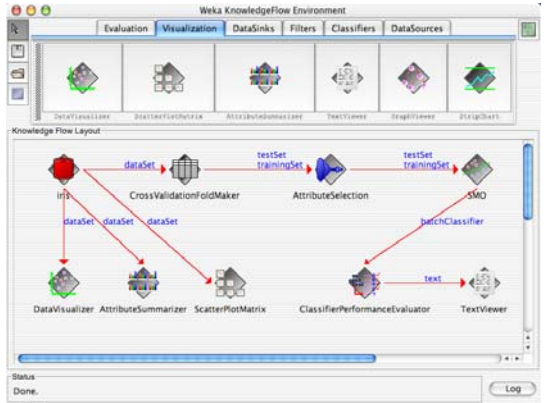

```
@relation heart-disease-simplified

@attribute age numeric
@attribute sex { female, male}
@attribute chest_pain_type { typ_angina, asympt, non_anginal, atyp_angina}
@attribute cholesterol numeric
@attribute exercise_induced_angina { no, yes}
@attribute class { present, not_present}

@data
63,male,typ_angina,233,no,not_present
67,male,asympt,286,yes,present
67,male,asympt,229,yes,present
38,female,non_anginal,?,no,not_present
...
```

java weka.gui.GUIChooser

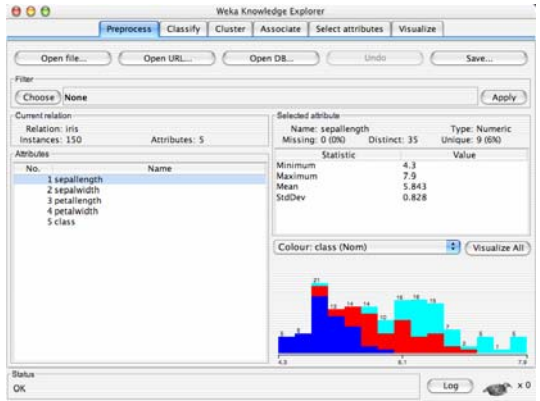


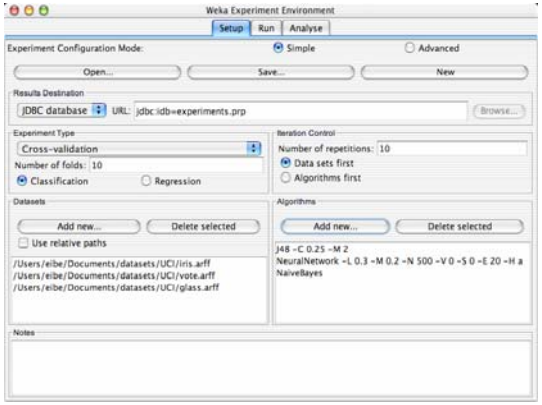


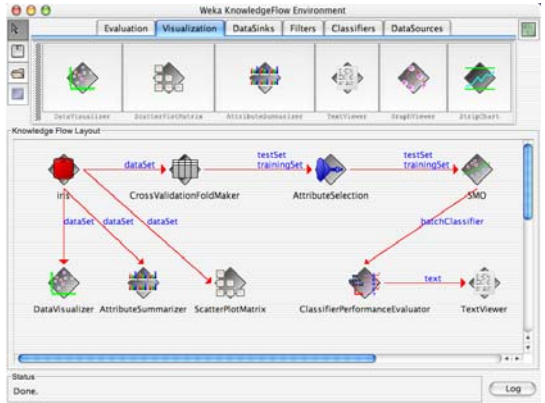




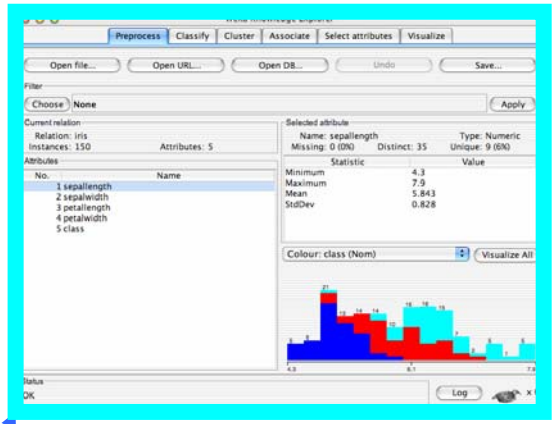

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log

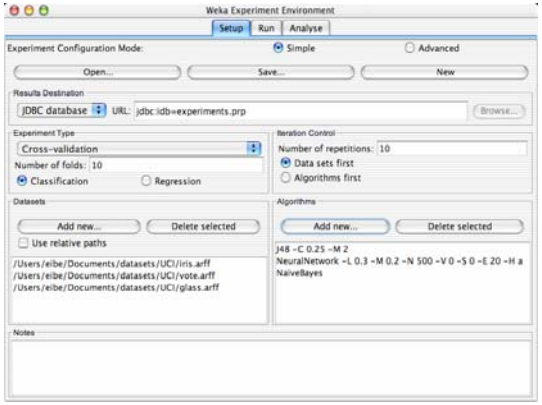

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize

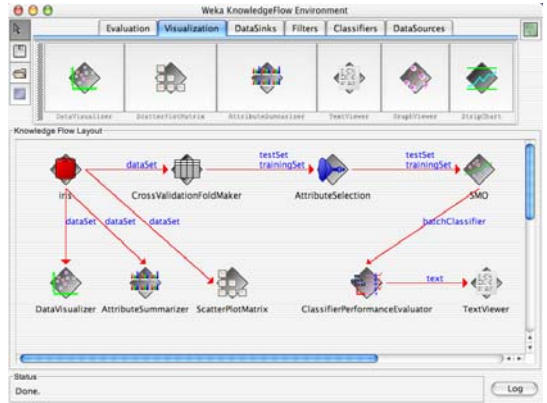

Weka Experiment Environment
Setup
Run
Analyse


Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources

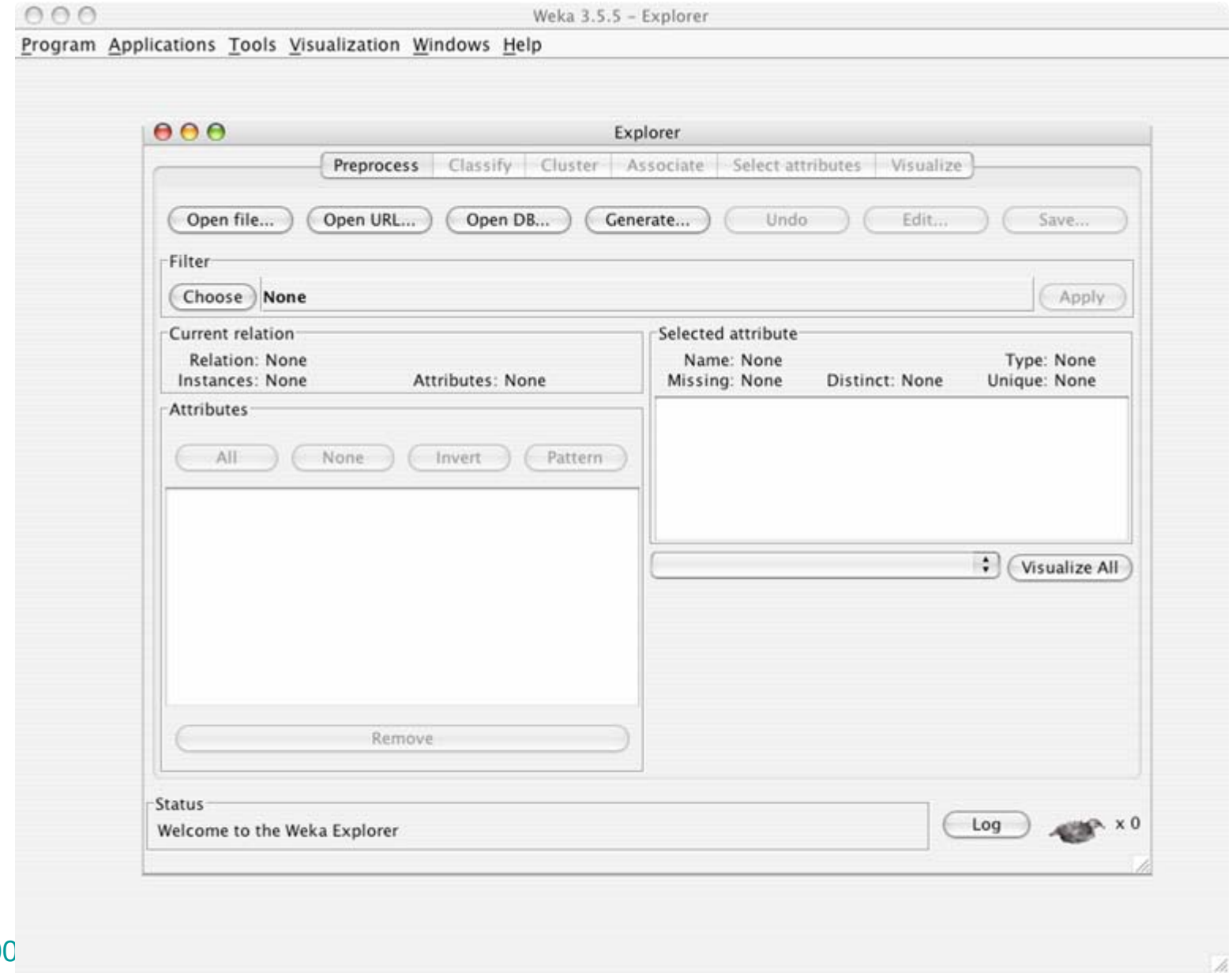

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log





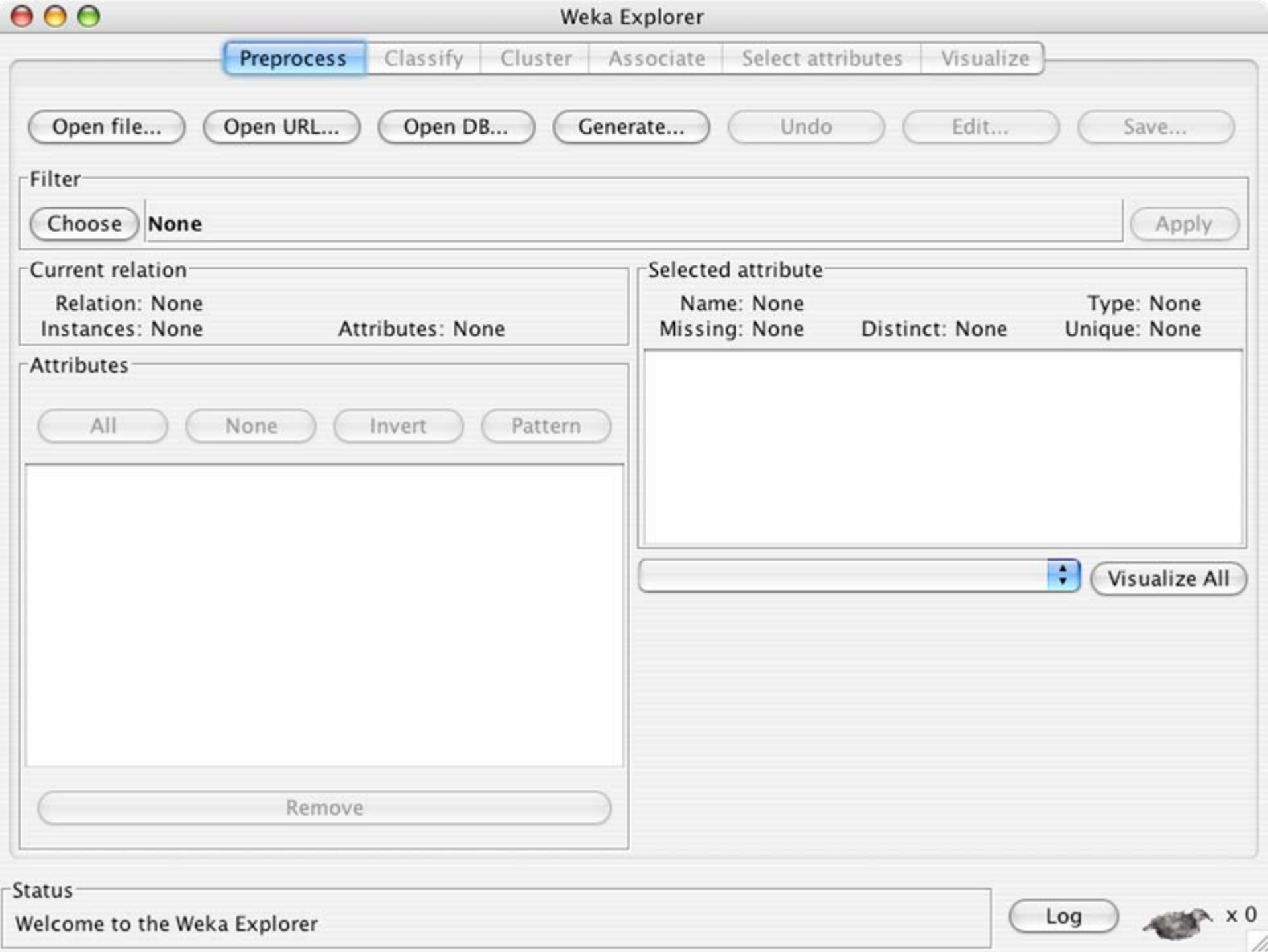


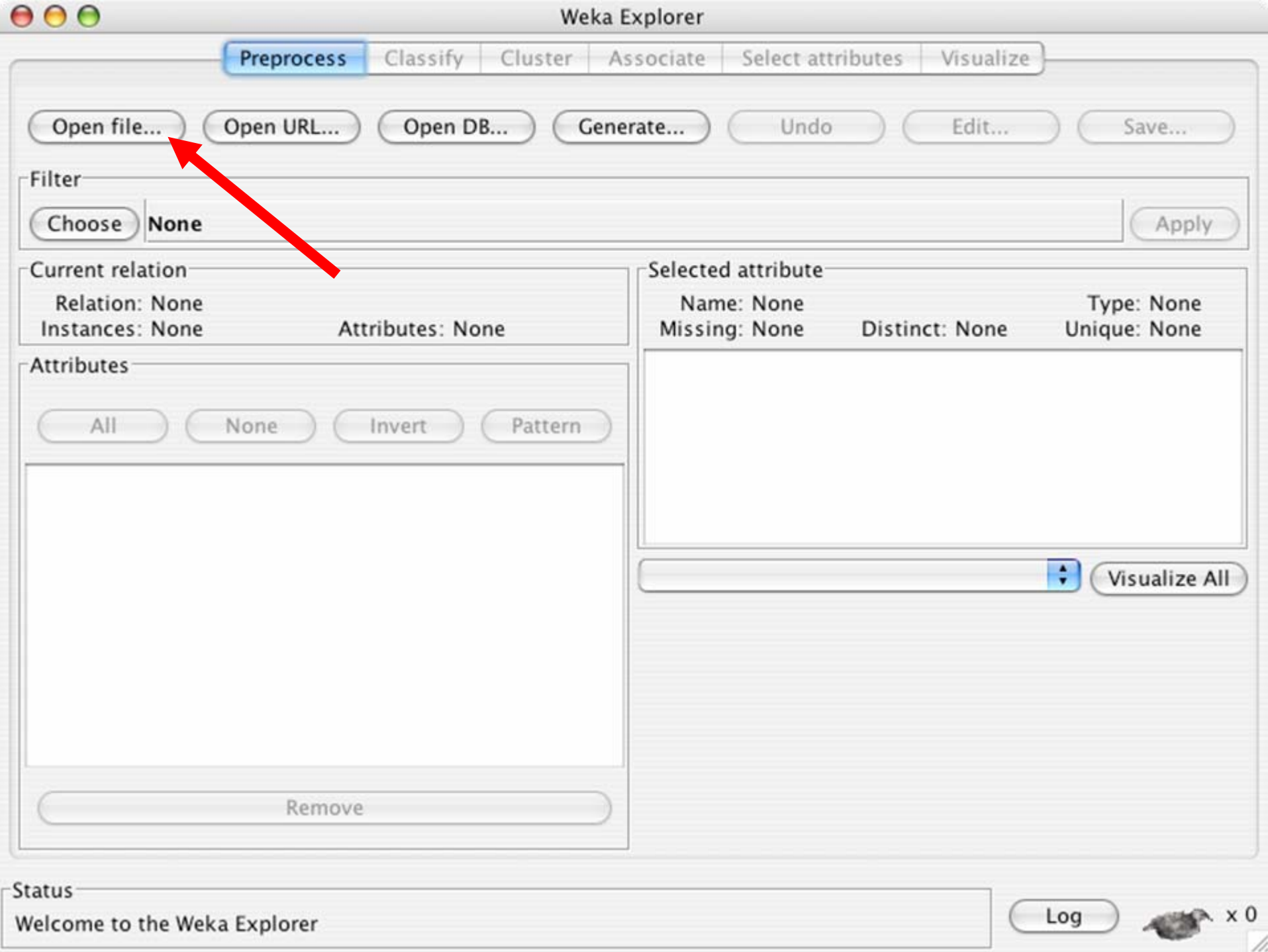
java -jar weka.jar

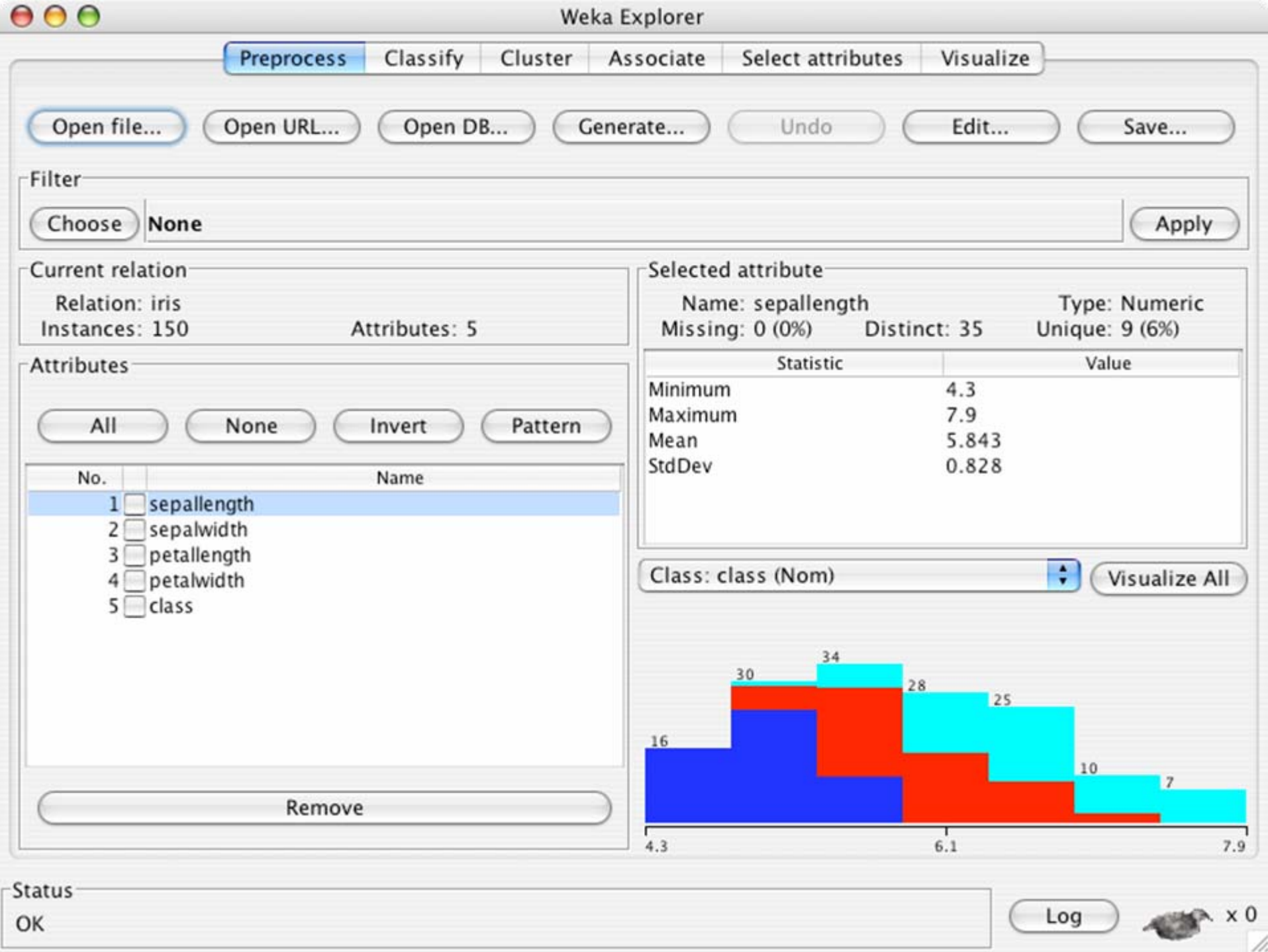


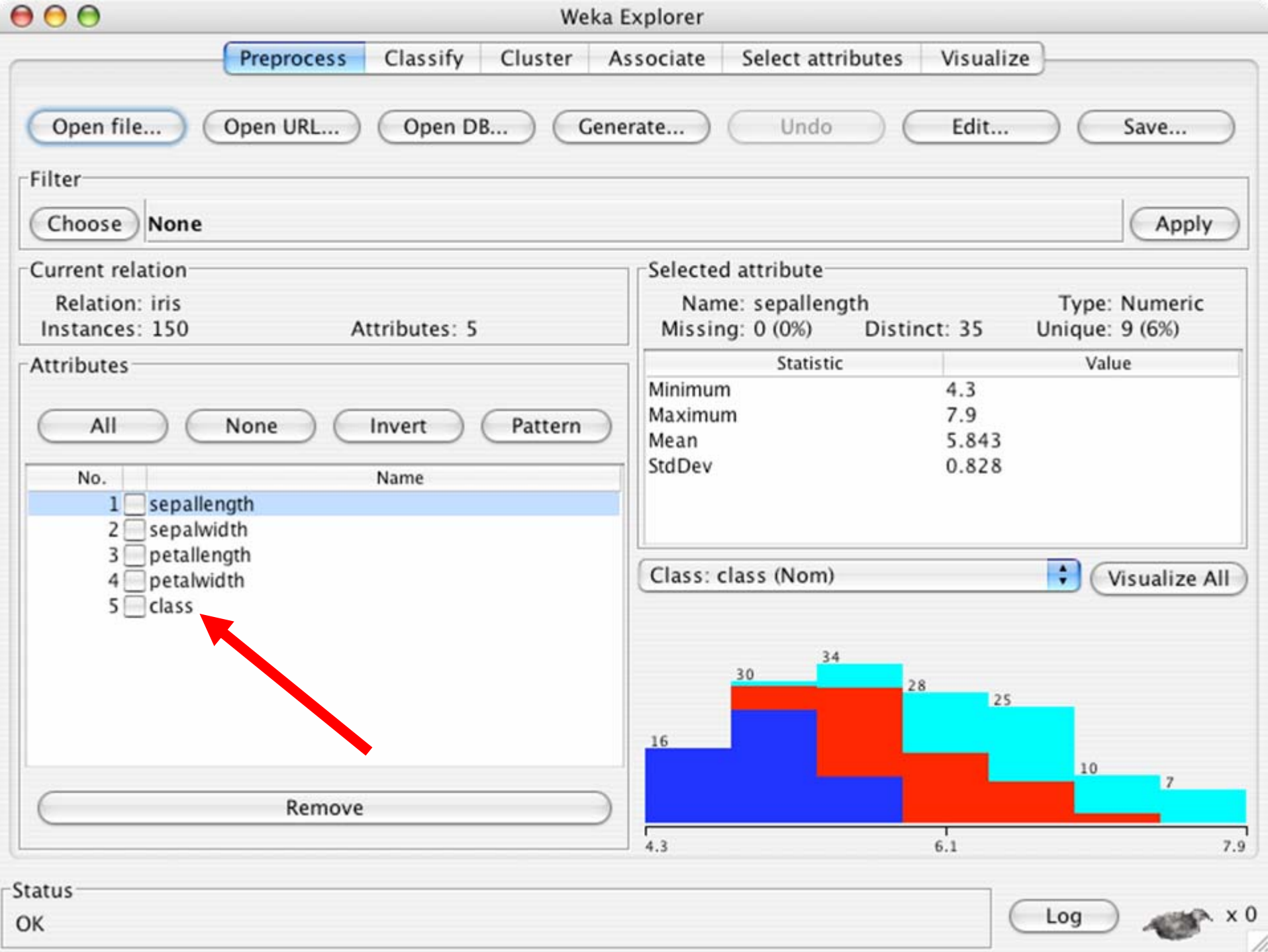
Explorer: pre-processing the data

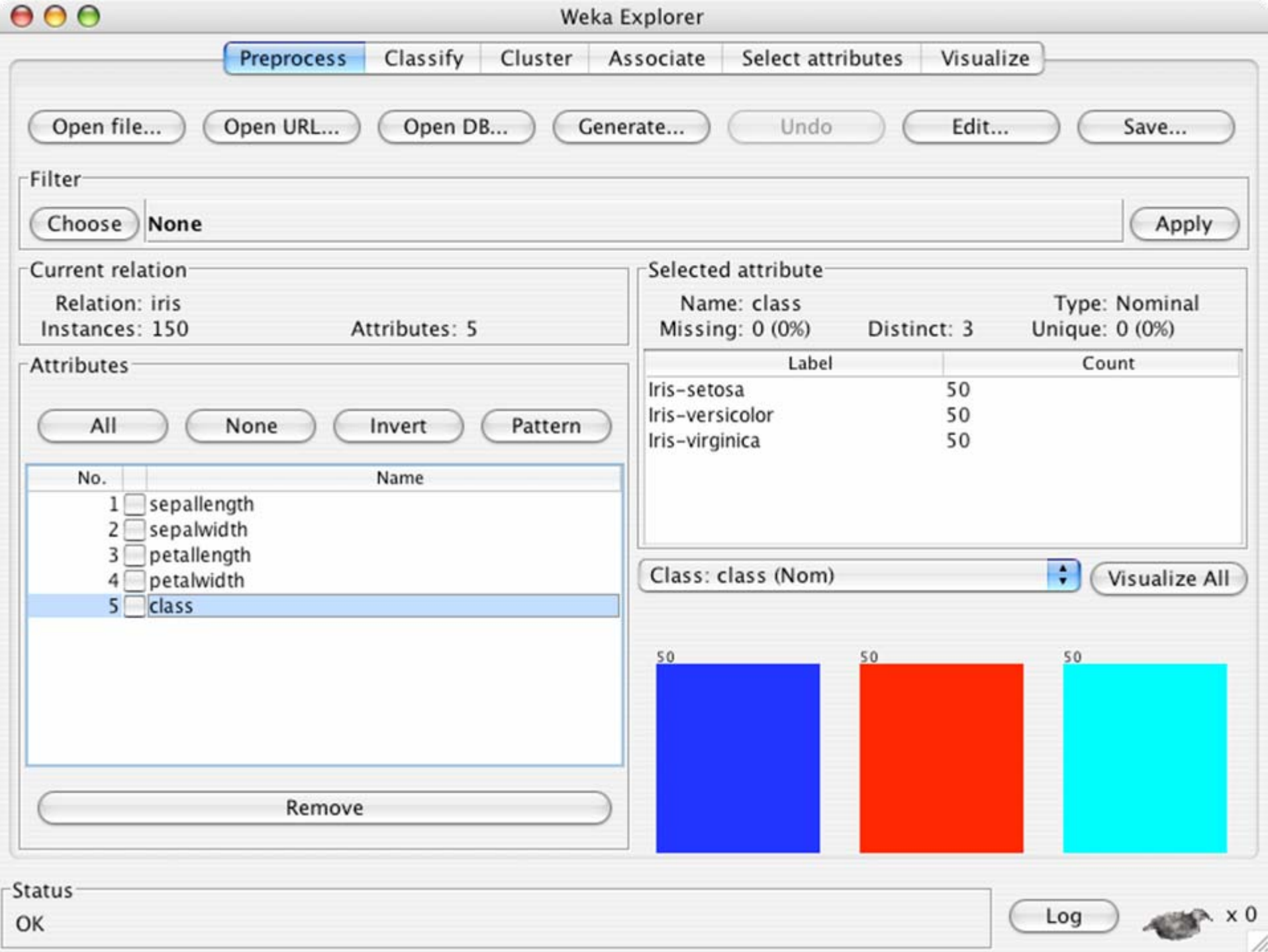
- Data can be imported from a file in various formats: ARFF, CSV, C4.5, binary
- Data can also be read from a URL or from an SQL database (using JDBC)
- Pre-processing tools in WEKA are called “filters”
- WEKA contains filters for:
  - Discretization, normalization, resampling, attribute selection, transforming and combining attributes, …

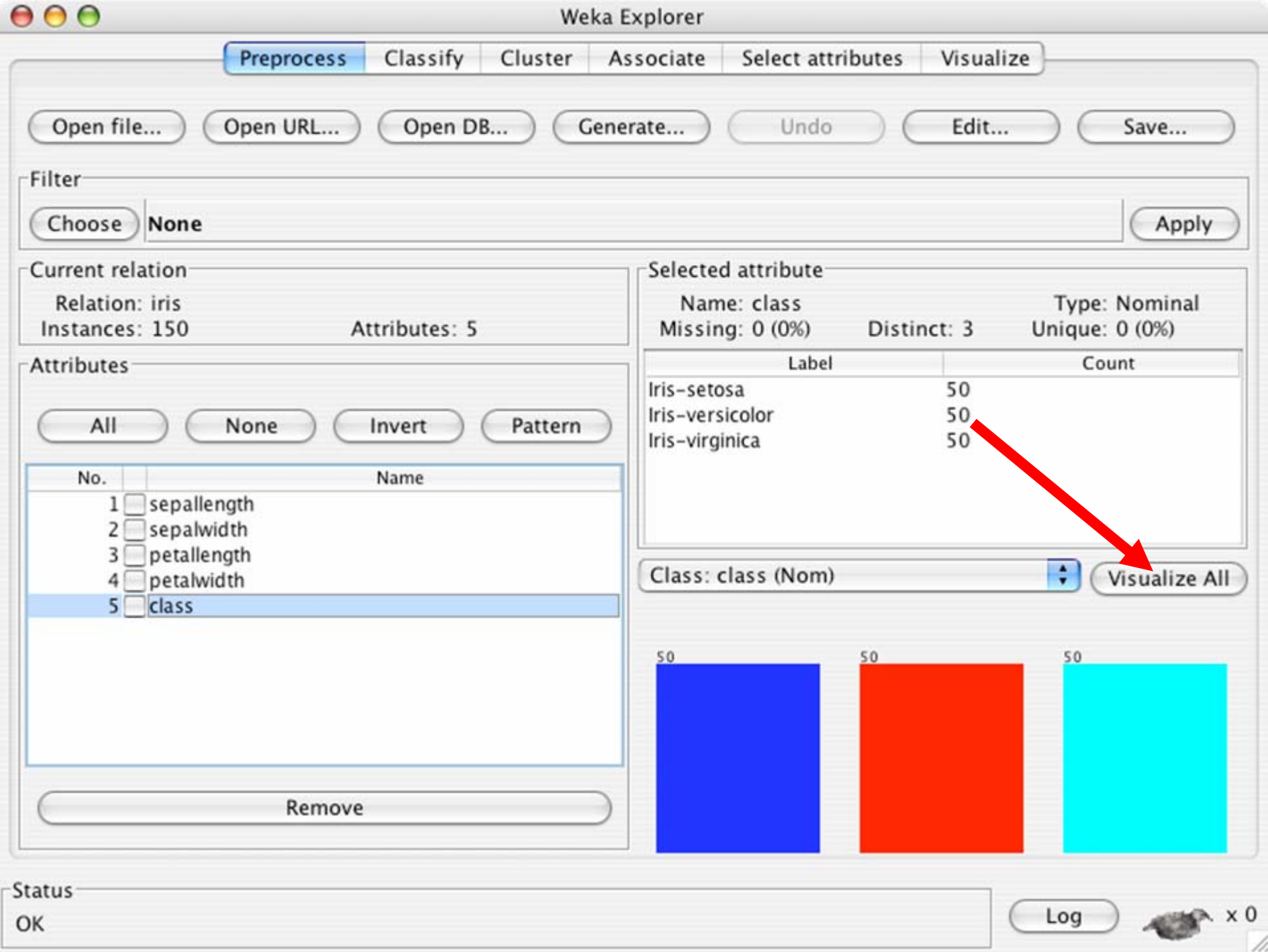

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: None
Instances: None
Attributes: None
Selected attribute
Name: None
Type: None
Missing: None
Distinct: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

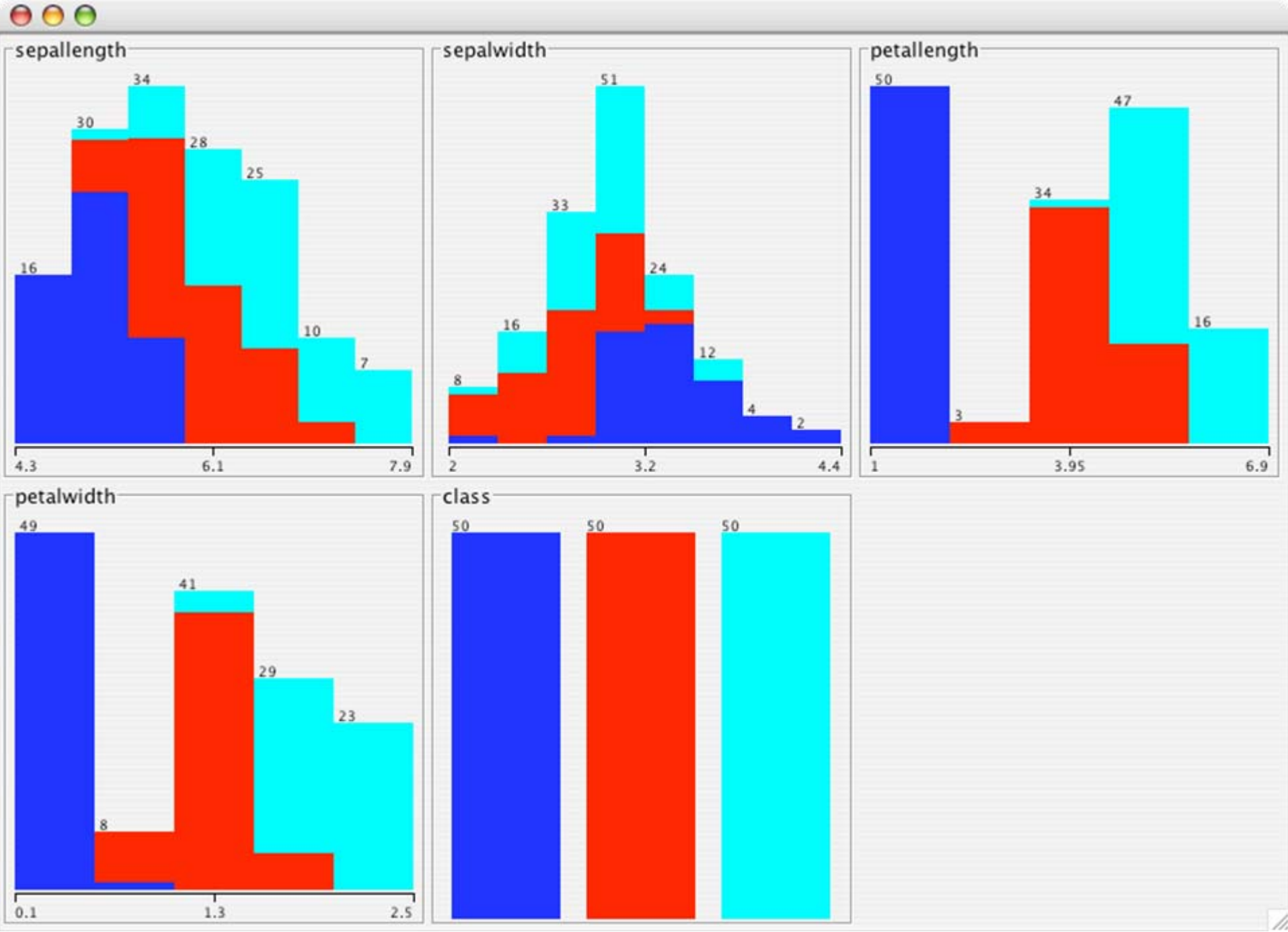

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: None
Instances: None
Attributes: None
Selected attribute
Name: None
Type: None
Missing: None
Distinct: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

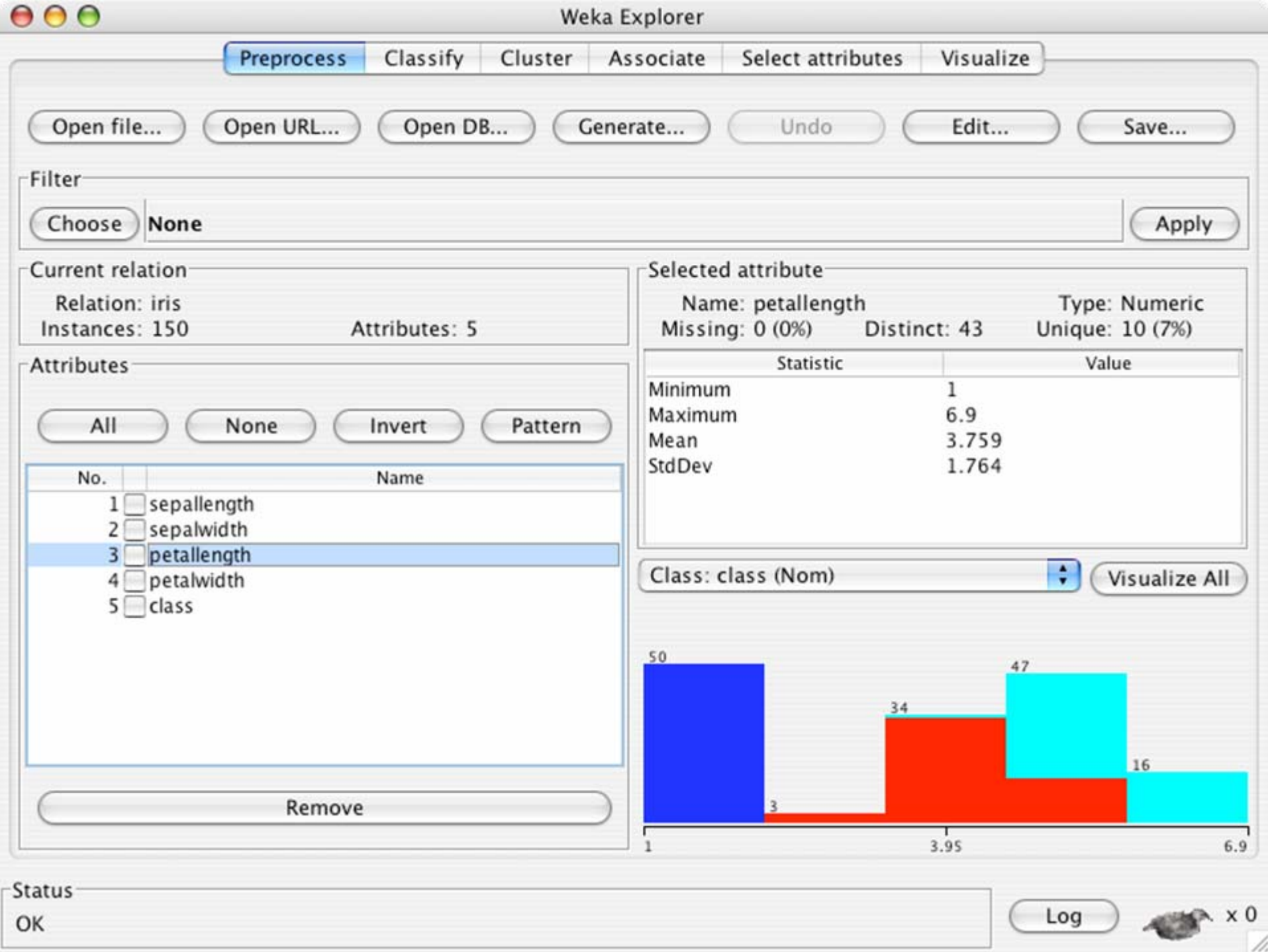

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
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

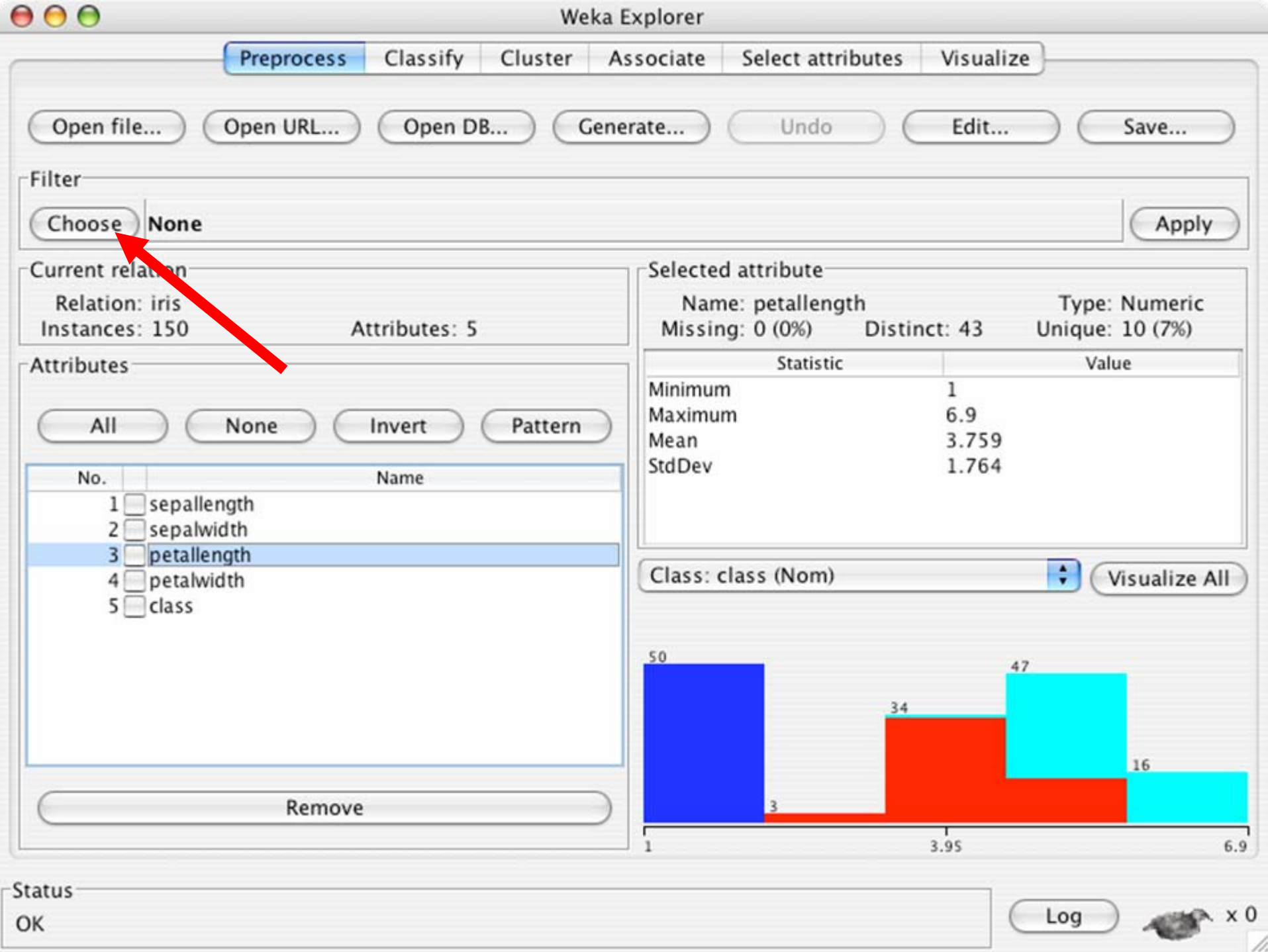

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
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

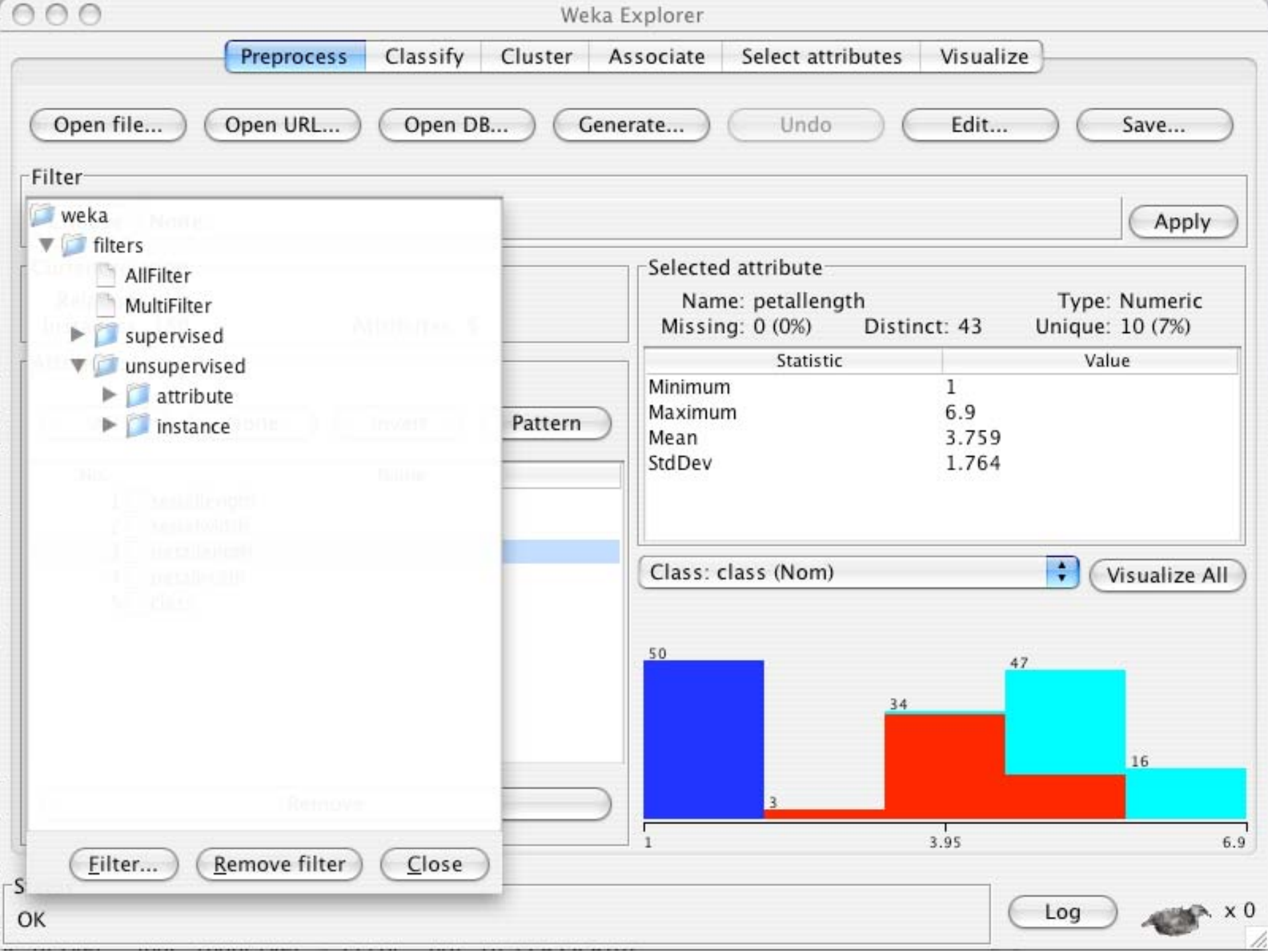

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: class
Type: Nominal
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)
Label
Count
Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
Class: class (Nom)
Visualize All
50
50
50
Status
OK
Log
x 0

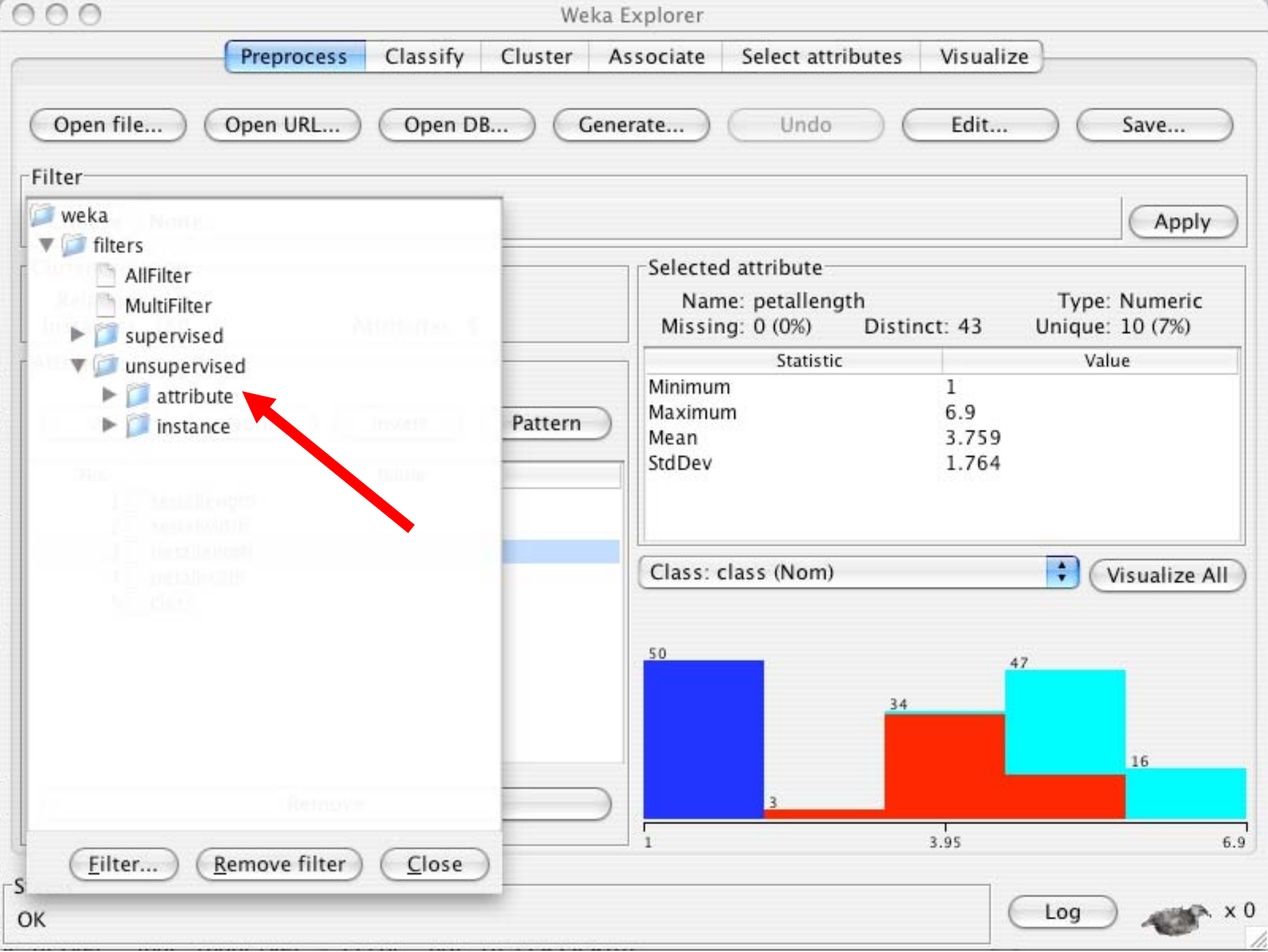

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: class
Type: Nominal
Missing: 0 (0%)
Distinct: 3
Unique: 0 (0%)
Label
Count
Iris-setosa 50
Iris-versicolor 50
Iris-virginica 50
Class: class (Nom)
Visualize All
50
50
50
Status
OK
Log
x 0

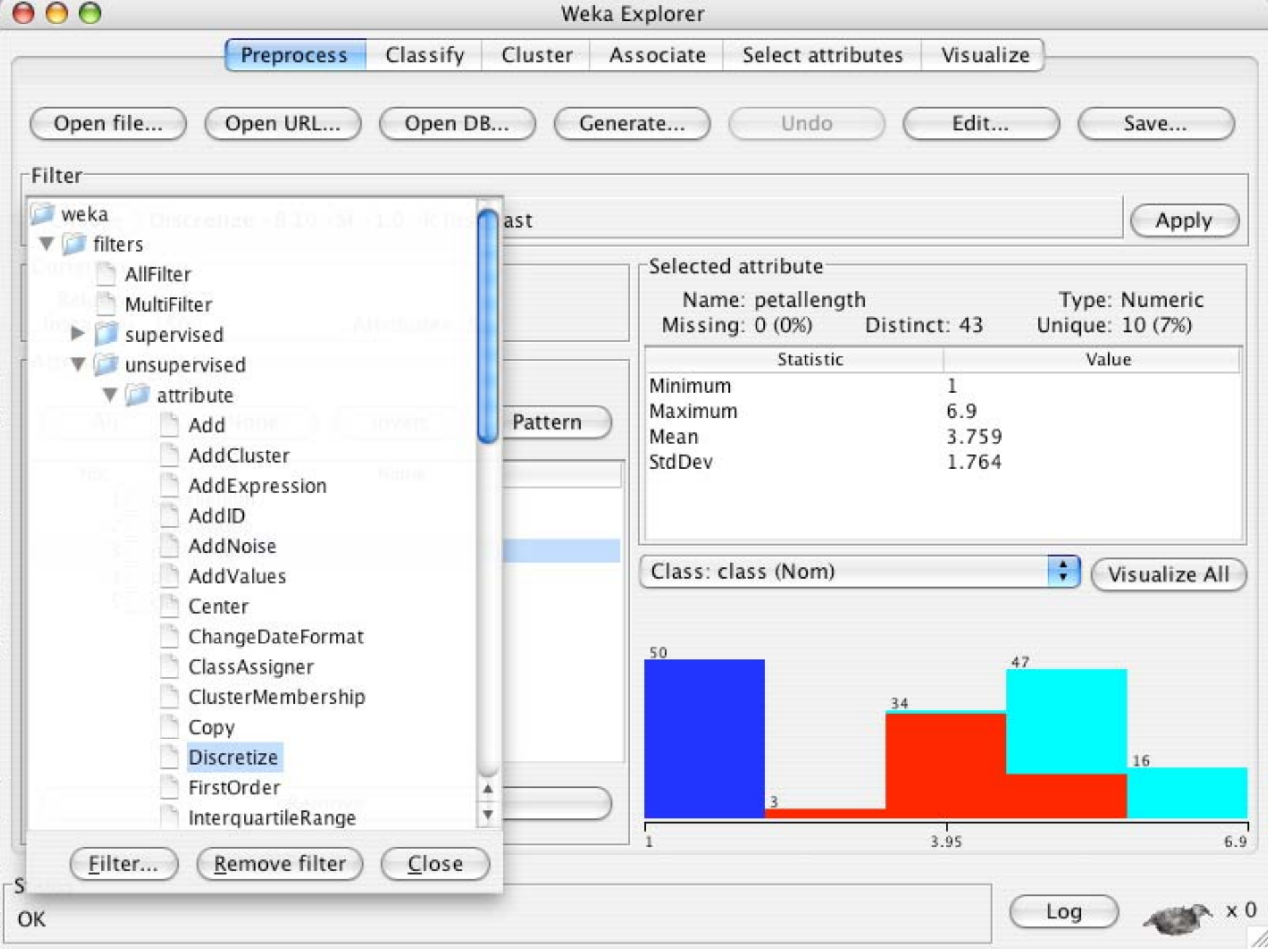

sepallength
34
30
28
25
16
10
7
4.3
6.1
7.9
sepalwidth
51
33
16
24
8
12
4
2
2
3.2
4.4
petallength
50
47
34
3
16
1
3.95
6.9
petalwidth
49
41
29
23
8
0.1
1.3
2.5
class
50
50
50


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: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum 1
Maximum 6.9
Mean 3.759
StdDev 1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
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
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum 1
Maximum 6.9
Mean 3.759
StdDev 1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
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
Apply
weka
filters
AllFilter
MultiFilter
supervised
unsupervised
attribute
instance
Pattern
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum
1
Maximum
6.9
Mean
3.759
StdDev
1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
Filter...
Remove filter
Close
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
weka
filters
AllFilter
MultiFilter
supervised
unsupervised
attribute
instance
Pattern
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum
1
Maximum
6.9
Mean
3.759
StdDev
1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
Filter...
Remove filter
Close
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
weka
filters
AllFilter
MultiFilter
supervised
unsupervised
attribute
Add
AddCluster
AddExpression
AddID
AddNoise
AddValues
Center
ChangeDateFormat
ClassAssigner
ClusterMembership
Copy
Discretize
FirstOrder
InterquartileRange
Filter...
Remove filter
Close
Pattern
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum
1
Maximum
6.9
Mean
3.759
StdDev
1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
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 | **Discretize** -B 10 -M -1.0 -R first-last | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

All | None | Invert | Pattern

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

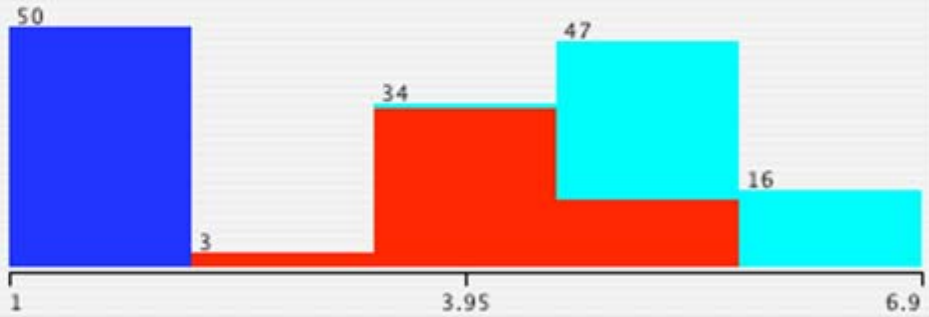
Remove

Selected attribute

Name: petallength Type: Numeric
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

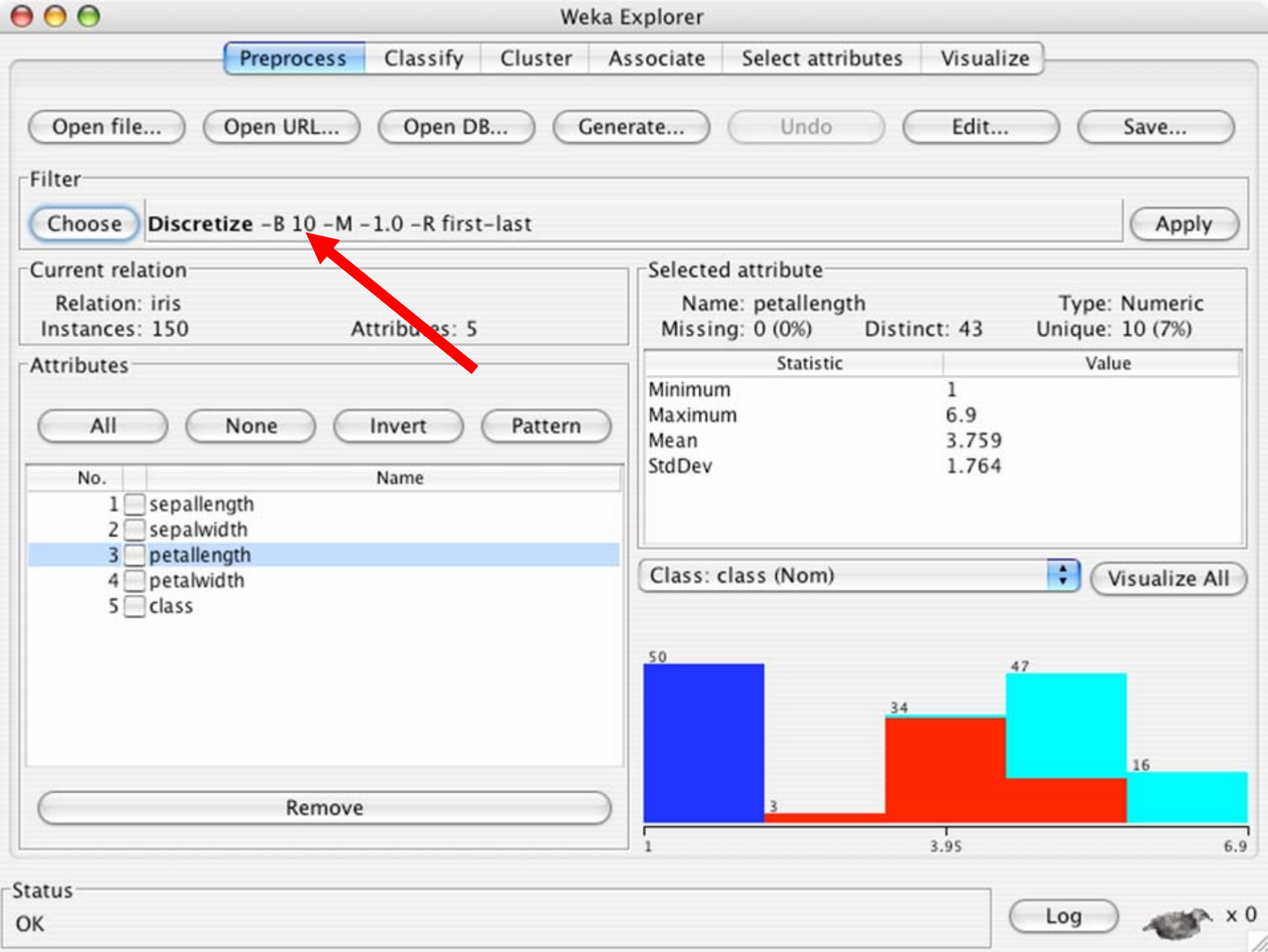
Class: class (Nom) | Visualize All

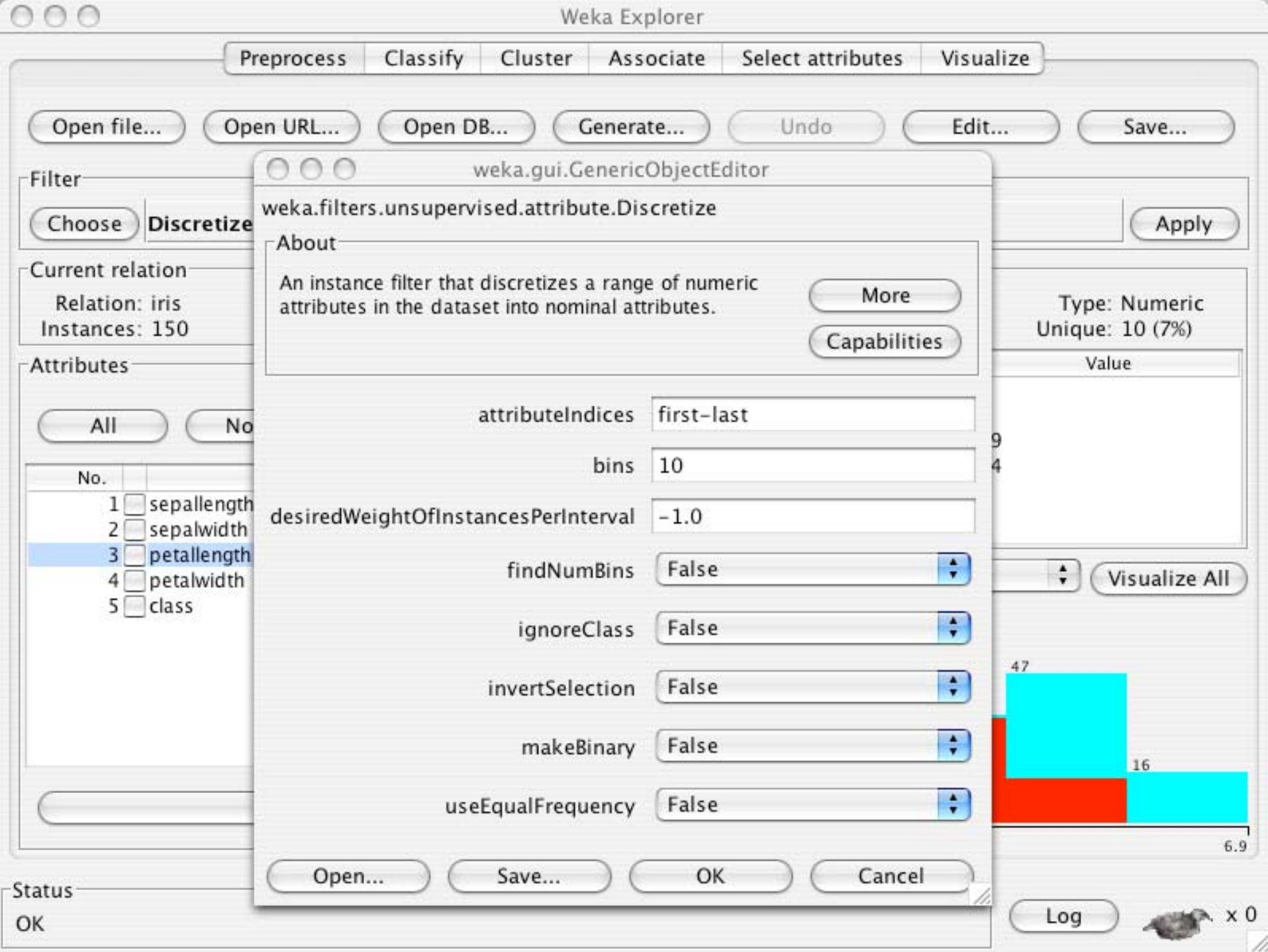


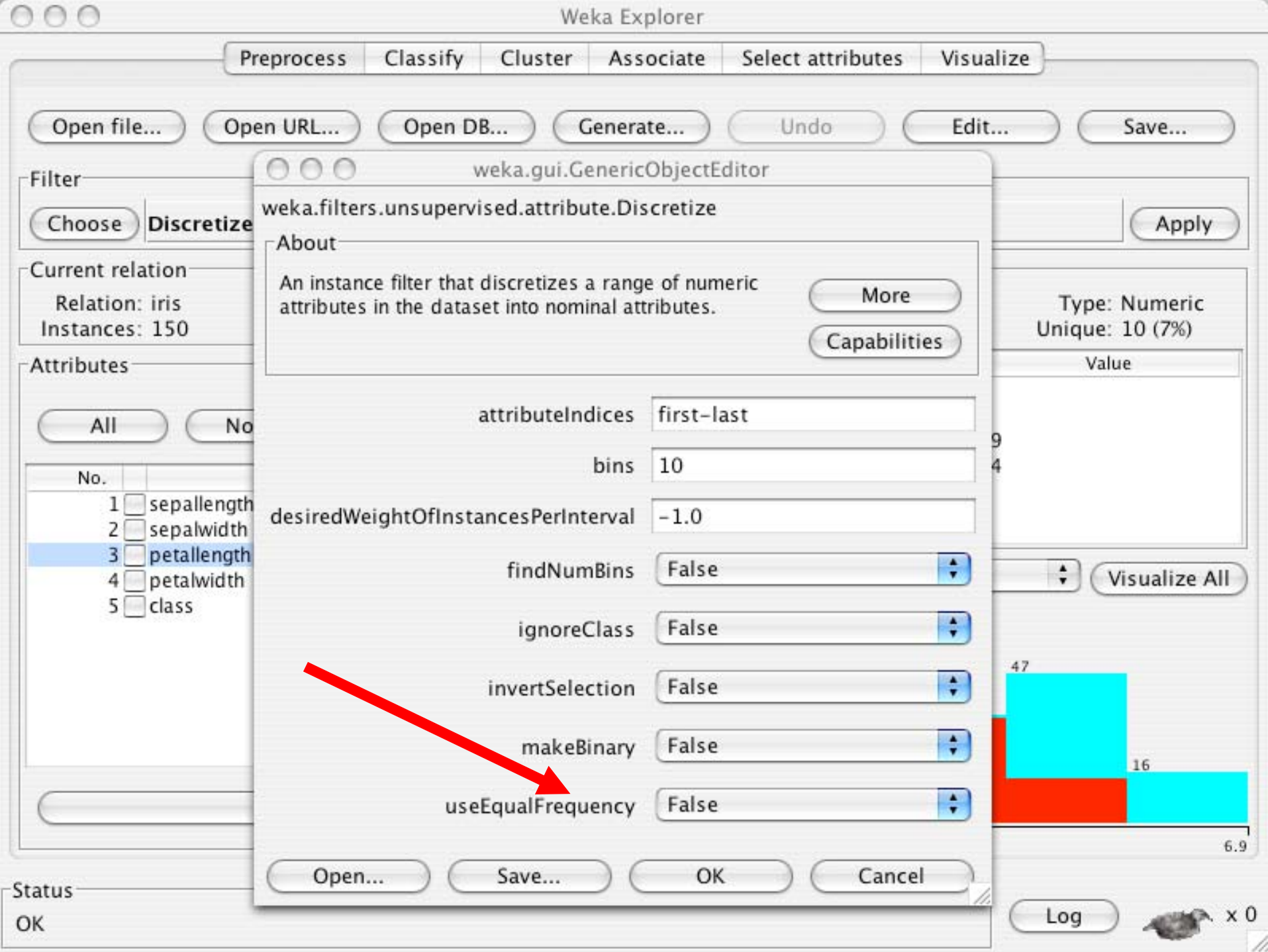
Status

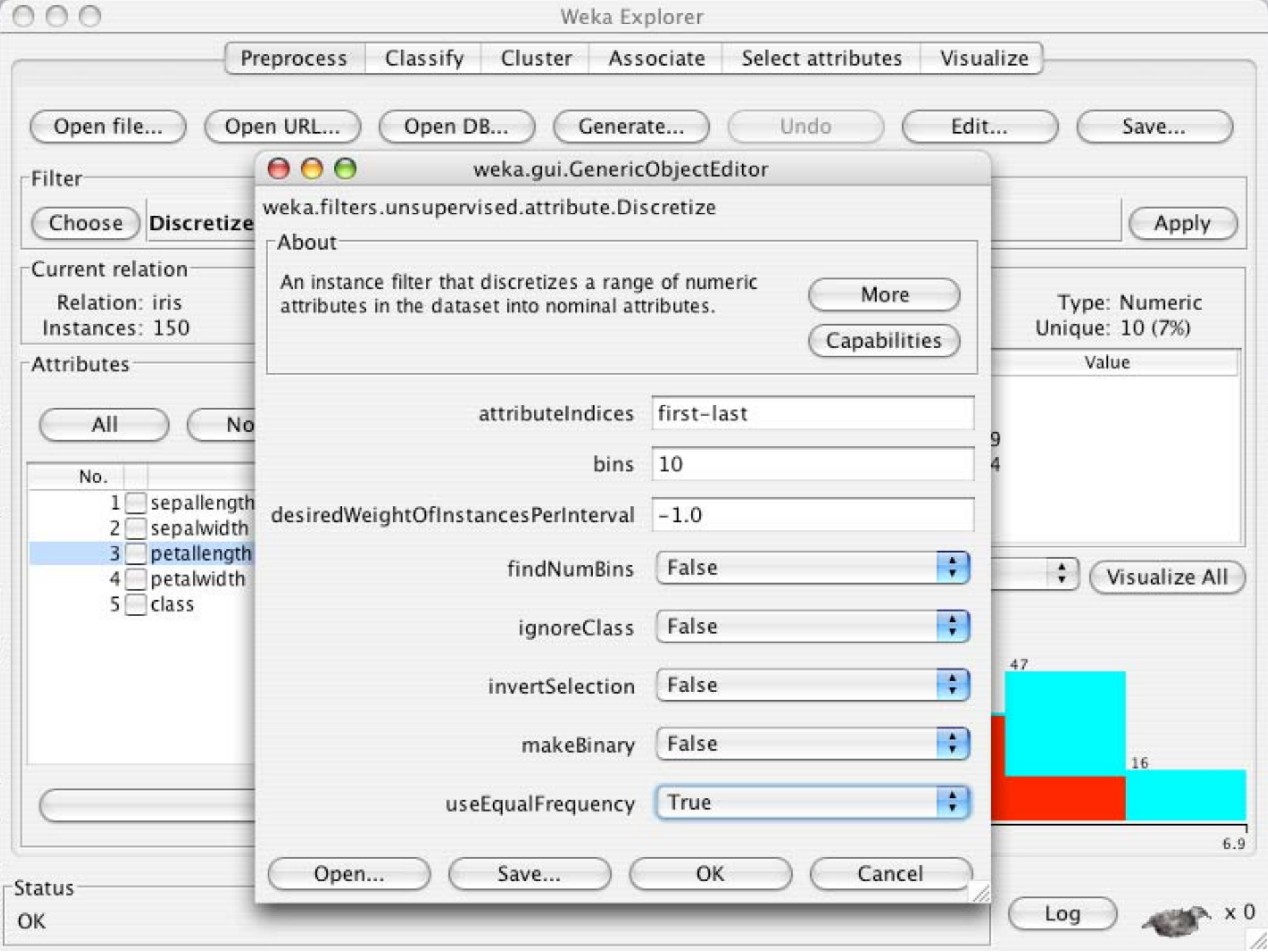
OK

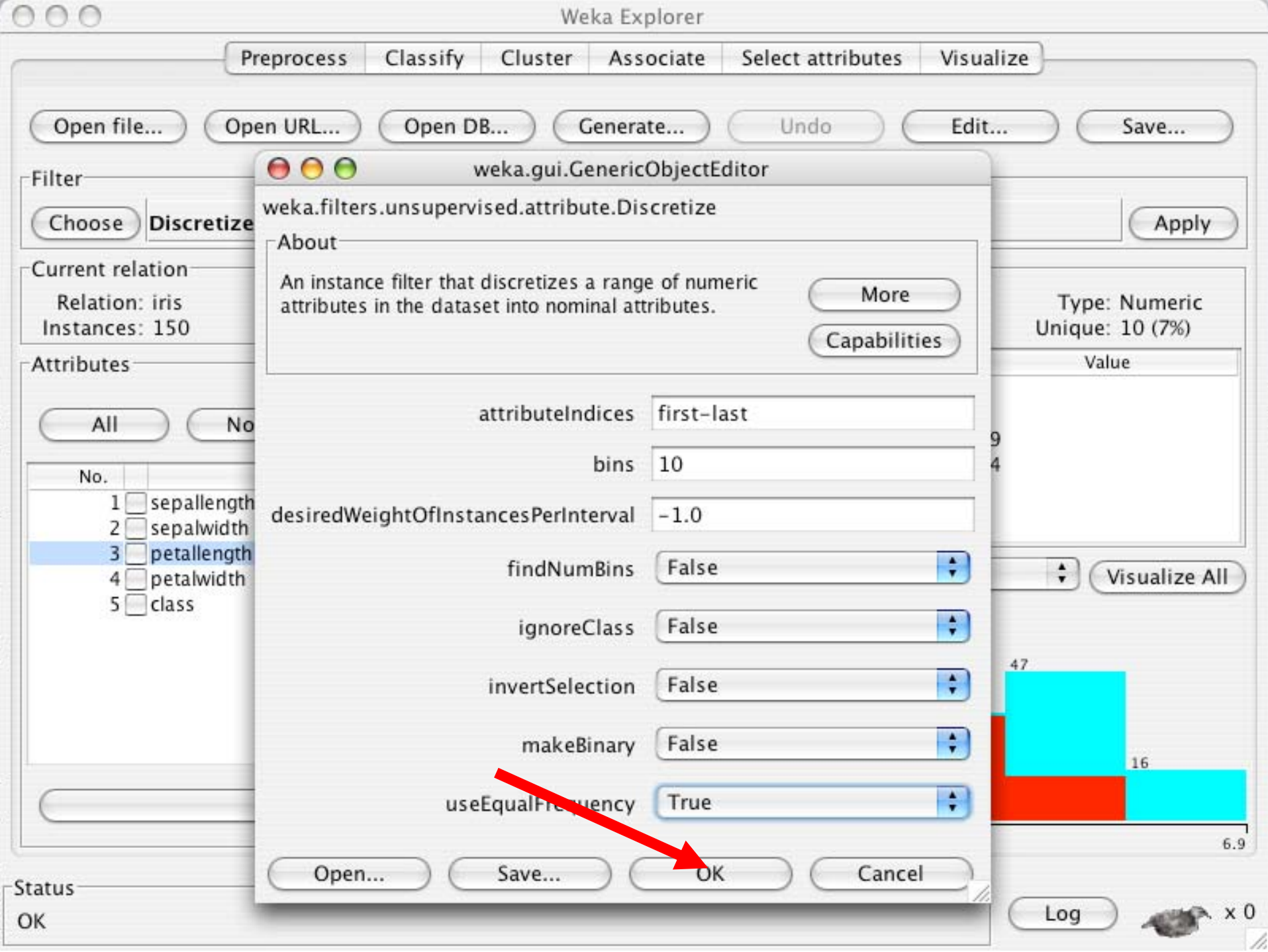
Log | x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize -B 10 -M -1.0 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum 1
Maximum 6.9
Mean 3.759
StdDev 1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
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
Discretize
Apply
Current relation
Relation: iris
Instances: 150
Attributes
All
No.
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Type: Numeric
Unique: 10 (7%)
Value
Visualize All
47
16
6.9
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices
first-last
bins
10
desiredWeightOfInstancesPerInterval
-1.0
findNumBins
False
ignoreClass
False
invertSelection
False
makeBinary
False
useEqualFrequency
False
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize
Apply
Current relation
Relation: iris
Instances: 150
Attributes
All
No.
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Type: Numeric
Unique: 10 (7%)
Value
Visualize All
47
16
6.9
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices
first-last
bins
10
desiredWeightOfInstancesPerInterval
-1.0
findNumBins
False
ignoreClass
False
invertSelection
False
makeBinary
False
useEqualFrequency
False
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize
Apply
Current relation
Relation: iris
Instances: 150
Attributes
All
No.
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Type: Numeric
Unique: 10 (7%)
Value
Visualize All
47
16
6.9
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices
first-last
bins
10
desiredWeightOfInstancesPerInterval
-1.0
findNumBins
False
ignoreClass
False
invertSelection
False
makeBinary
False
useEqualFrequency
True
Open...
Save...
OK
Cancel


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
Discretize
Apply
Current relation
Relation: iris
Instances: 150
Attributes
All
No.
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Type: Numeric
Unique: 10 (7%)
Value
Visualize All
47
16
6.9
Status
OK
Log
x 0
weka.gui.GenericObjectEditor
weka.filters.unsupervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
Capabilities
attributeIndices
first-last
bins
10
desiredWeightOfInstancesPerInterval
-1.0
findNumBins
False
ignoreClass
False
invertSelection
False
makeBinary
False
useEqualFrequency
True
Open...
Save...
OK
Cancel

Weka Explorer

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

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

Filter

Choose | **Discretize** -F -B 10 -M -1.0 -R first-last | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

All | None | Invert | Pattern

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

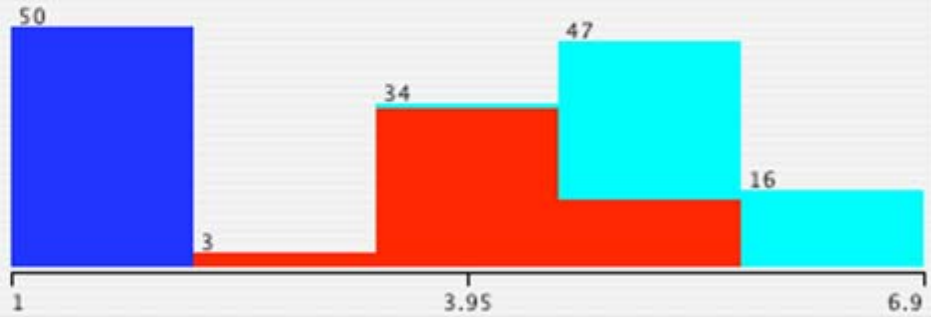
Remove

Selected attribute

Name: petallength Type: Numeric
Missing: 0 (0%) Distinct: 43 Unique: 10 (7%)

Statistic	Value
Minimum	1
Maximum	6.9
Mean	3.759
StdDev	1.764

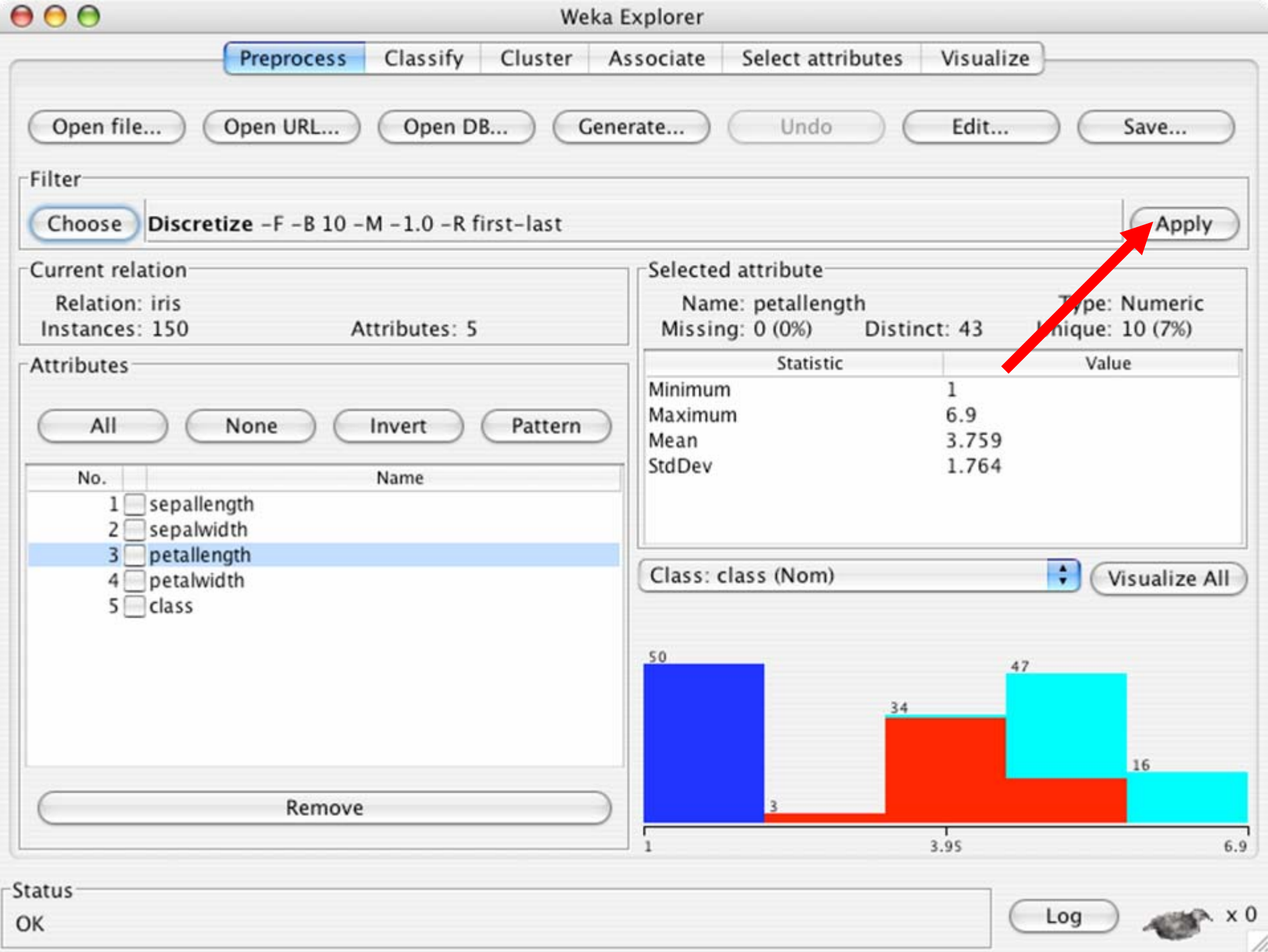
Class: class (Nom) | Visualize All



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
Discretize -F -B 10 -M -1.0 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
All
None
Invert
Pattern
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Remove
Selected attribute
Name: petallength
Type: Numeric
Missing: 0 (0%)
Distinct: 43
Unique: 10 (7%)
Statistic
Value
Minimum 1
Maximum 6.9
Mean 3.759
StdDev 1.764
Class: class (Nom)
Visualize All
50
3
34
47
16
1
3.95
6.9
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 | **Discretize** -F -B 10 -M -1.0 -R first-last | Apply

Current relation

Relation: iris-weka.filters.unsupervised.attribute.Dis...
Instances: 150 Attributes: 5

Attributes

All | None | Invert | Pattern

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

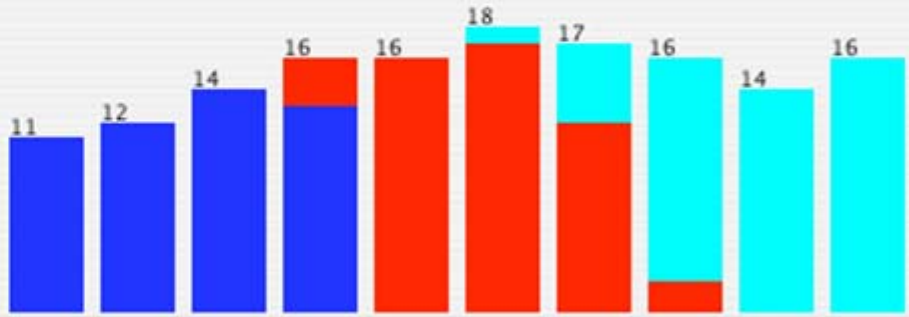
Remove

Selected attribute

Name: petallength Type: Nominal
Missing: 0 (0%) Distinct: 10 Unique: 0 (0%)

Label	Count
'(-inf-1.35]'	11
'(1.35-1.45]'	12
'(1.45-1.55]'	14
'(1.55-3.4]'	16
'(3.4-4.15]'	16
'(4.15-4.55]'	18
'(4.55-4.95]'	17

Class: class (Nom) | Visualize All



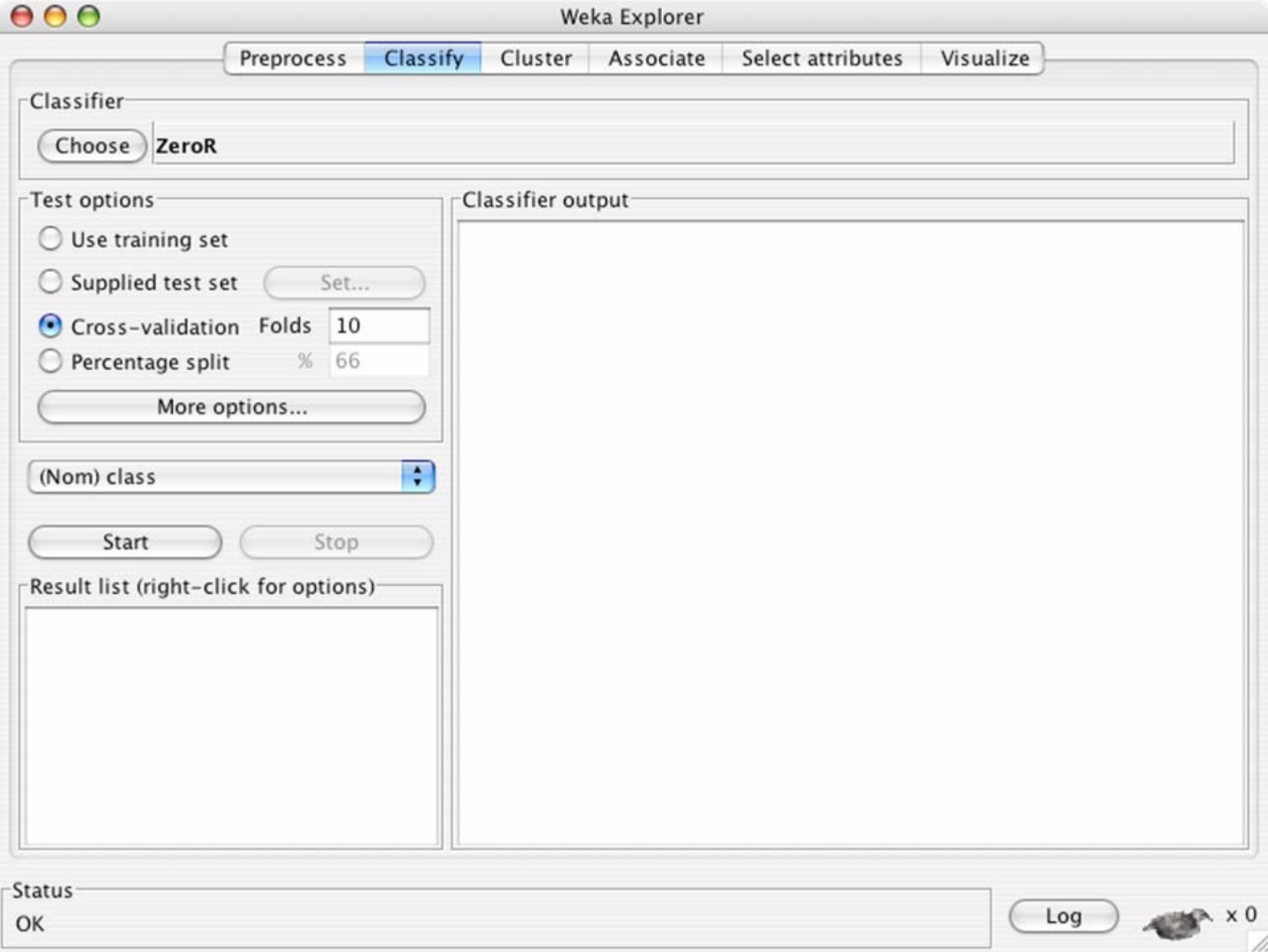
Status

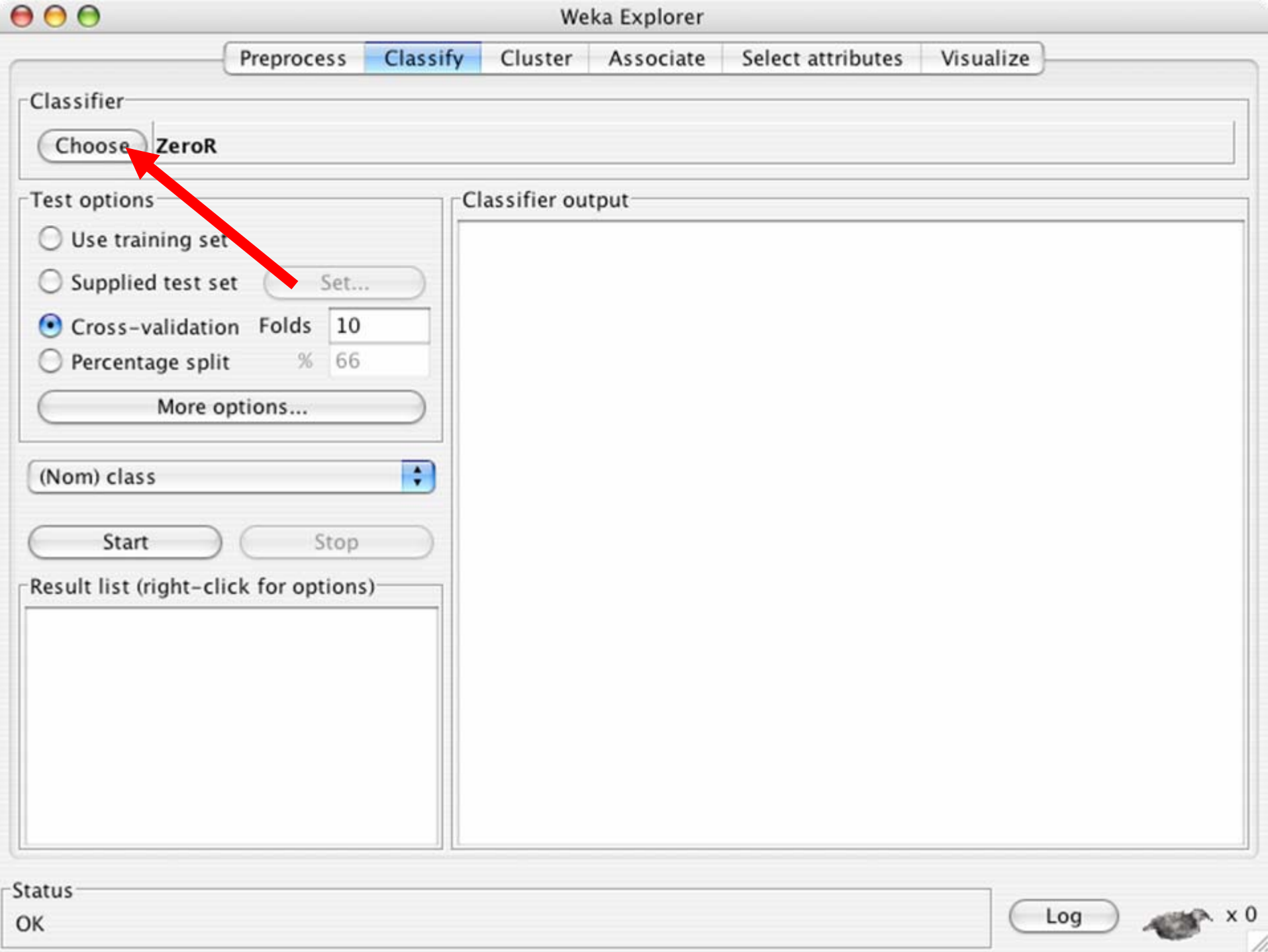
OK

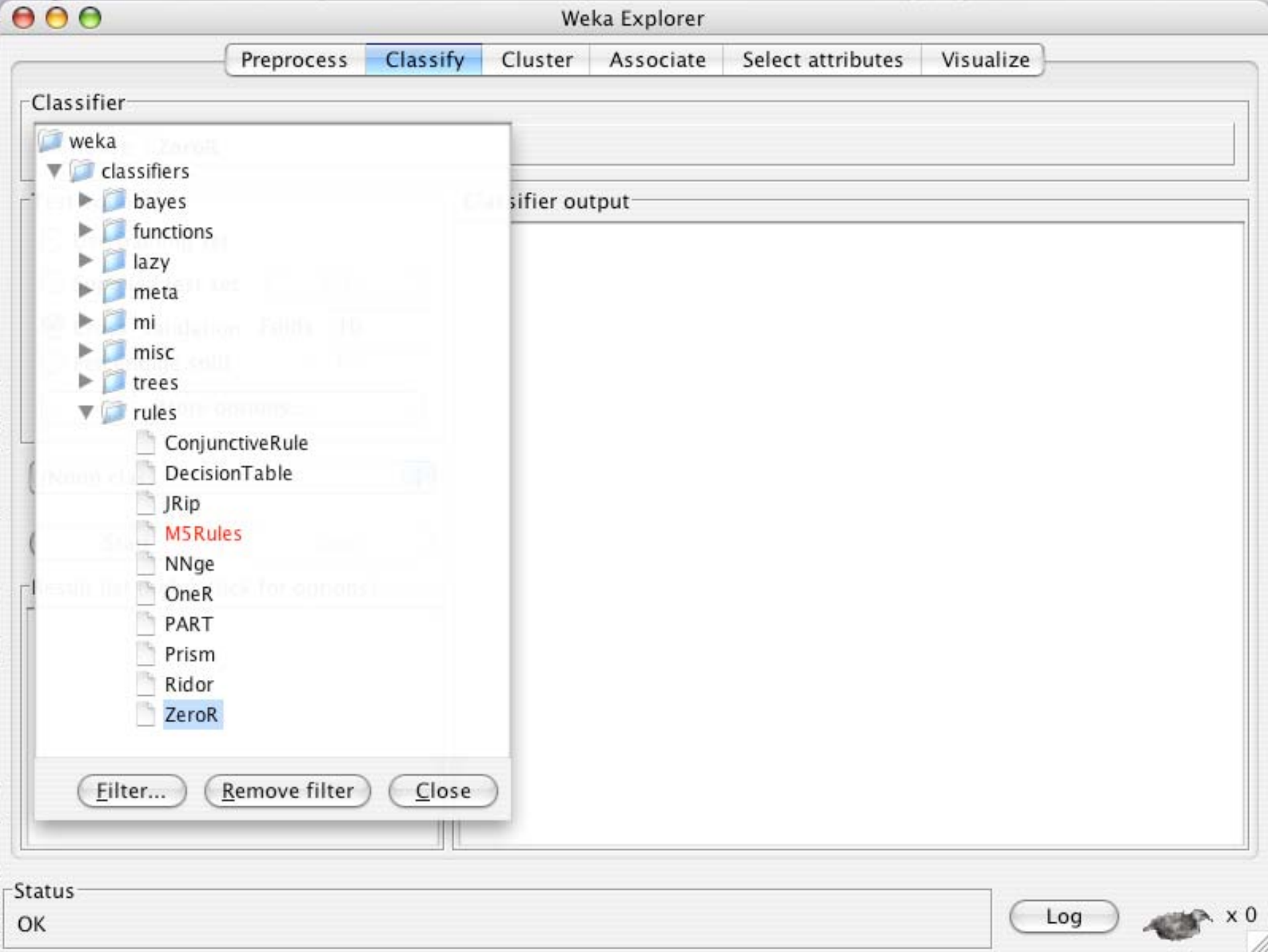
Log | x 0

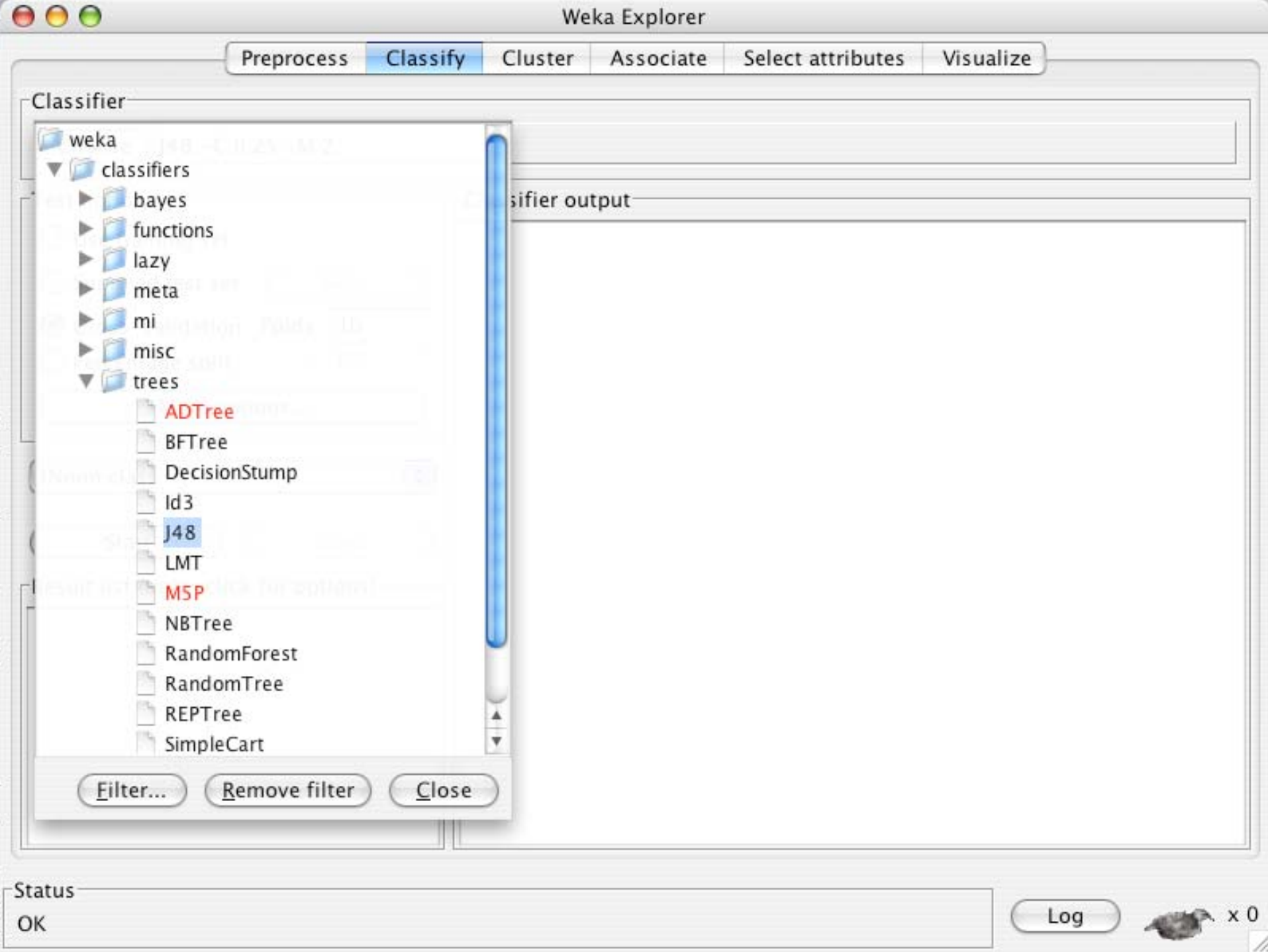
Explorer: building “classifiers”

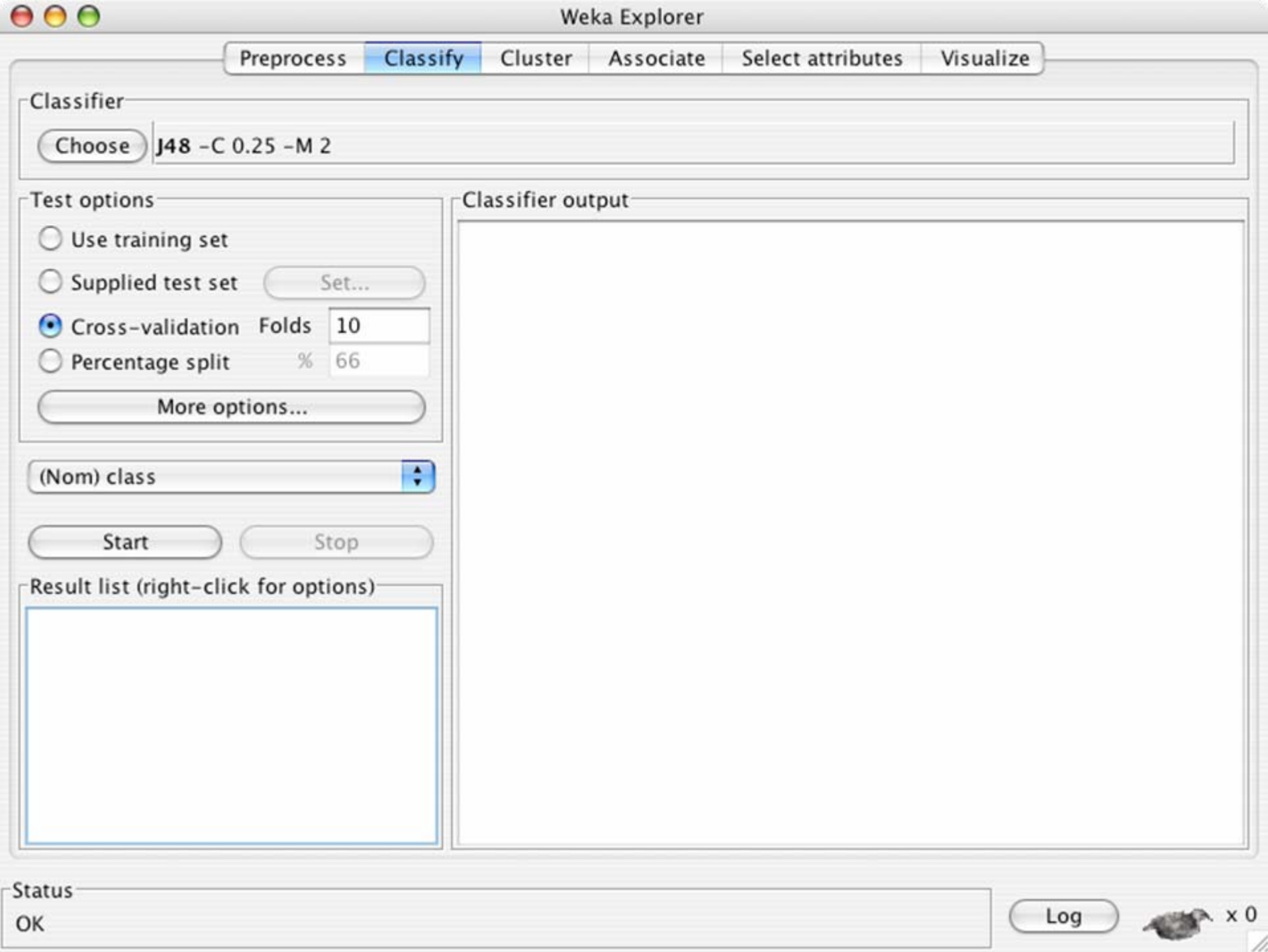
- Classifiers in WEKA are models for predicting nominal or numeric quantities
- Implemented learning schemes include:
  - Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes’ nets, …
- “Meta”-classifiers include:
  - Bagging, boosting, stacking, error-correcting output codes, locally weighted learning, …

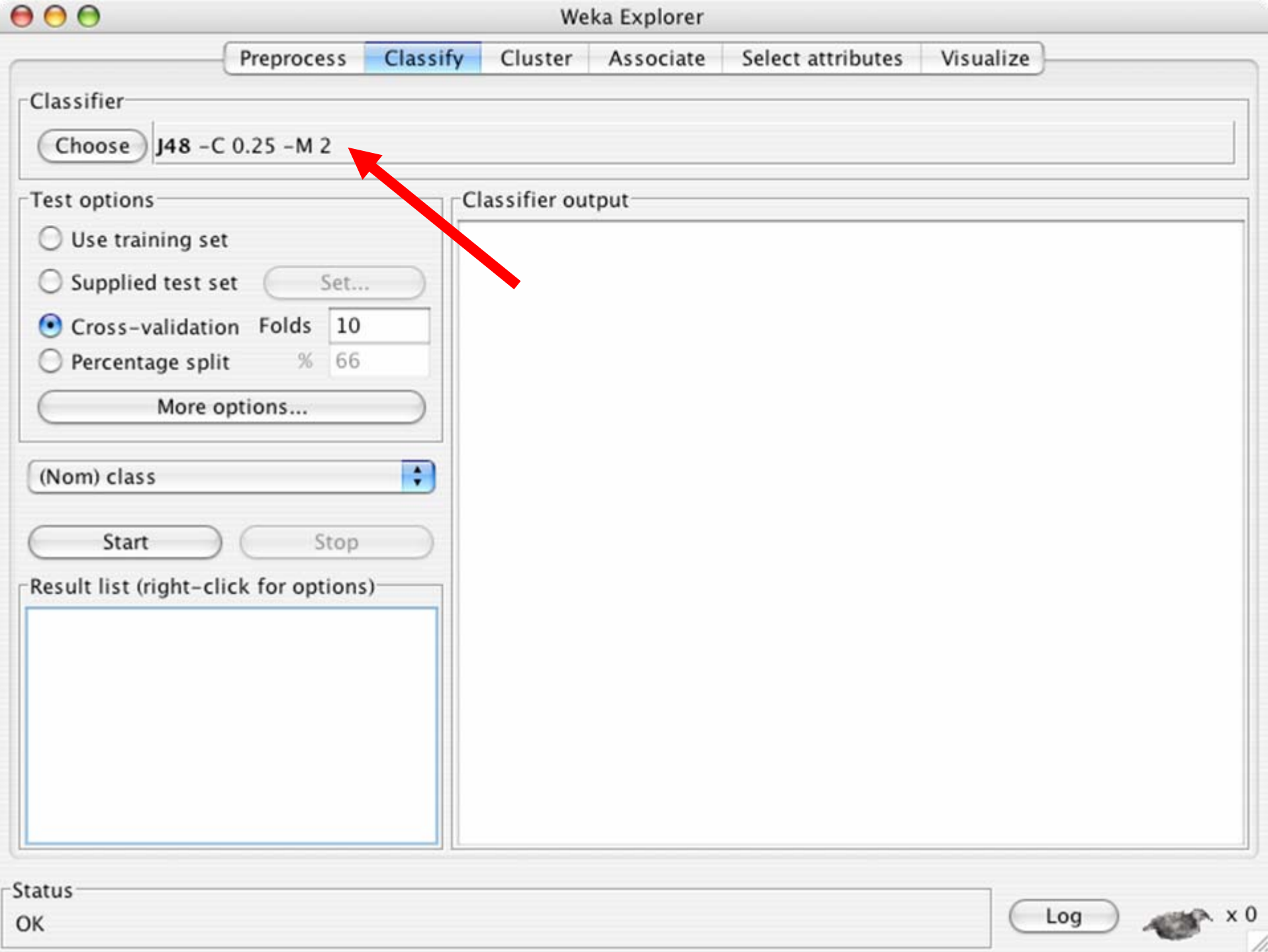

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

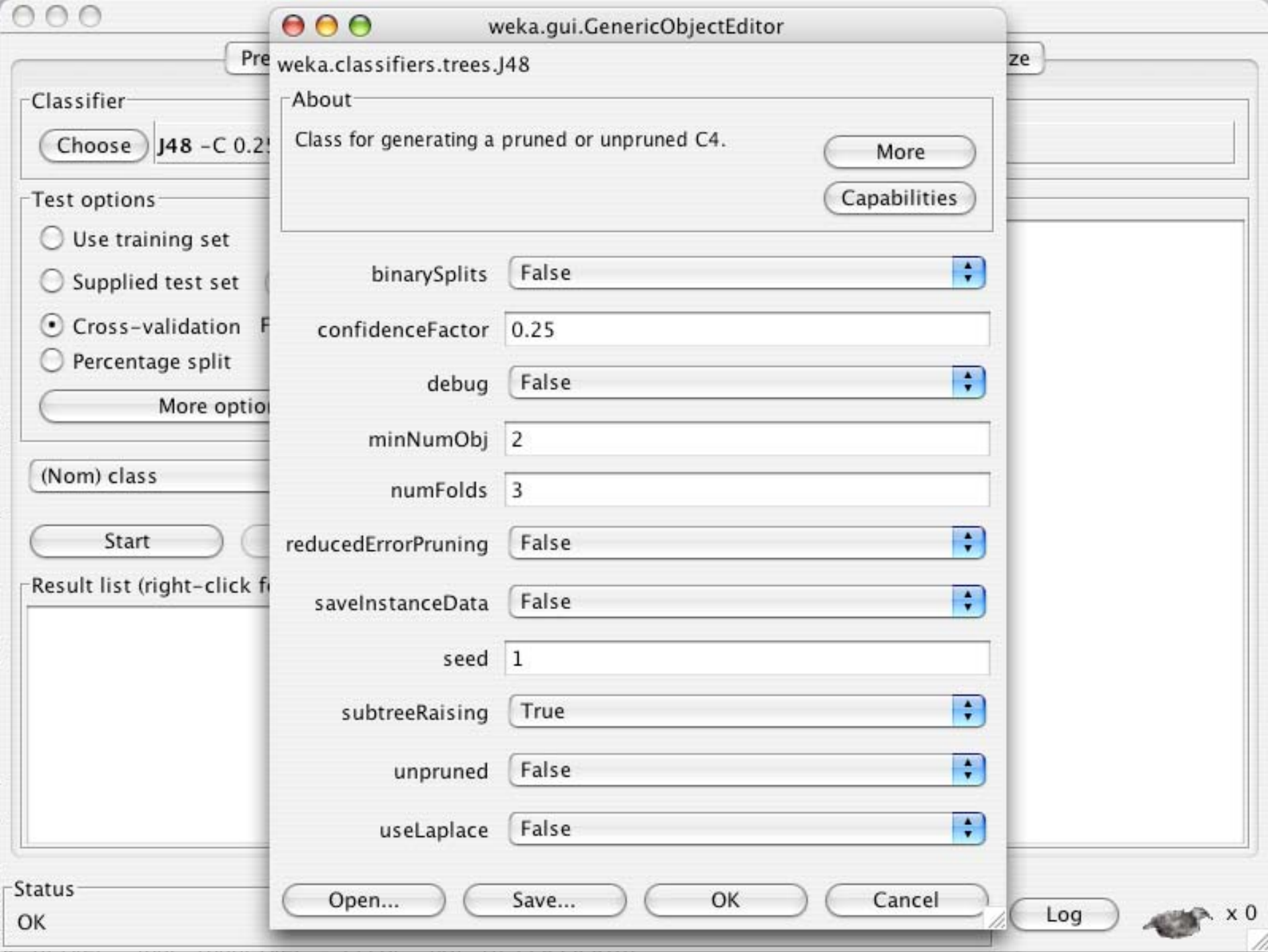

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

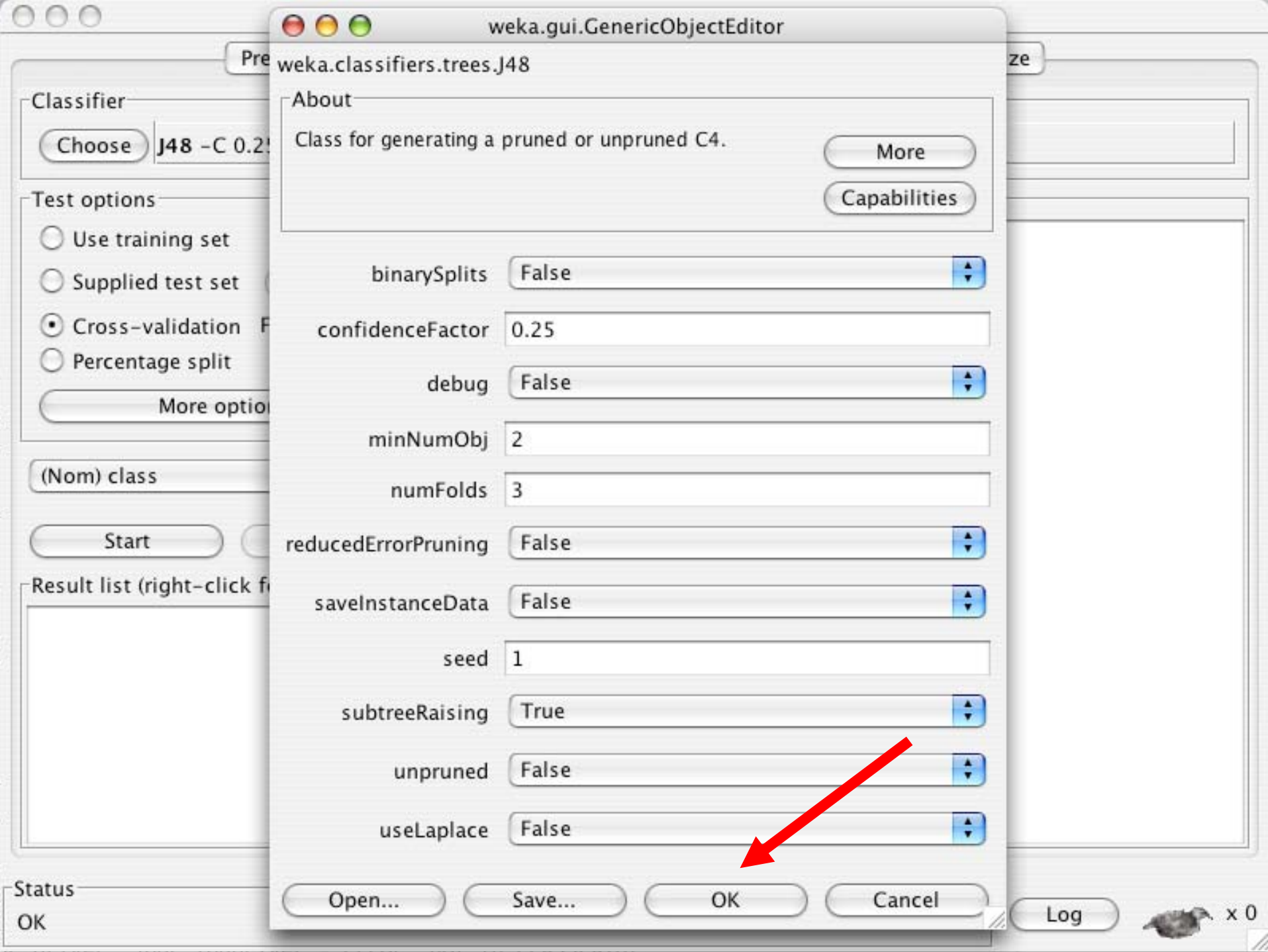

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
mi
misc
trees
rules
ConjunctiveRule
DecisionTable
JRip
M5Rules
NNge
OneR
PART
Prism
Ridor
ZeroR
Filter...
Remove filter
Close
Classifier output
Status
OK
Log
x 0

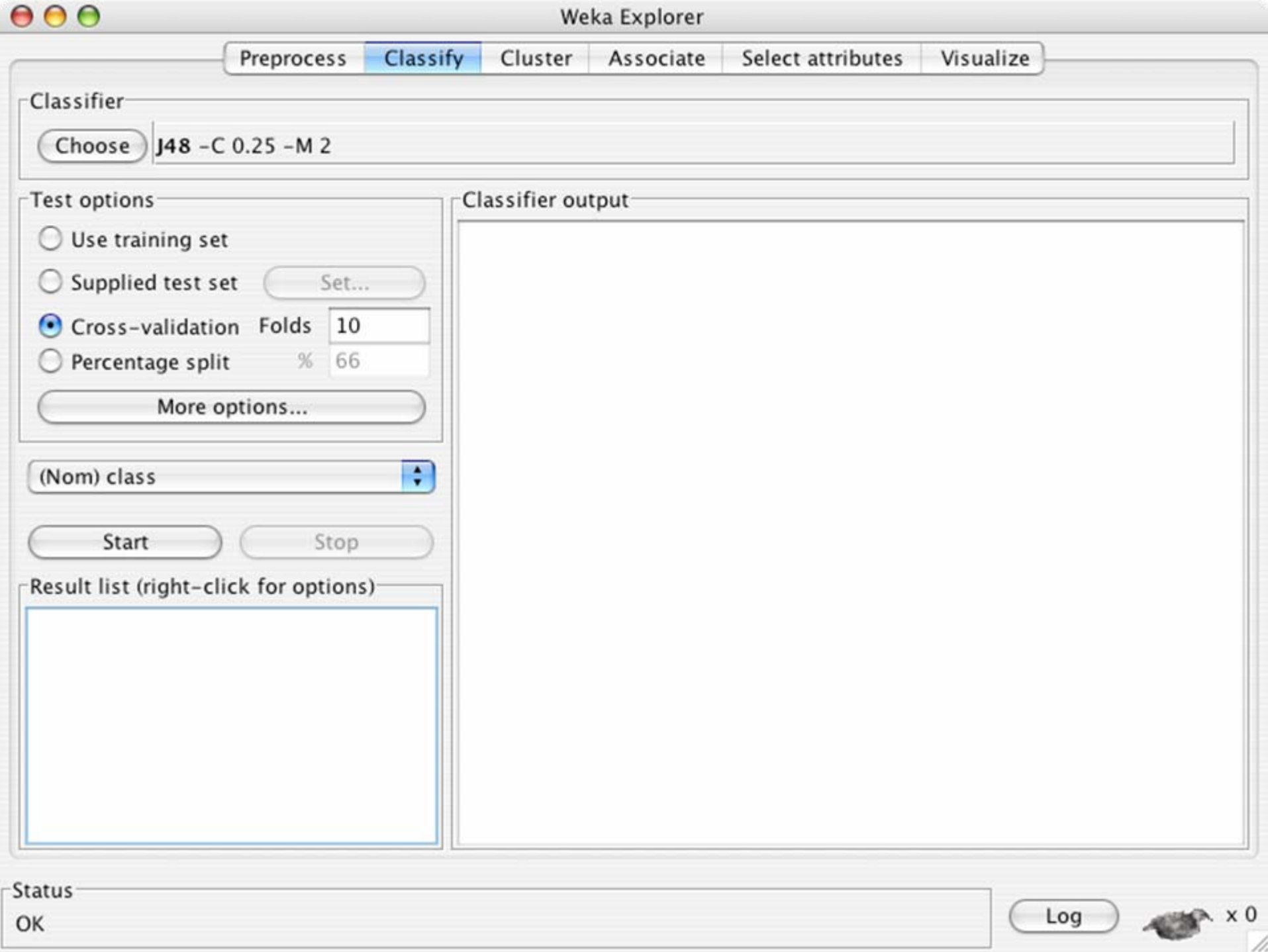

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
mi
misc
trees
ADTree
BFTree
DecisionStump
Id3
J48
LMT
M5P
NBTree
RandomForest
RandomTree
REPTree
SimpleCart
Filter...
Remove filter
Close
sifier output
Status
OK
Log
x 0

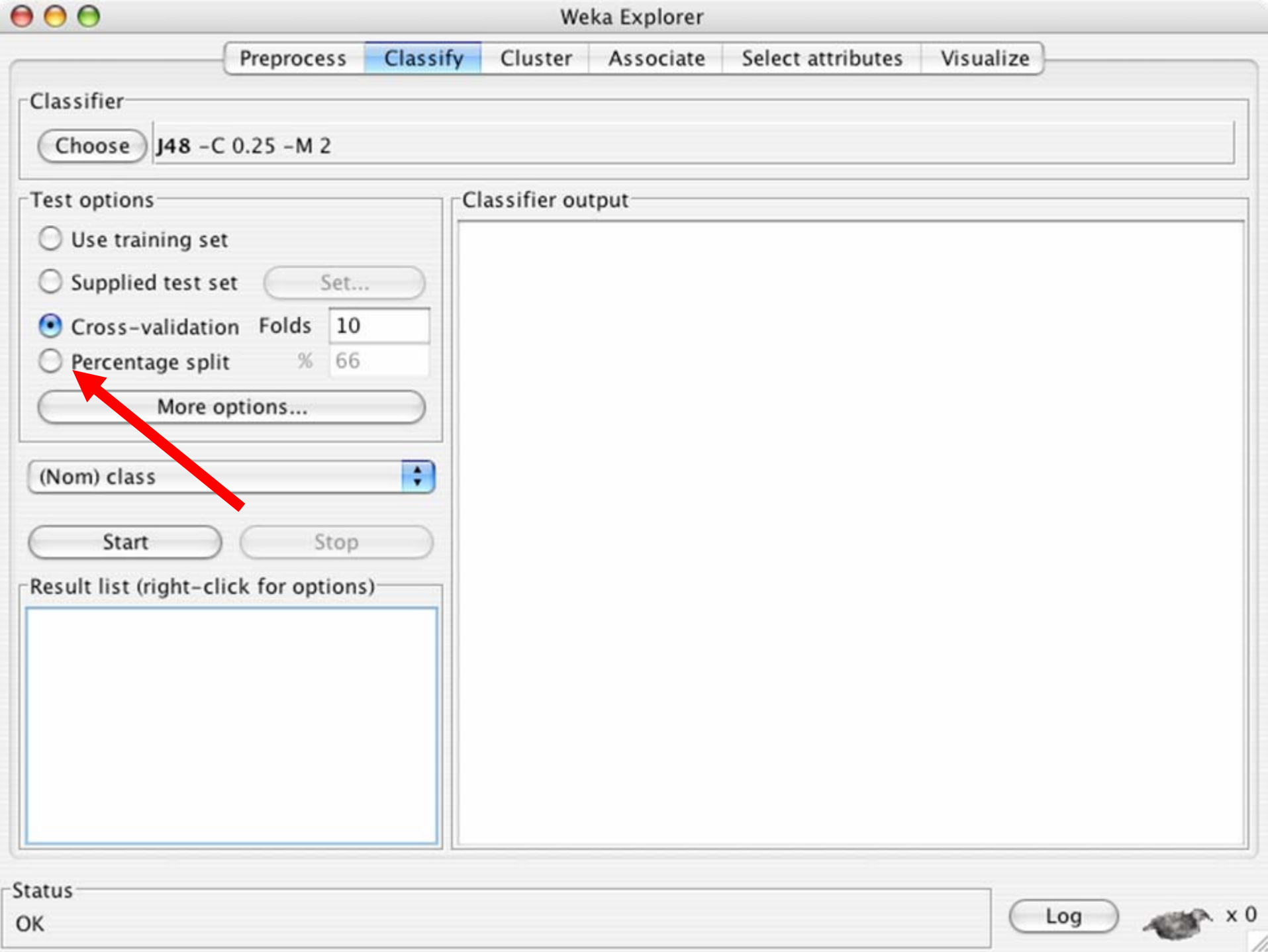

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

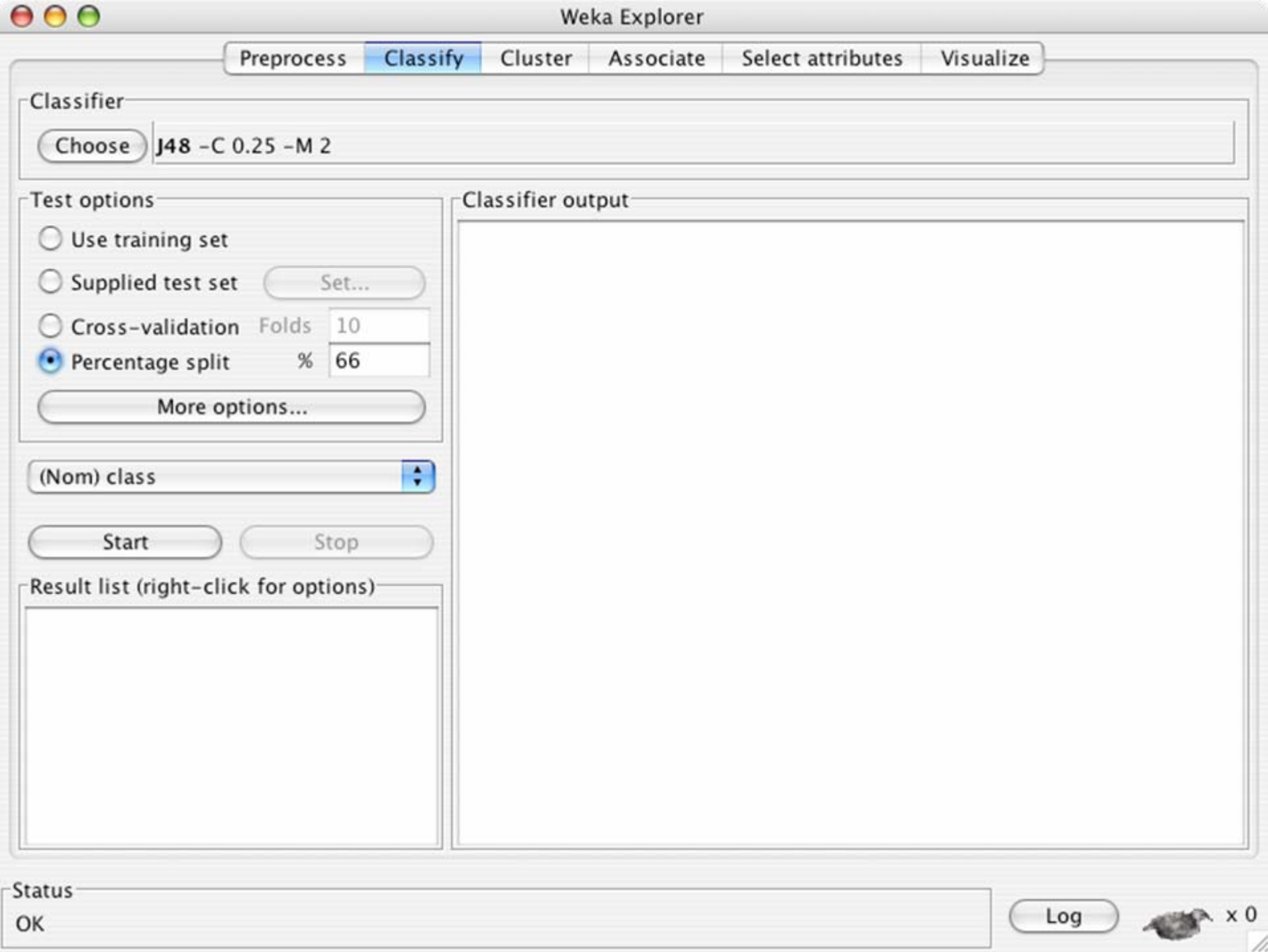

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

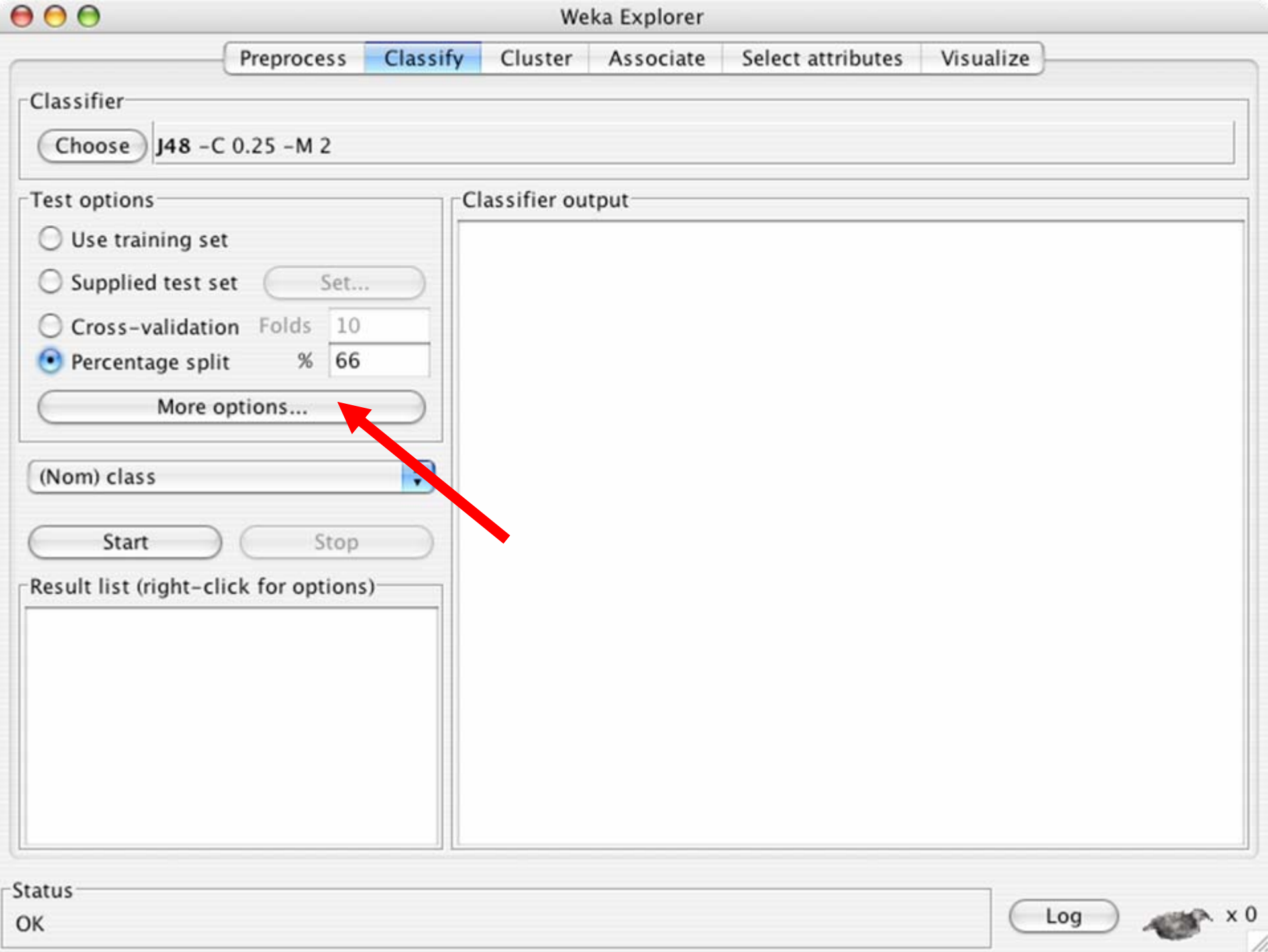

weka.gui.GenericObjectEditor
weka.classifiers.trees.J48
About
Class for generating a pruned or unpruned C4.
More
Capabilities
binarySplits
False
confidenceFactor
0.25
debug
False
minNumObj
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
seed
1
subtreeRaising
True
unpruned
False
useLaplace
False
Open...
Save...
OK
Cancel
Classifier
Choose
J48 -C 0.2
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optio
(Nom) class
Start
Result list (right-click f
Status
OK
Log
x 0

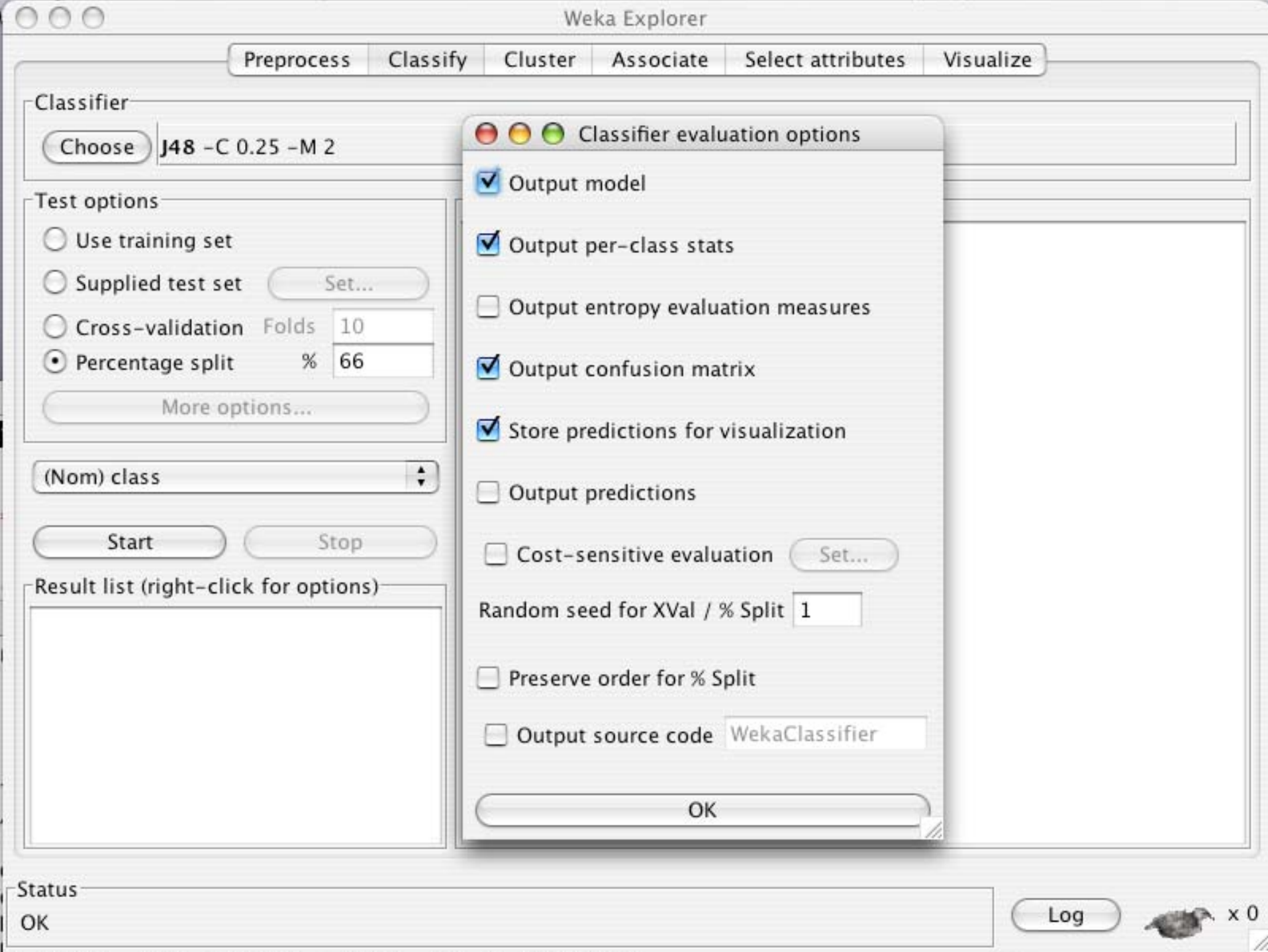

weka.gui.GenericObjectEditor
weka.classifiers.trees.J48
About
Class for generating a pruned or unpruned C4.
More
Capabilities
binarySplits
False
confidenceFactor
0.25
debug
False
minNumObj
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
seed
1
subtreeRaising
True
unpruned
False
useLaplace
False
Open...
Save...
OK
Cancel
Classifier
Choose
J48 -C 0.2
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optio
(Nom) class
Start
Result list (right-click f
Status
OK
Log
x 0

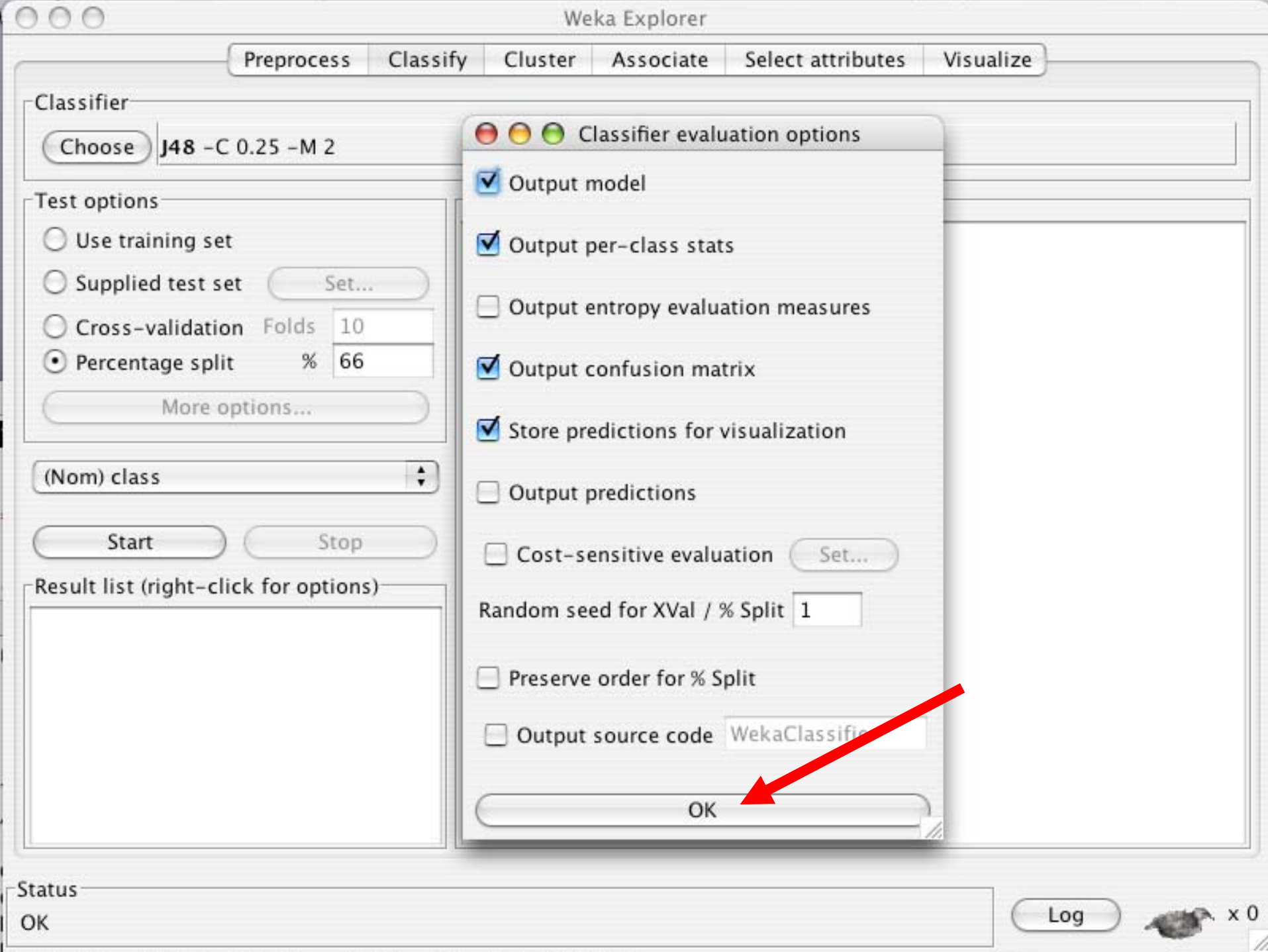

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

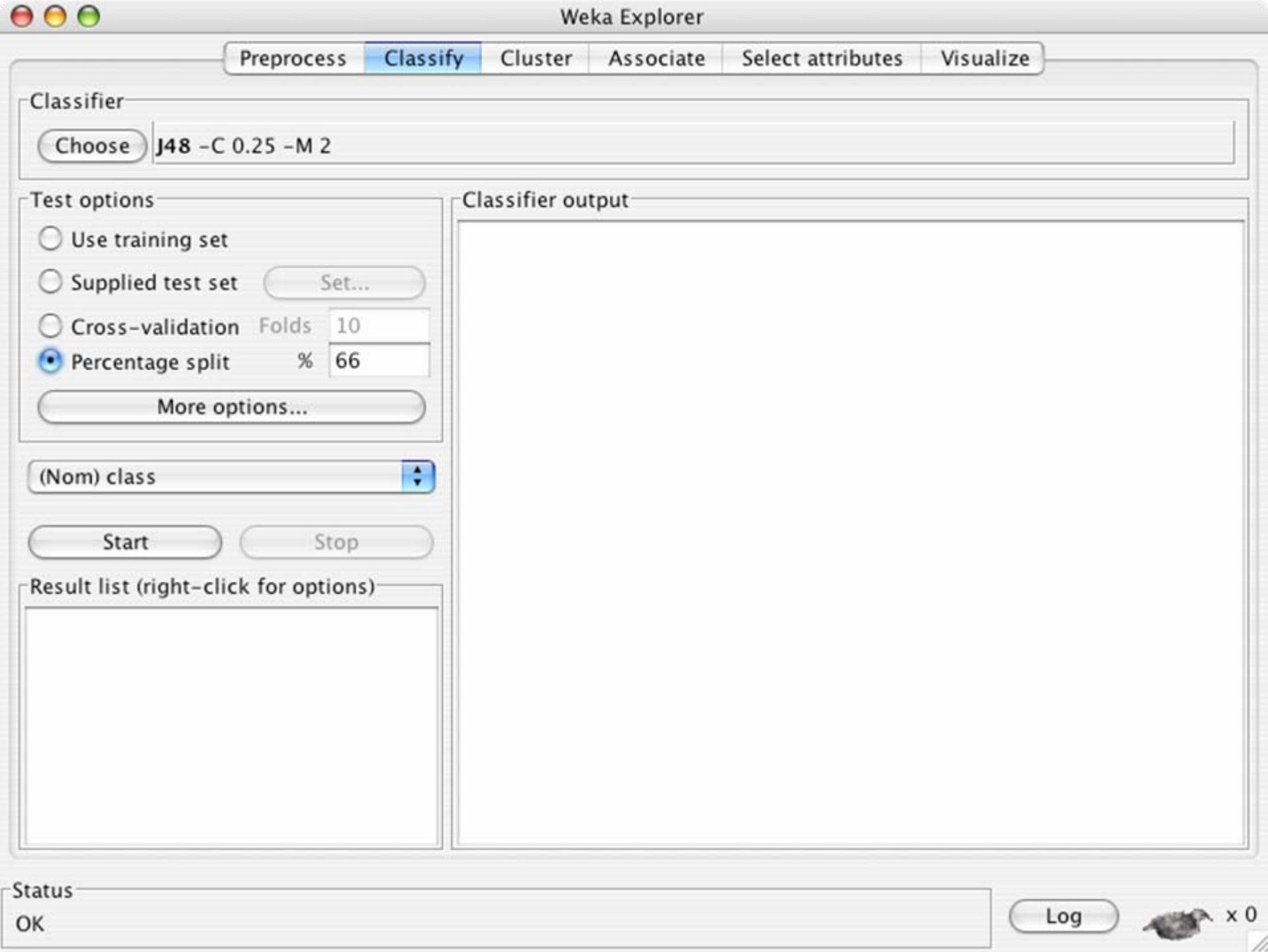

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

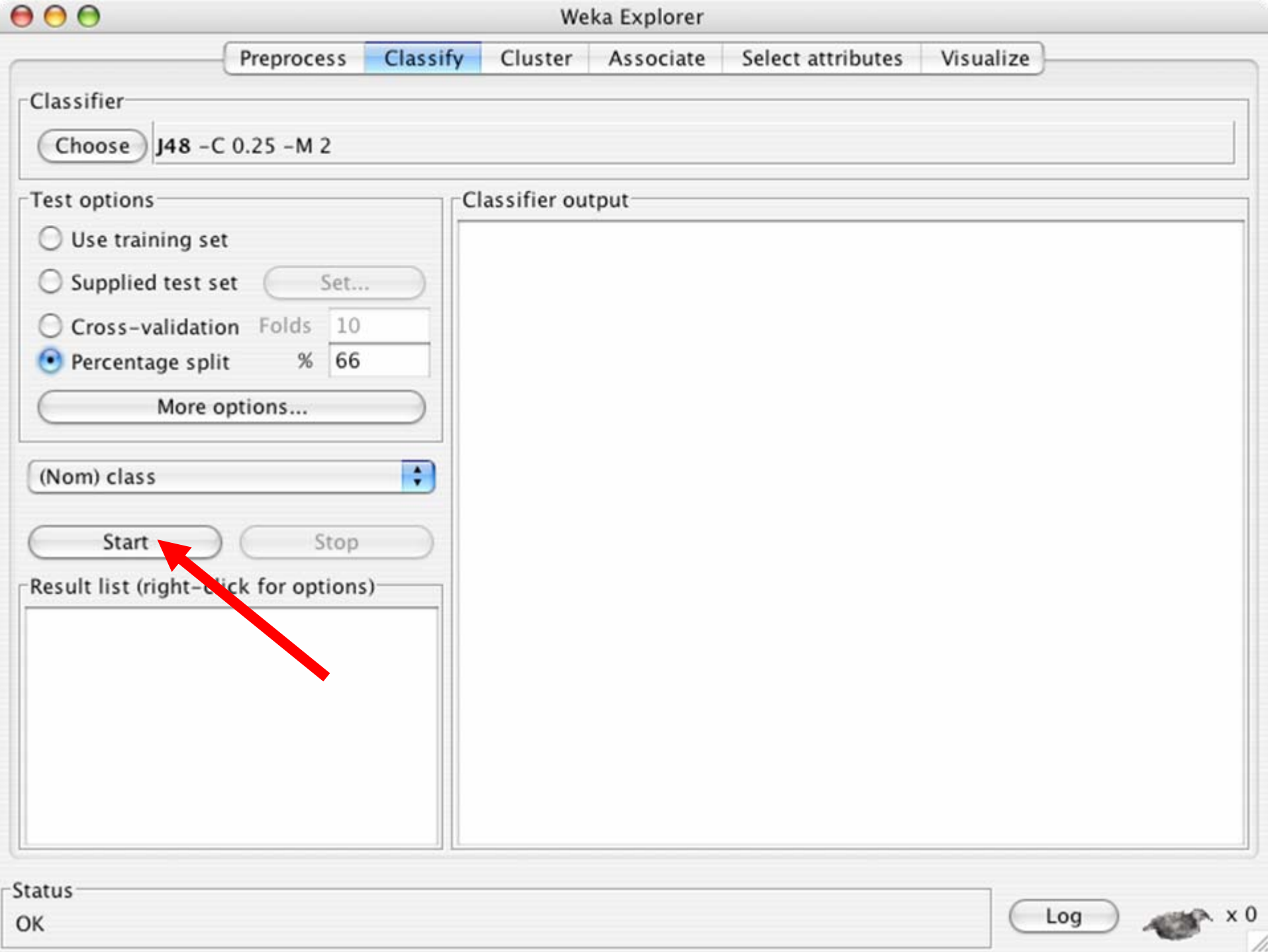

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

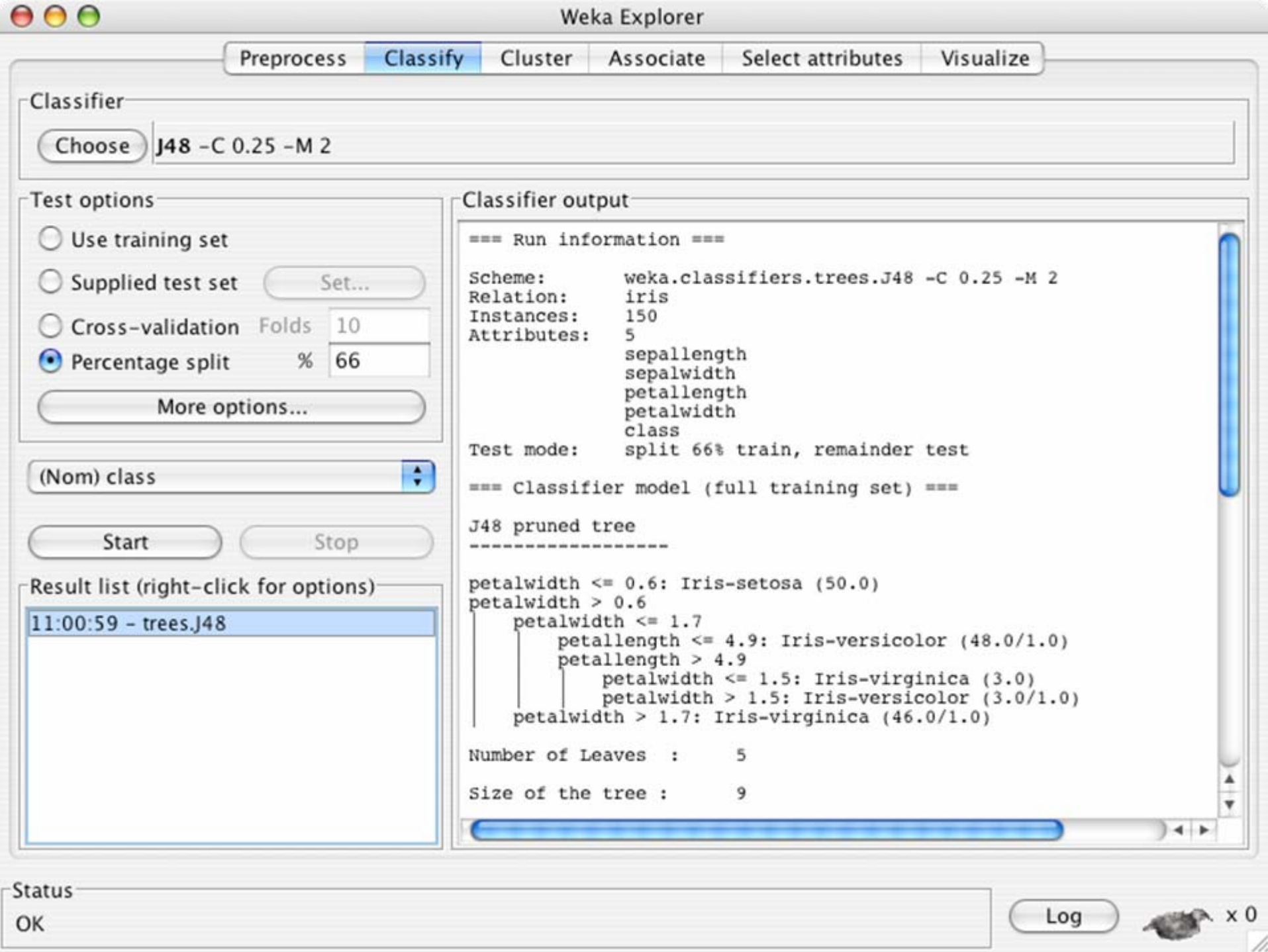

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

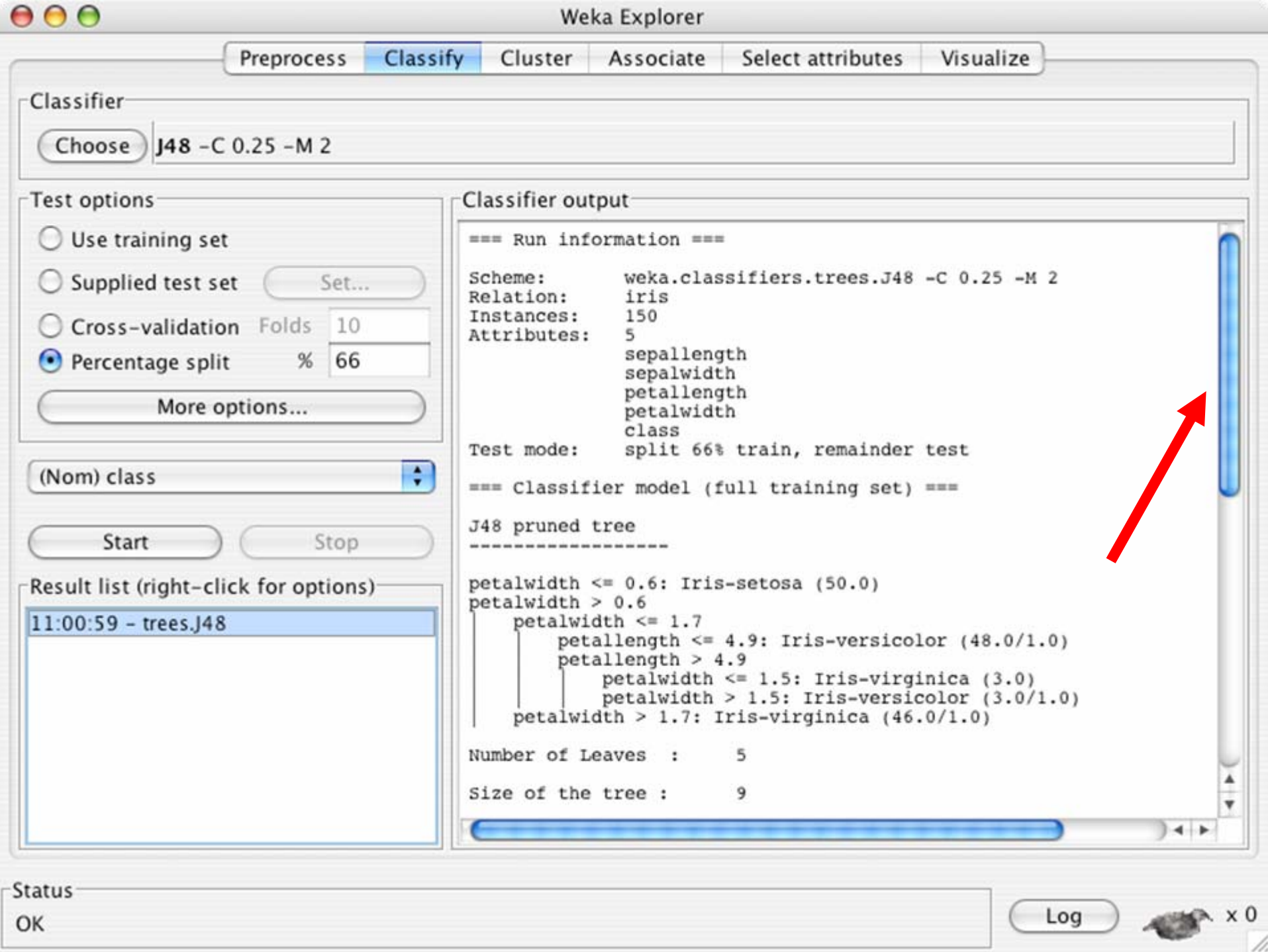

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
Classifier evaluation options
Output model
Output per-class stats
Output entropy evaluation measures
Output confusion matrix
Store predictions for visualization
Output predictions
Cost-sensitive evaluation
Set...
Random seed for XVal / % Split
1
Preserve order for % Split
Output source code
WekaClassifier
OK
Status
OK
Log
x 0

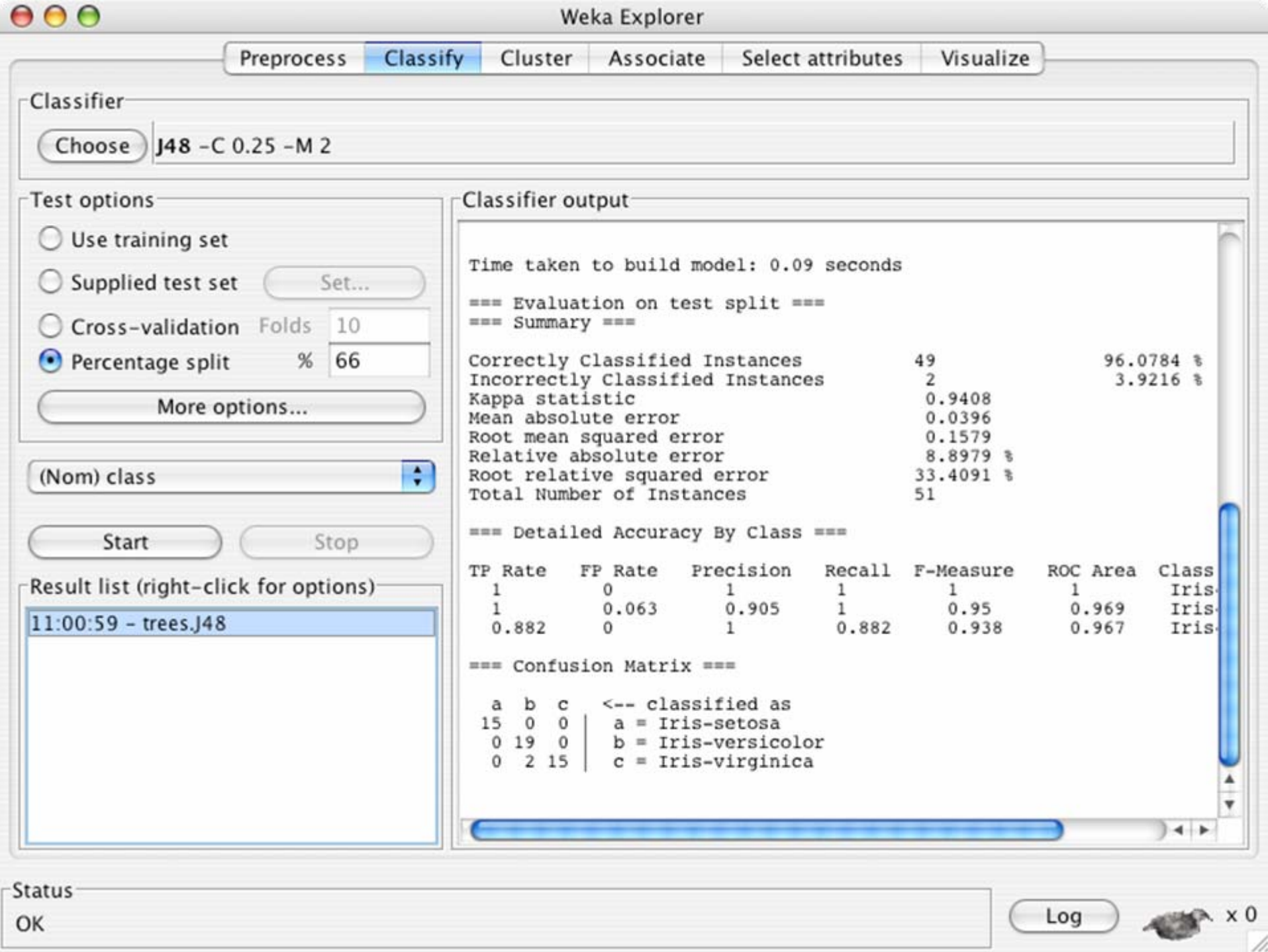

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
Classifier evaluation options
Output model
Output per-class stats
Output entropy evaluation measures
Output confusion matrix
Store predictions for visualization
Output predictions
Cost-sensitive evaluation
Set...
Random seed for XVal / % Split
1
Preserve order for % Split
Output source code
WekaClassifi
OK
Status
OK
Log
x 0

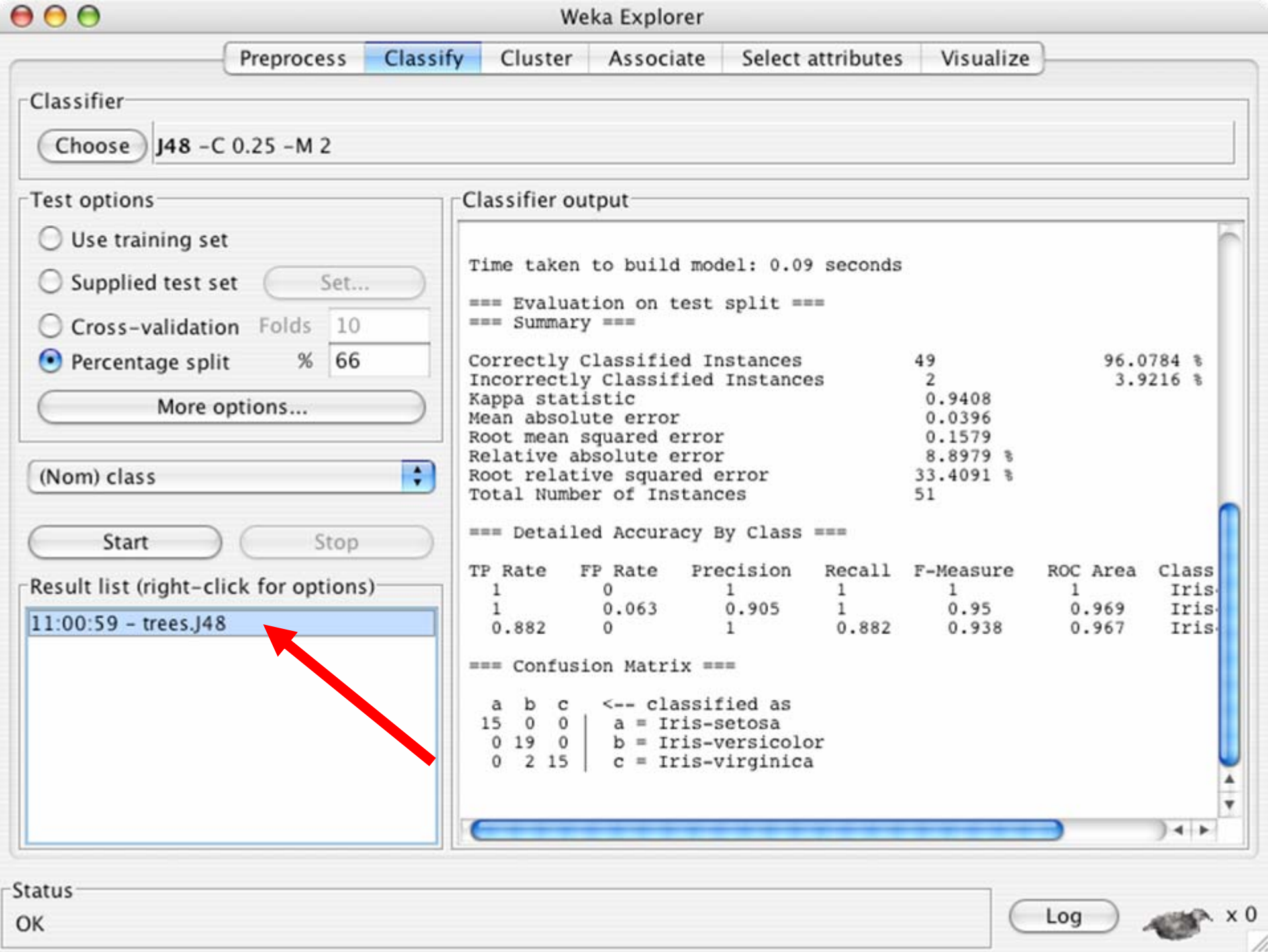

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

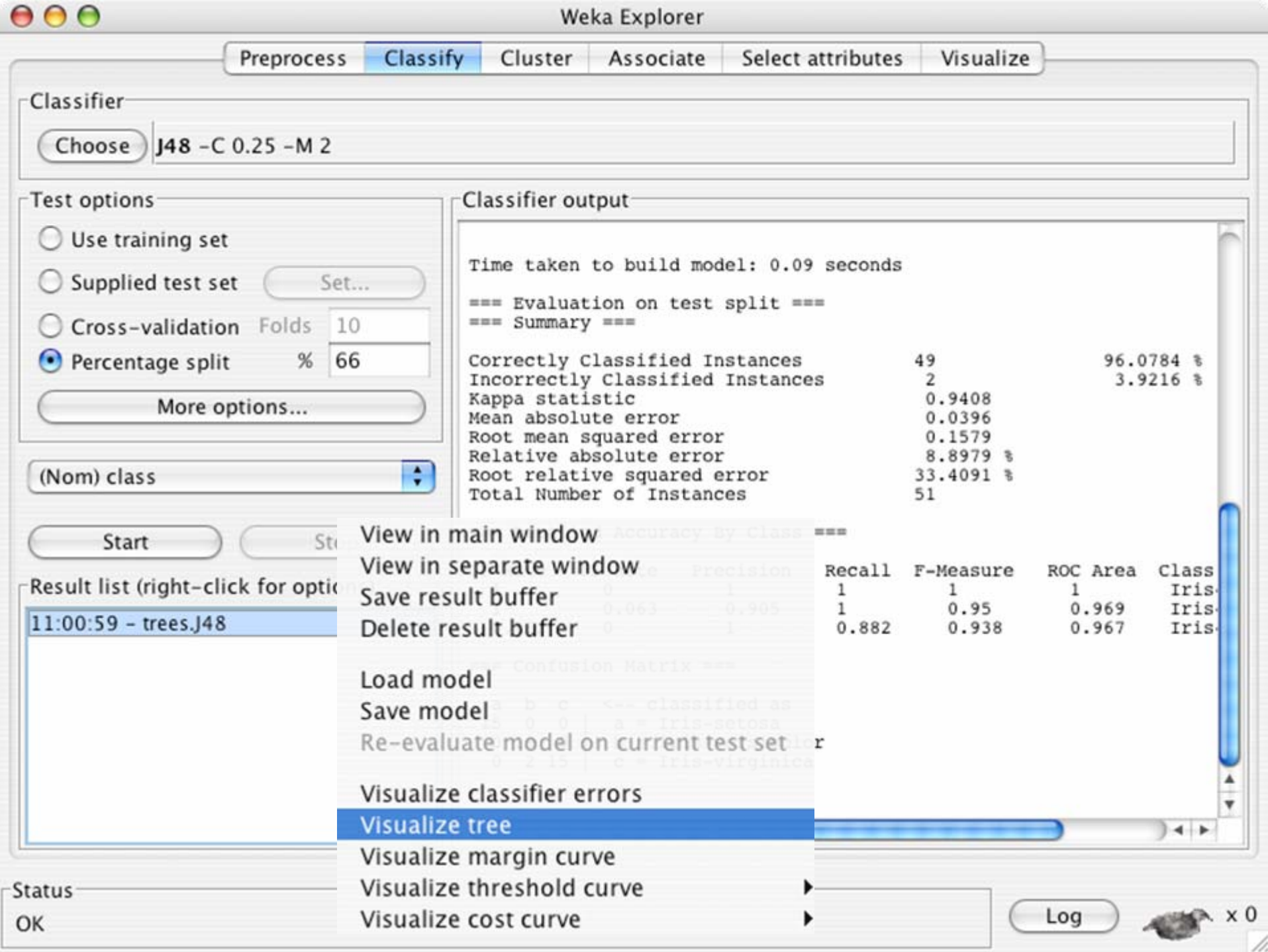

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0

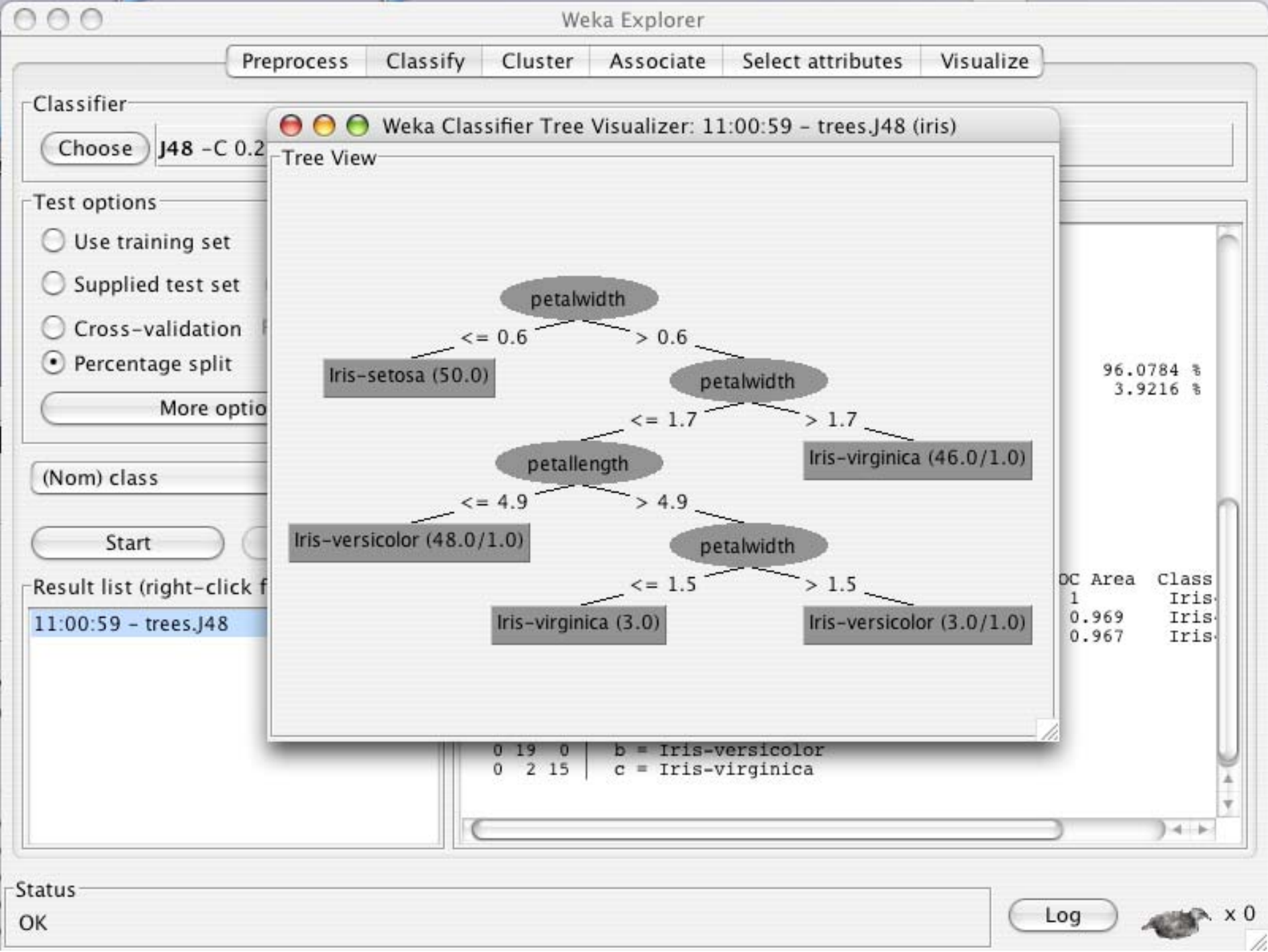

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
class
Test mode: split 66% train, remainder test
=== Classifier model (full training set) ===
J48 pruned tree
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
petalwidth <= 1.7
petallength <= 4.9: Iris-versicolor (48.0/1.0)
petallength > 4.9
petalwidth <= 1.5: Iris-virginica (3.0)
petalwidth > 1.5: Iris-versicolor (3.0/1.0)
petalwidth > 1.7: Iris-virginica (46.0/1.0)
Number of Leaves : 5
Size of the tree : 9
Status
OK
Log
x 0

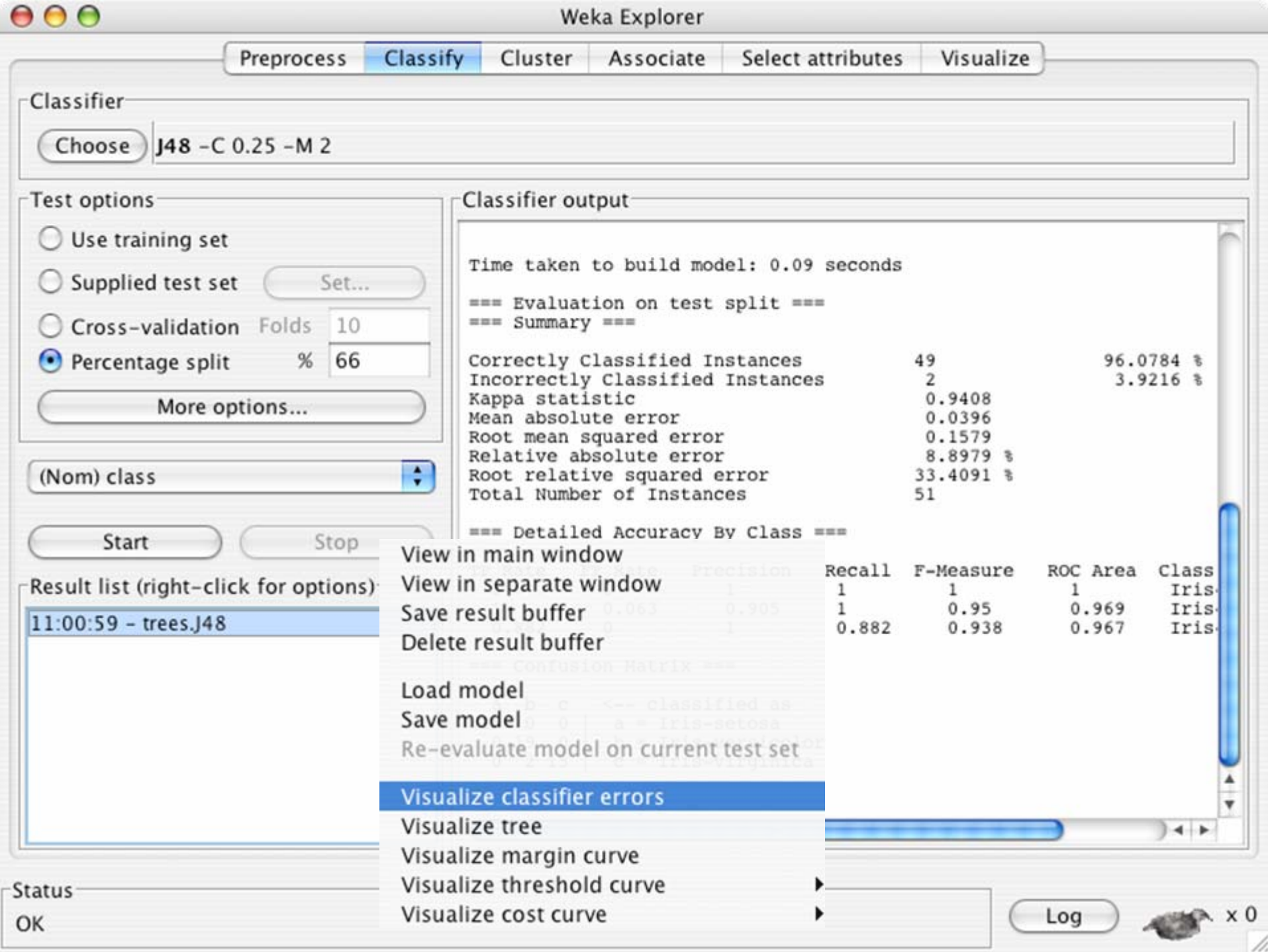

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
class
Test mode: split 66% train, remainder test
=== Classifier model (full training set) ===
J48 pruned tree
petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
petalwidth <= 1.7
petallength <= 4.9: Iris-versicolor (48.0/1.0)
petallength > 4.9
petalwidth <= 1.5: Iris-virginica (3.0)
petalwidth > 1.5: Iris-versicolor (3.0/1.0)
petalwidth > 1.7: Iris-virginica (46.0/1.0)
Number of Leaves : 5
Size of the tree : 9
Status
OK
Log
x 0

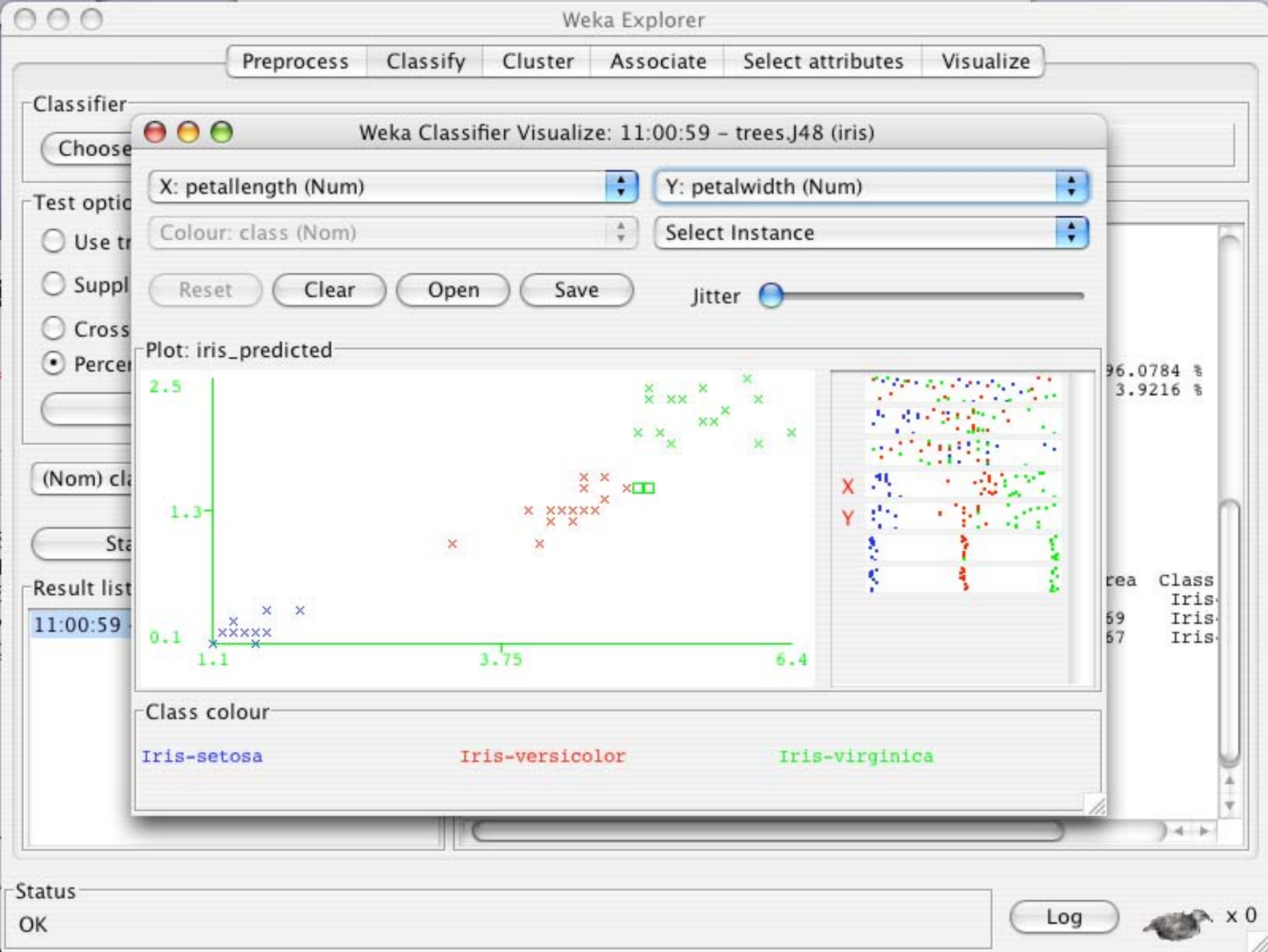

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

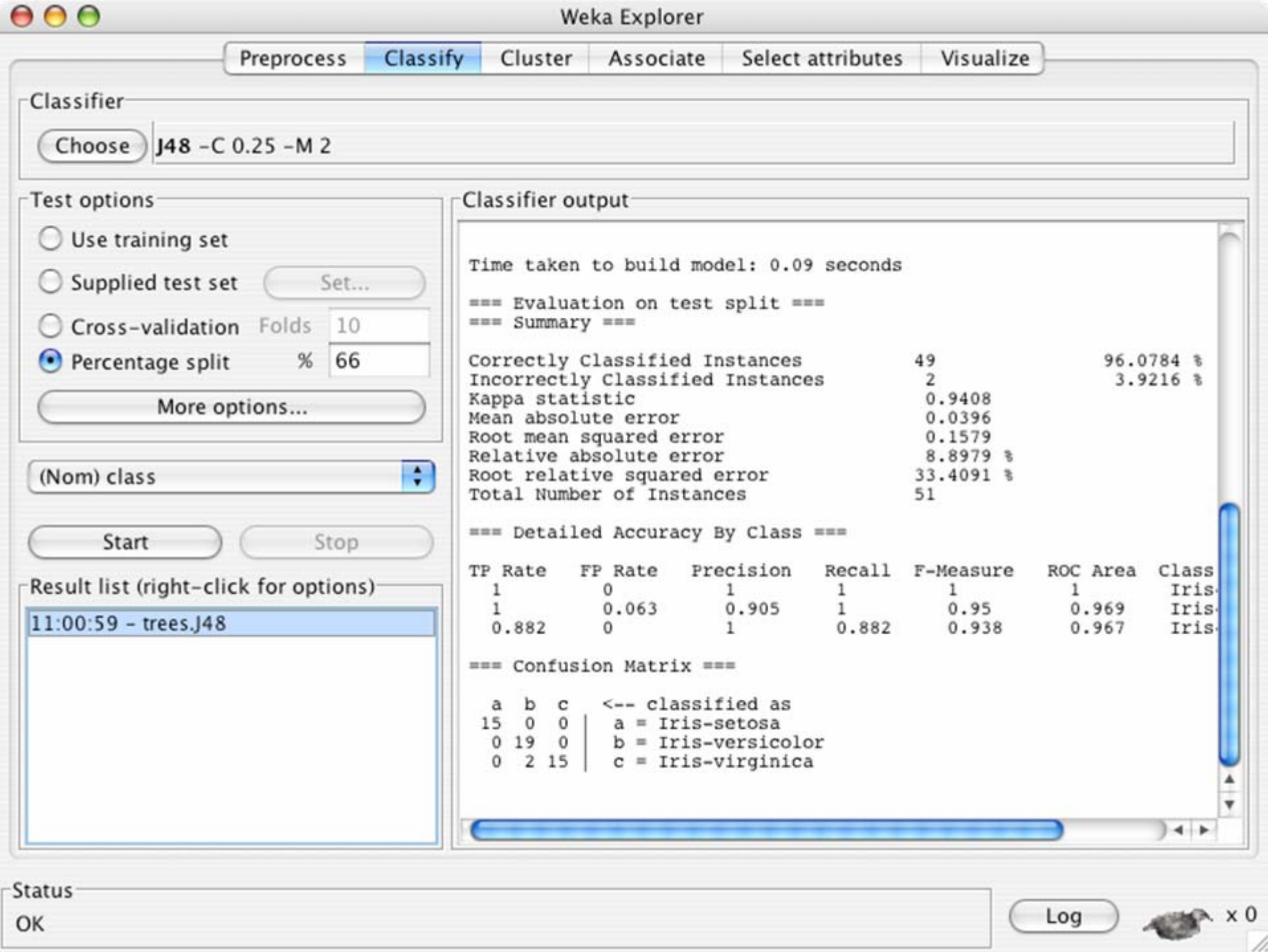

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

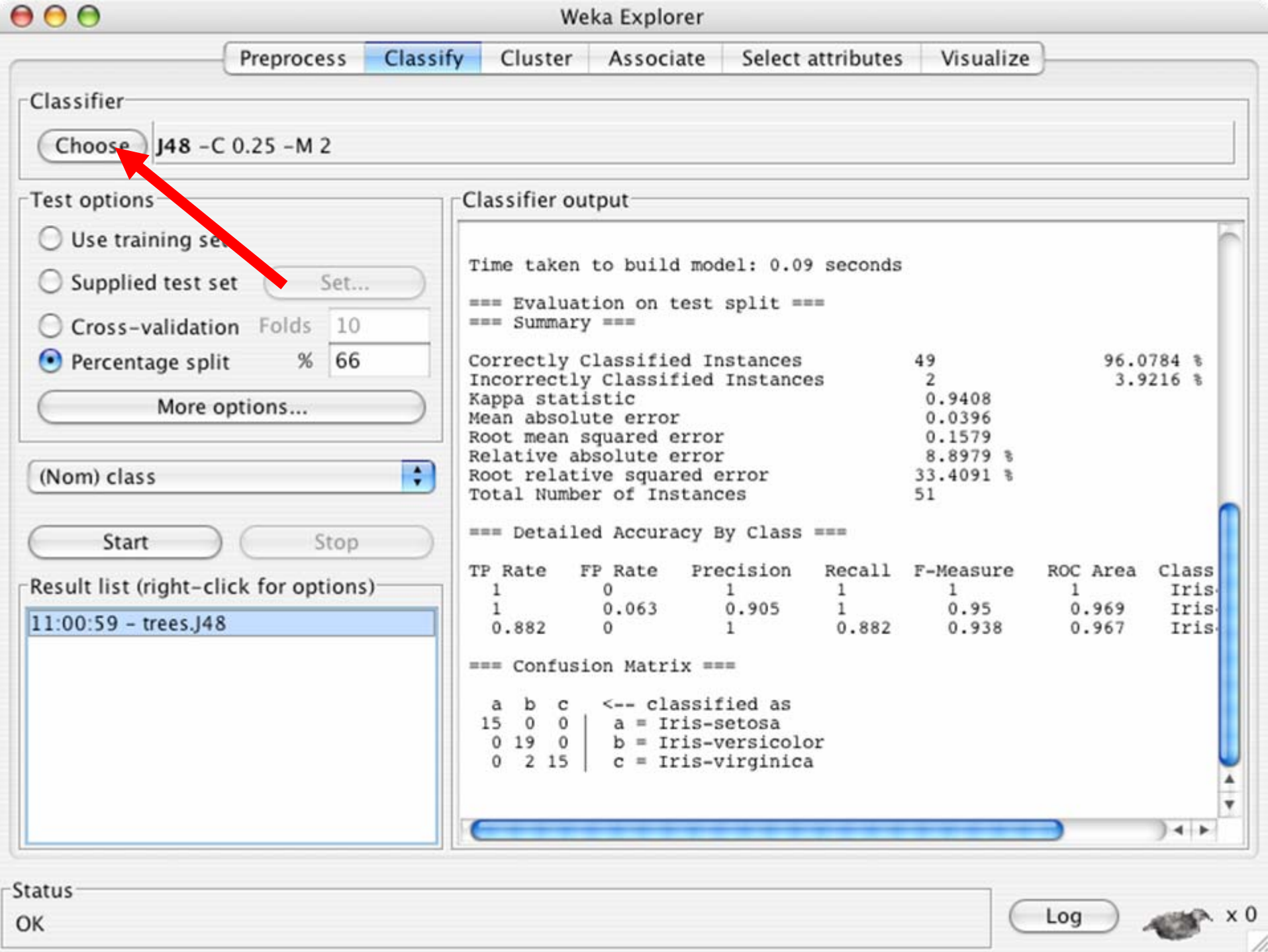

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Result list (right-click for opti
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
Recall F-Measure ROC Area Class
1 1 1 Iris
1 0.95 0.969 Iris
0.882 0.938 0.967 Iris
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Status
OK
Log
x 0

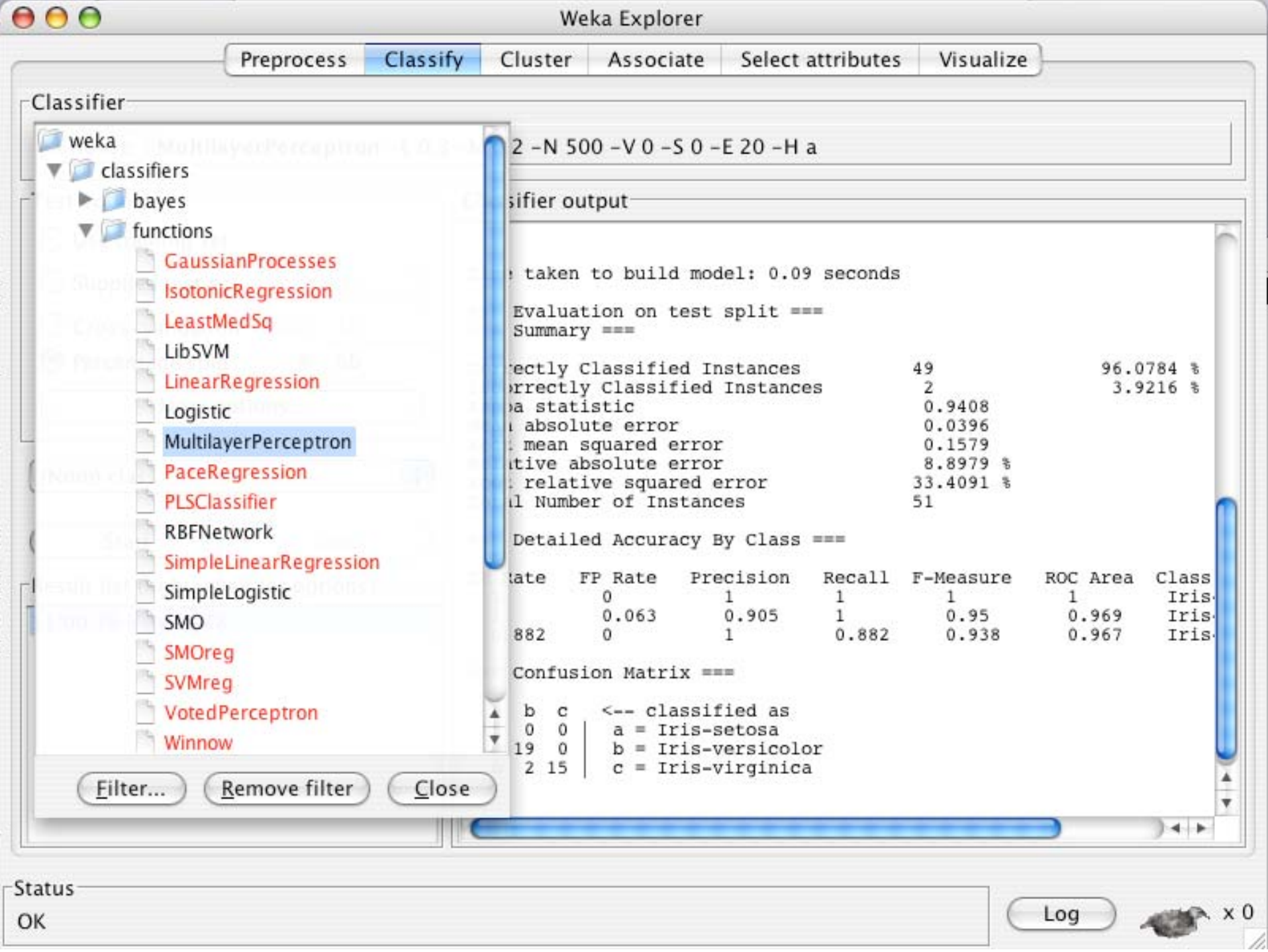

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.2
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optio
(Nom) class
Start
Result list (right-click f
11:00:59 - trees.J48
Status
OK
Log
x 0
Weka Classifier Tree Visualizer: 11:00:59 - 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)
96.0784 %
3.9216 %
OC Area Class
1 Iris
0.969 Iris
0.967 Iris
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica

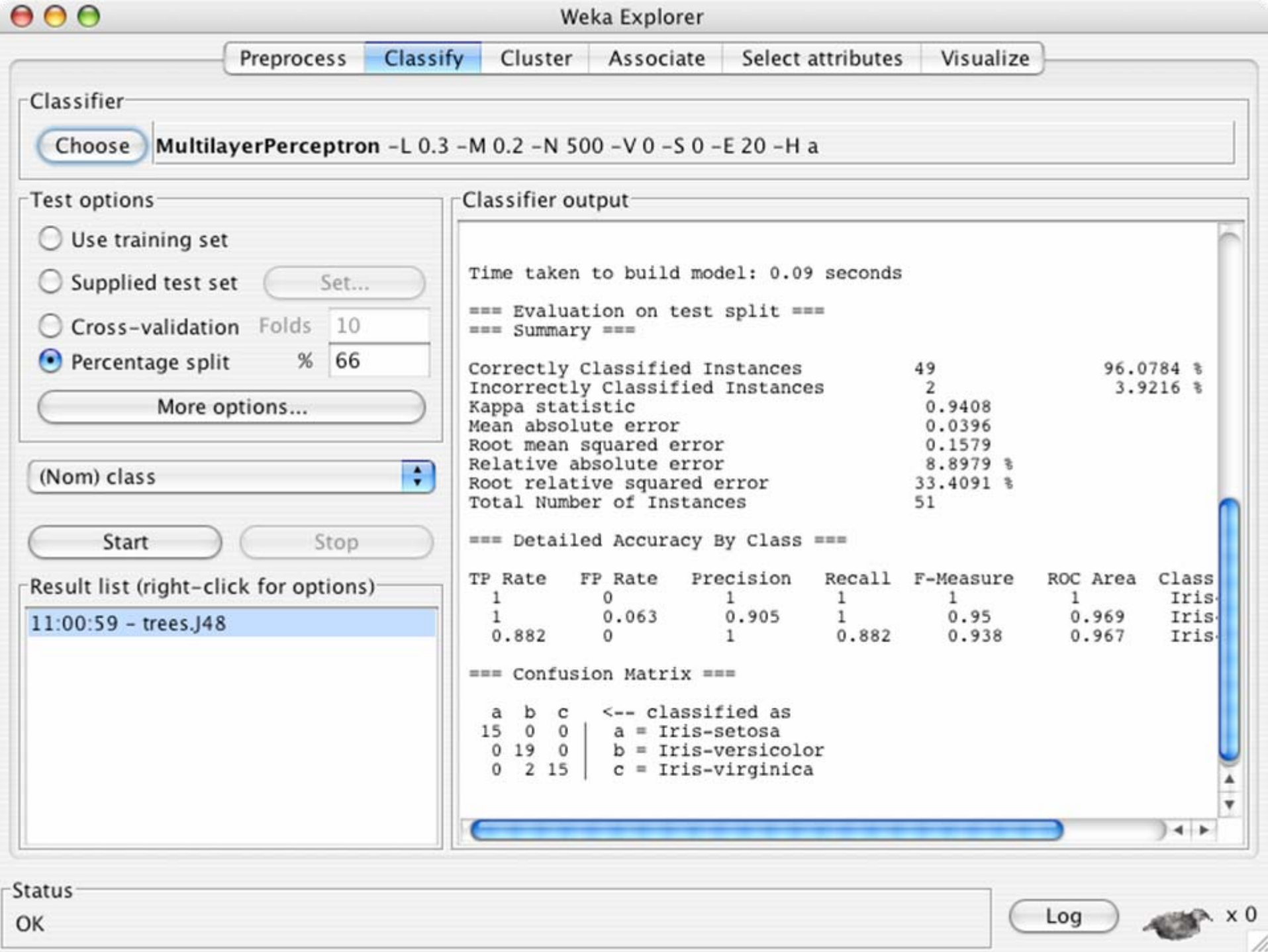

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
Recall F-Measure ROC Area Class
1 1 1 Iris
1 0.95 0.969 Iris
0.882 0.938 0.967 Iris
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Status
OK
Log
x 0

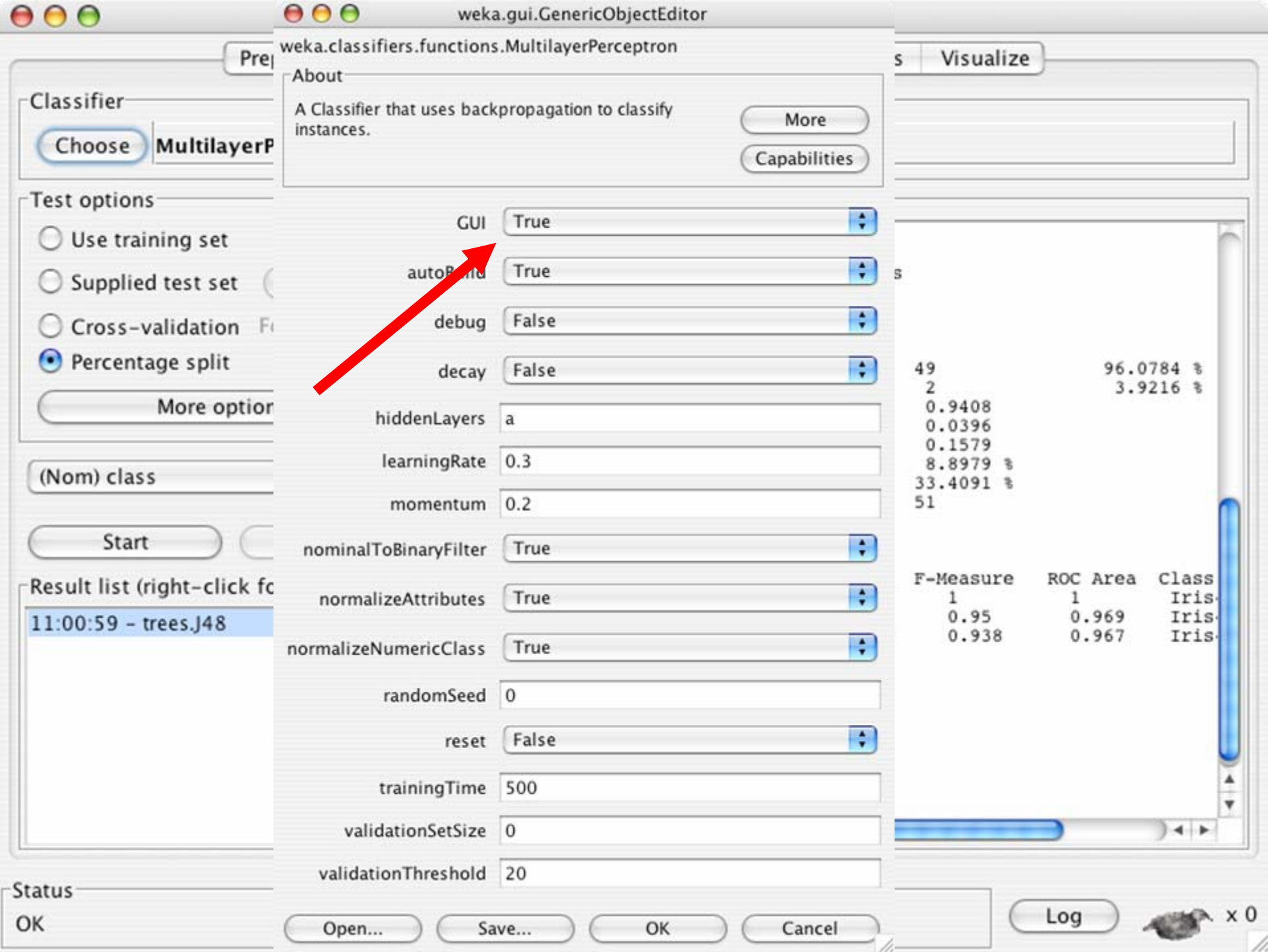

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
Test optic
Use tr
Suppl
Cross
Percer
(Nom) cl
Sta
Result list
11:00:59
Weka Classifier Visualize: 11:00:59 - trees.J48 (iris)
X: petallength (Num)
Y: petalwidth (Num)
Colour: class (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_predicted
2.5
1.3
0.1
1.1
3.75
6.4
X
Y
Class colour
Iris-setosa
Iris-versicolor
Iris-virginica
96.0784 %
3.9216 %
rea Class
Iris
69 Iris
67 Iris
Status
OK
Log
x 0

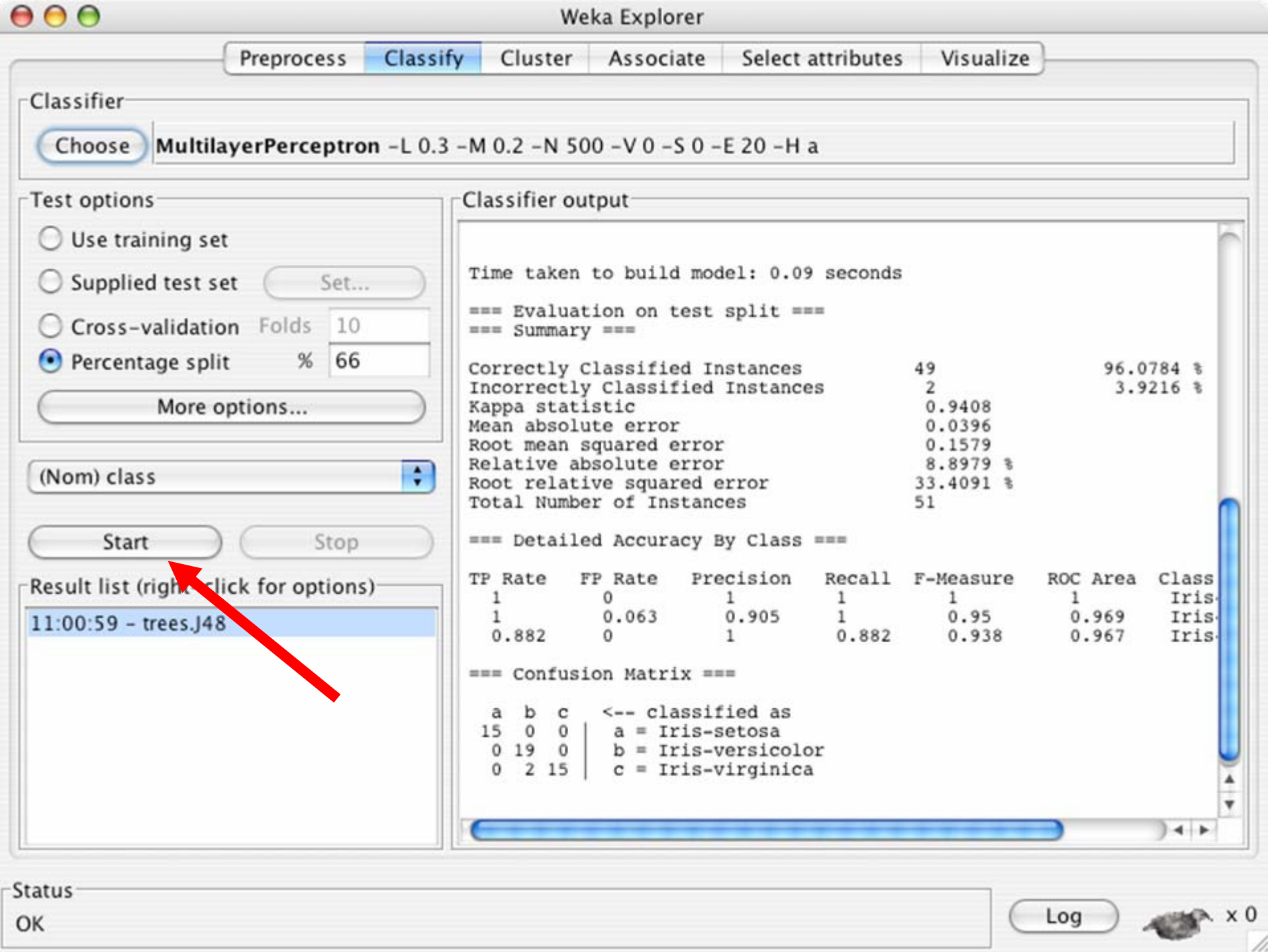

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

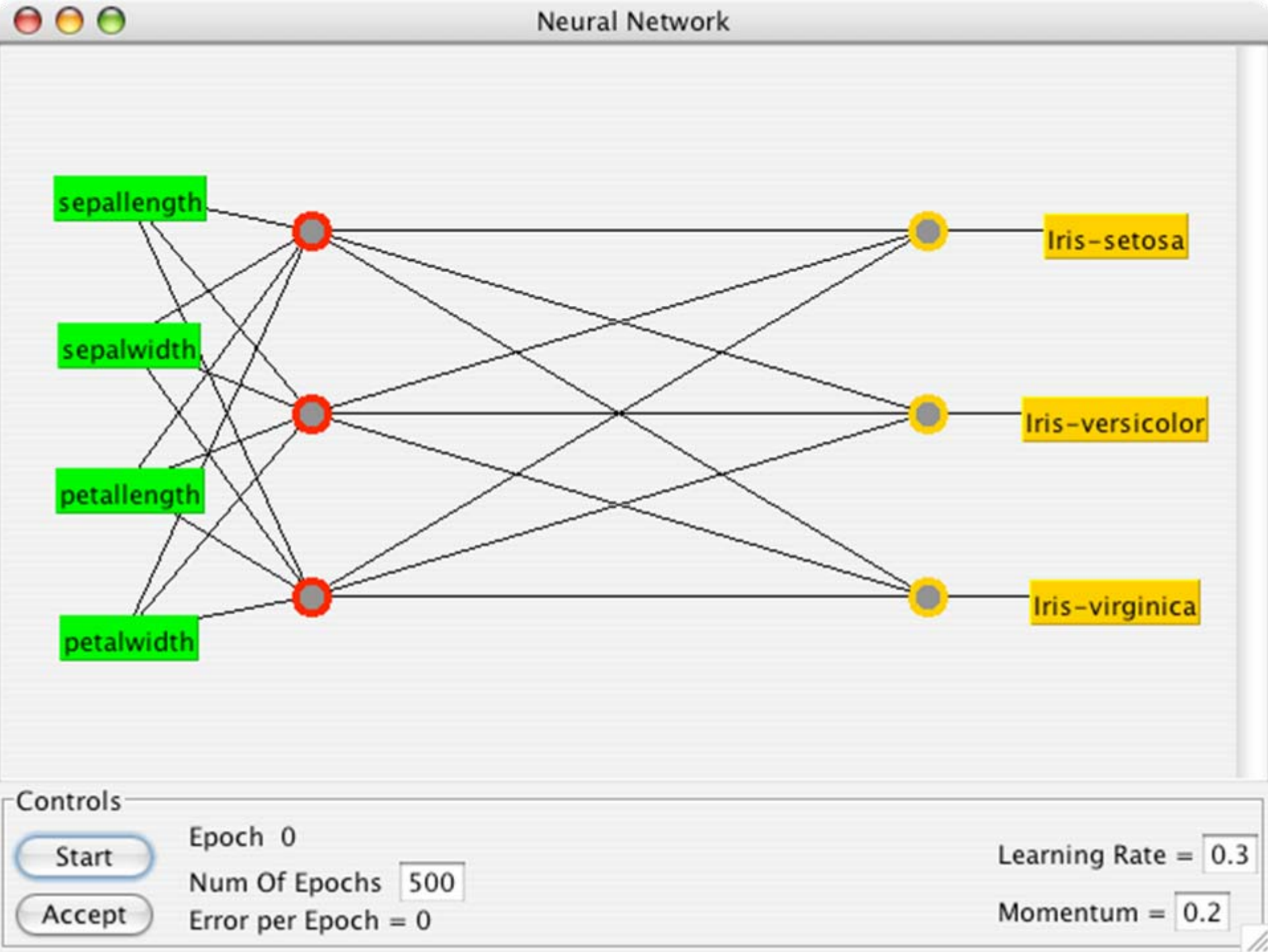

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

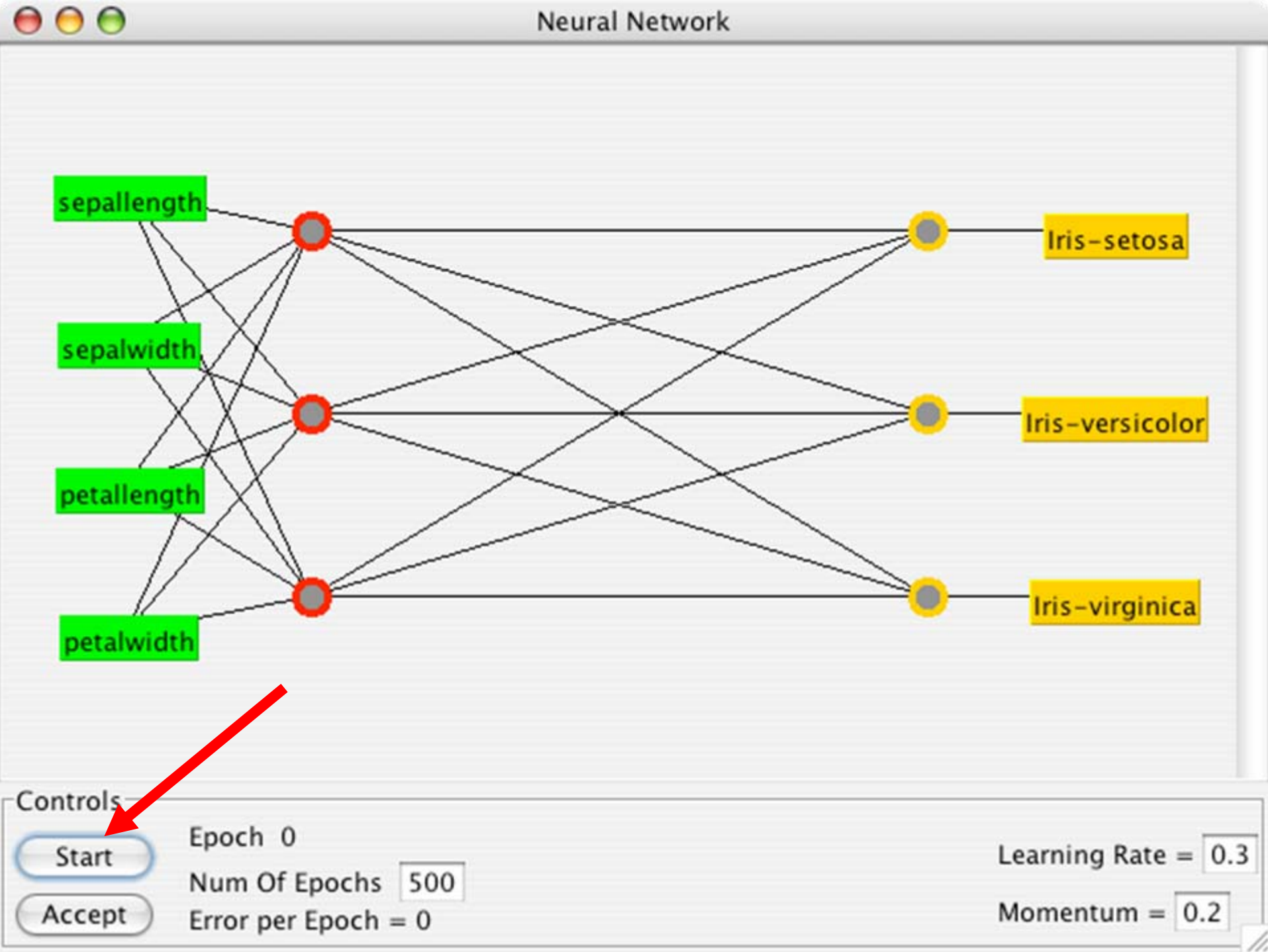

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
GaussianProcesses
IsotonicRegression
LeastMedSq
LibSVM
LinearRegression
Logistic
MultilayerPerceptron
PaceRegression
PLSClassifier
RBFNetwork
SimpleLinearRegression
SimpleLogistic
SMO
SMOreg
SVMreg
VotedPerceptron
Winnow
2 -N 500 -V 0 -S 0 -E 20 -H a
sifier output
taken to build model: 0.09 seconds
Evaluation on test split ===
Summary ===
rectly Classified Instances 49 96.0784 %
rrectly Classified Instances 2 3.9216 %
a statistic 0.9408
absolute error 0.0396
mean squared error 0.1579
itive absolute error 8.8979 %
relative squared error 33.4091 %
l Number of Instances 51
Detailed Accuracy By Class ===
ate FP Rate Precision Recall F-Measure ROC Area Class
0 1 1 1 1 Iris
0.063 0.905 1 0.95 0.969 Iris
882 0 1 0.882 0.938 0.967 Iris
Confusion Matrix ===
b c <-- classified as
0 0 | a = Iris-setosa
19 0 | b = Iris-versicolor
2 15 | c = Iris-virginica
Filter...
Remove filter
Close
Status
OK
Log
x 0

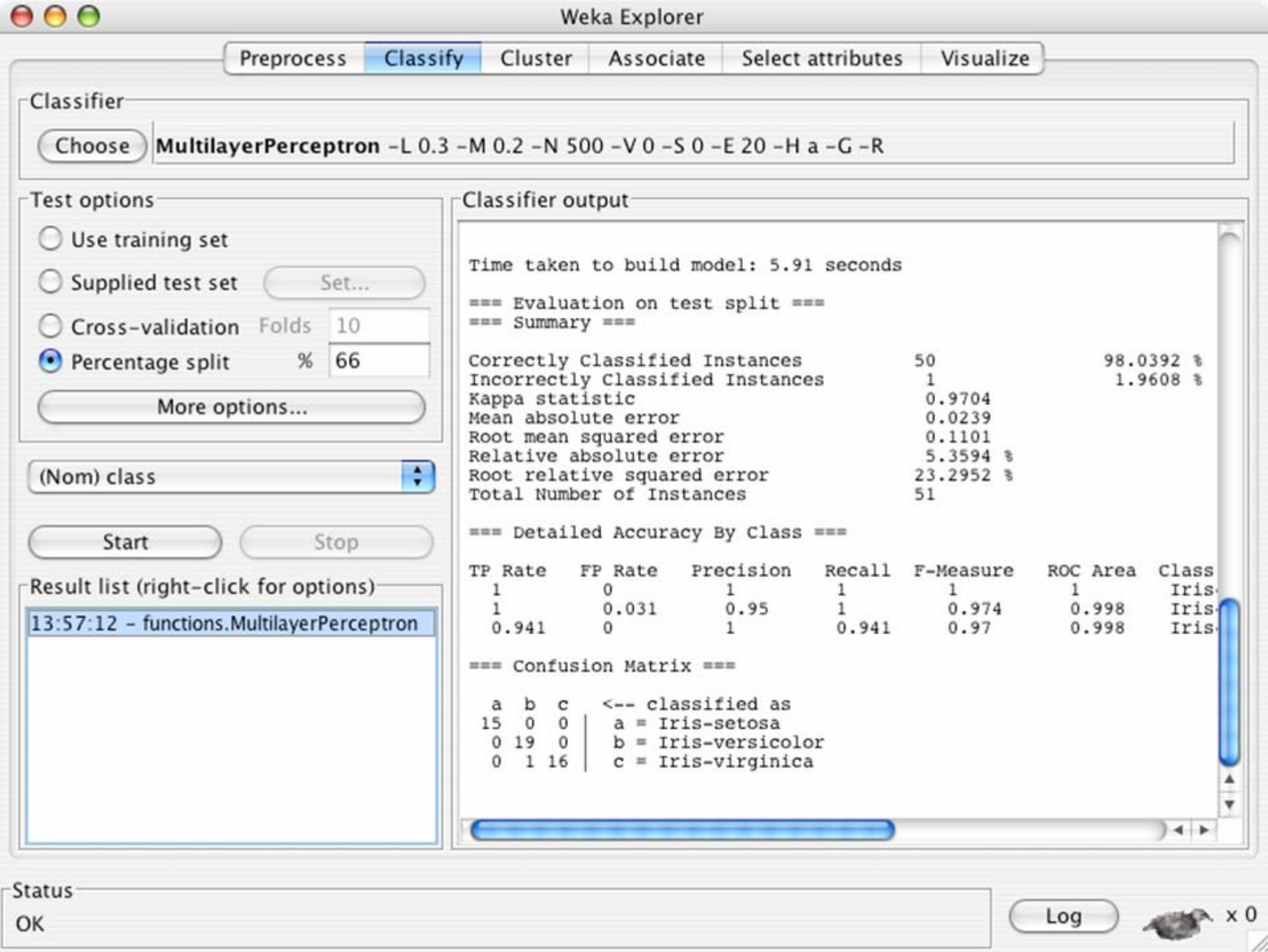

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

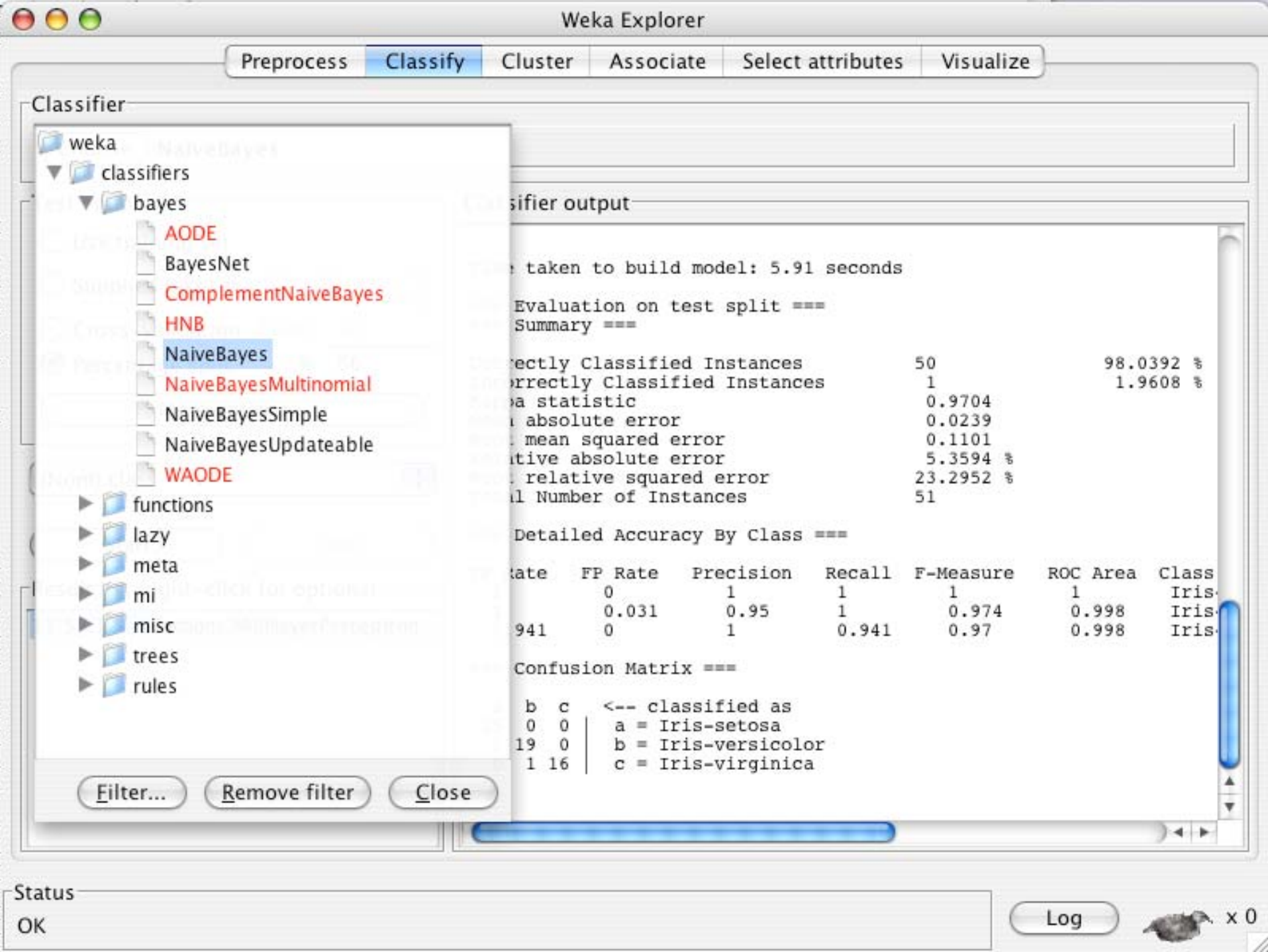

weka.gui.GenericObjectEditor
weka.classifiers.functions.MultilayerPerceptron
About
A Classifier that uses backpropagation to classify instances.
More
Capabilities
GUI
True
autoBuild
True
debug
False
decay
False
hiddenLayers
a
learningRate
0.3
momentum
0.2
nominalToBinaryFilter
True
normalizeAttributes
True
normalizeNumericClass
True
randomSeed
0
reset
False
trainingTime
500
validationSetSize
0
validationThreshold
20
Open...
Save...
OK
Cancel
Classifier
Choose
MultilayerP
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optior
(Nom) class
Start
Result list (right-click fc
11:00:59 - trees.J48
Status
OK
Visualize
96.0784 %
3.9216 %
0.9408
0.0396
0.1579
8.8979 %
33.4091 %
51
F-Measure
ROC Area
Class
1
1
Iris
0.95
0.969
Iris
0.938
0.967
Iris
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:00:59 - trees.J48
Classifier output
Time taken to build model: 0.09 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0396
Root mean squared error 0.1579
Relative absolute error 8.8979 %
Root relative squared error 33.4091 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.967 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0


Neural Network
sepallength
sepalwidth
petallength
petalwidth
Iris-setosa
Iris-versicolor
Iris-virginica
Controls
Start
Accept
Epoch 0
Num Of Epochs 500
Error per Epoch = 0
Learning Rate = 0.3
Momentum = 0.2


Neural Network
sepallength
sepalwidth
petallength
petalwidth
Iris-setosa
Iris-versicolor
Iris-virginica
Controls
Start
Accept
Epoch 0
Num Of Epochs 500
Error per Epoch = 0
Learning Rate = 0.3
Momentum = 0.2


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R
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)
13:57:12 - functions.MultilayerPerceptron
Classifier output
Time taken to build model: 5.91 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 50 98.0392 %
Incorrectly Classified Instances 1 1.9608 %
Kappa statistic 0.9704
Mean absolute error 0.0239
Root mean squared error 0.1101
Relative absolute error 5.3594 %
Root relative squared error 23.2952 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.031 0.95 1 0.974 0.998 Iris
0.941 0 1 0.941 0.97 0.998 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 1 16 | c = Iris-virginica
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
AODE
BayesNet
ComplementNaiveBayes
HNB
NaiveBayes
NaiveBayesMultinomial
NaiveBayesSimple
NaiveBayesUpdateable
WAODE
functions
lazy
meta
mi
misc
trees
rules
Filter...
Remove filter
Close
sifier output
taken to build model: 5.91 seconds
Evaluation on test split ===
Summary ===
rectly Classified Instances 50 98.0392 %
rrectly Classified Instances 1 1.9608 %
a statistic 0.9704
absolute error 0.0239
mean squared error 0.1101
tive absolute error 5.3594 %
relative squared error 23.2952 %
l Number of Instances 51
Detailed Accuracy By Class ===
tate FP Rate Precision Recall F-Measure ROC Area Class
0 1 1 1 1 Iris
0.031 0.95 1 0.974 0.998 Iris
941 0 1 0.941 0.97 0.998 Iris
Confusion Matrix ===
b c <-- classified as
0 0 | a = Iris-setosa
19 0 | b = Iris-versicolor
1 16 | c = Iris-virginica
Status
OK
Log
x 0

Weka Knowledge 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)

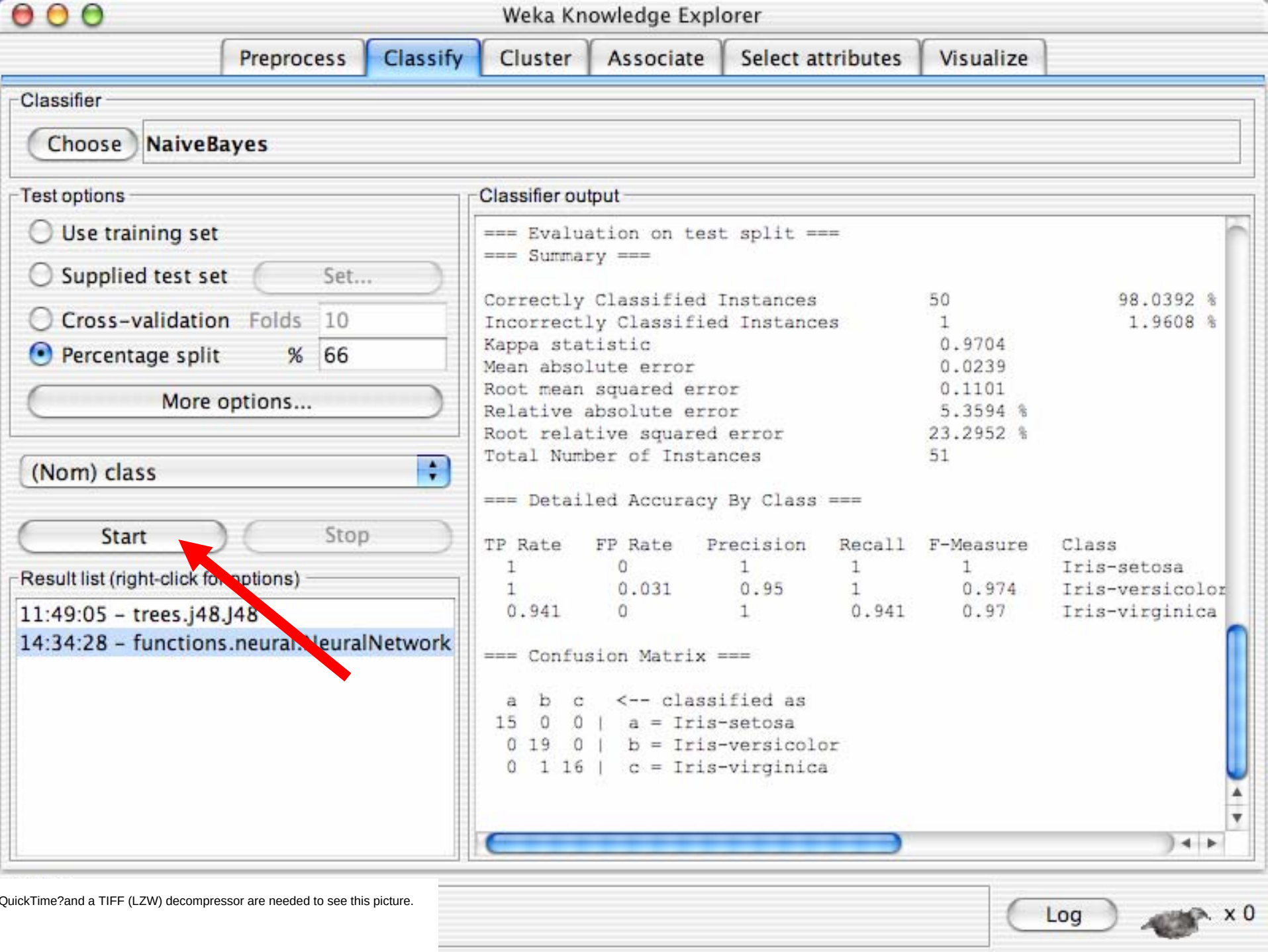
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork

Classifier output

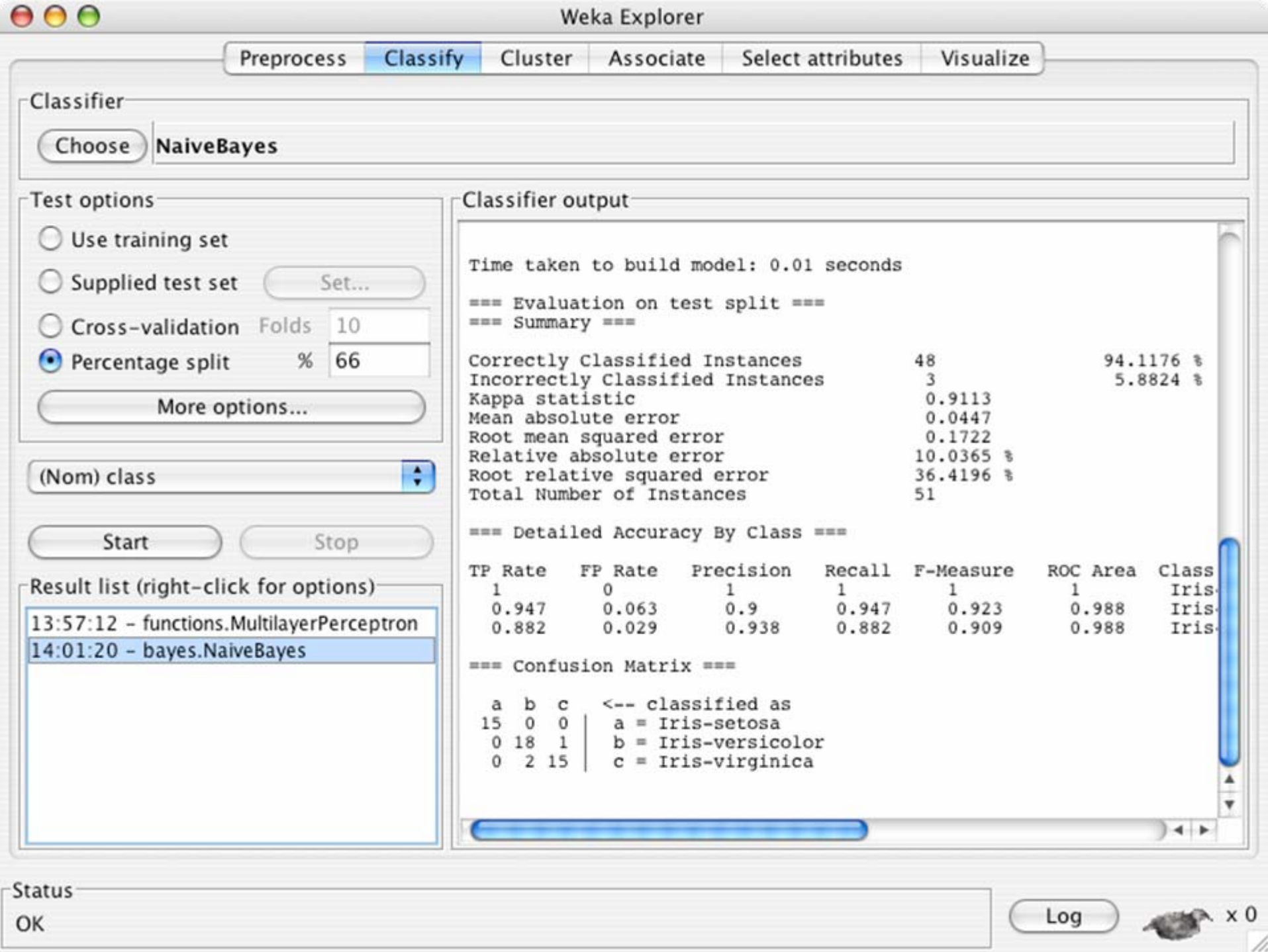
```
=== Evaluation on test split ===
=== Summary ===

Correctly Classified Instances          50               98.0392 %
Incorrectly Classified Instances         1                1.9608 %
Kappa statistic                          0.9704
Mean absolute error                      0.0239
Root mean squared error                  0.1101
Relative absolute error                  5.3594 %
Root relative squared error             23.2952 %
Total Number of Instances               51

=== Detailed Accuracy By Class ===

TP Rate   FP Rate   Precision   Recall  F-Measure   Class
  1          0          1          1          1        Iris-setosa
  1          0.031      0.95       1          0.974    Iris-versicolor
  0.941      0          1          0.941      0.97     Iris-virginica

=== Confusion Matrix ===

  a  b  c   <-- classified as
 15  0  0 |  a = Iris-setosa
  0 19  0 |  b = Iris-versicolor
  0  1 16 |  c = Iris-virginica
```

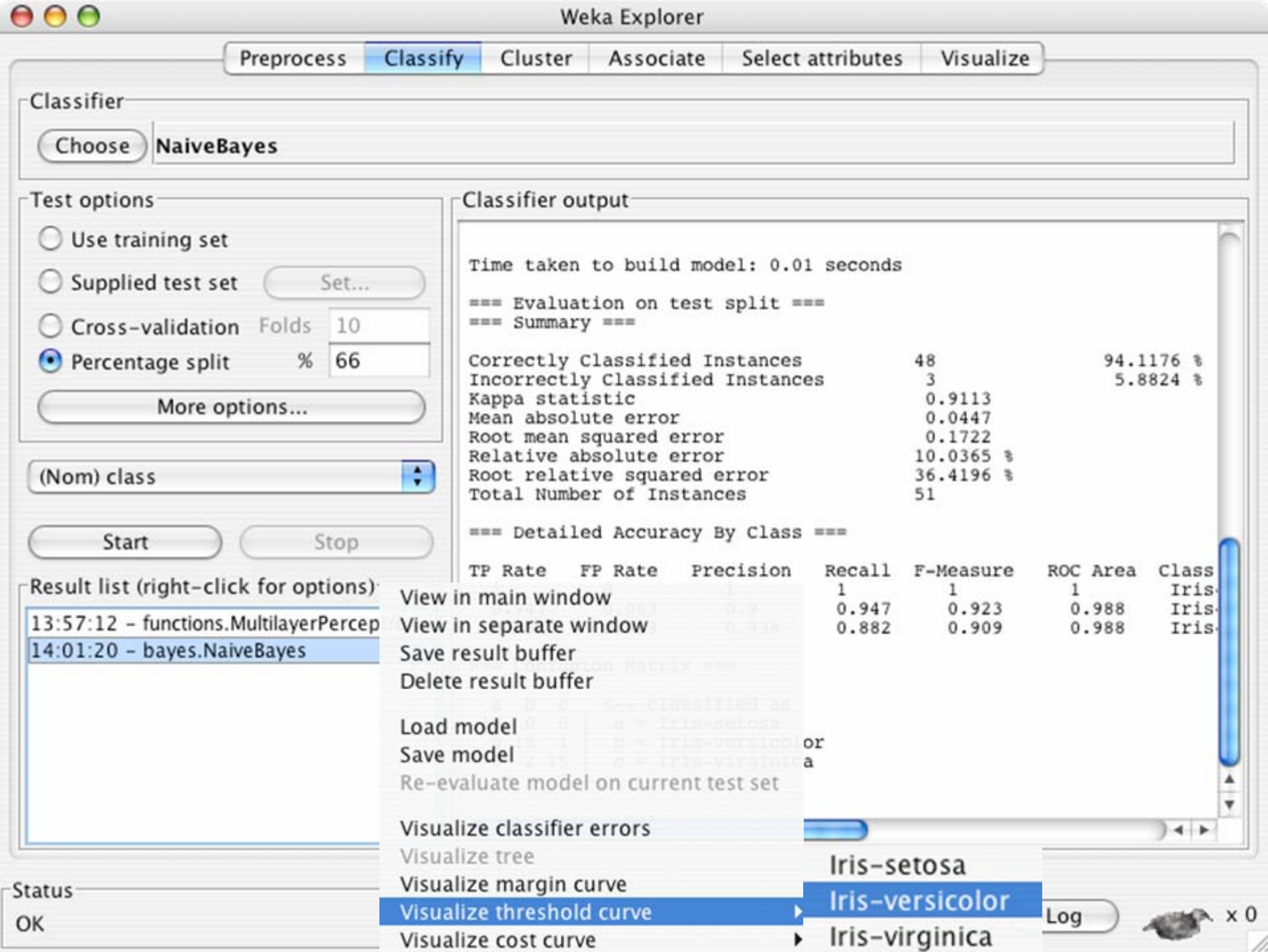
Log x 0

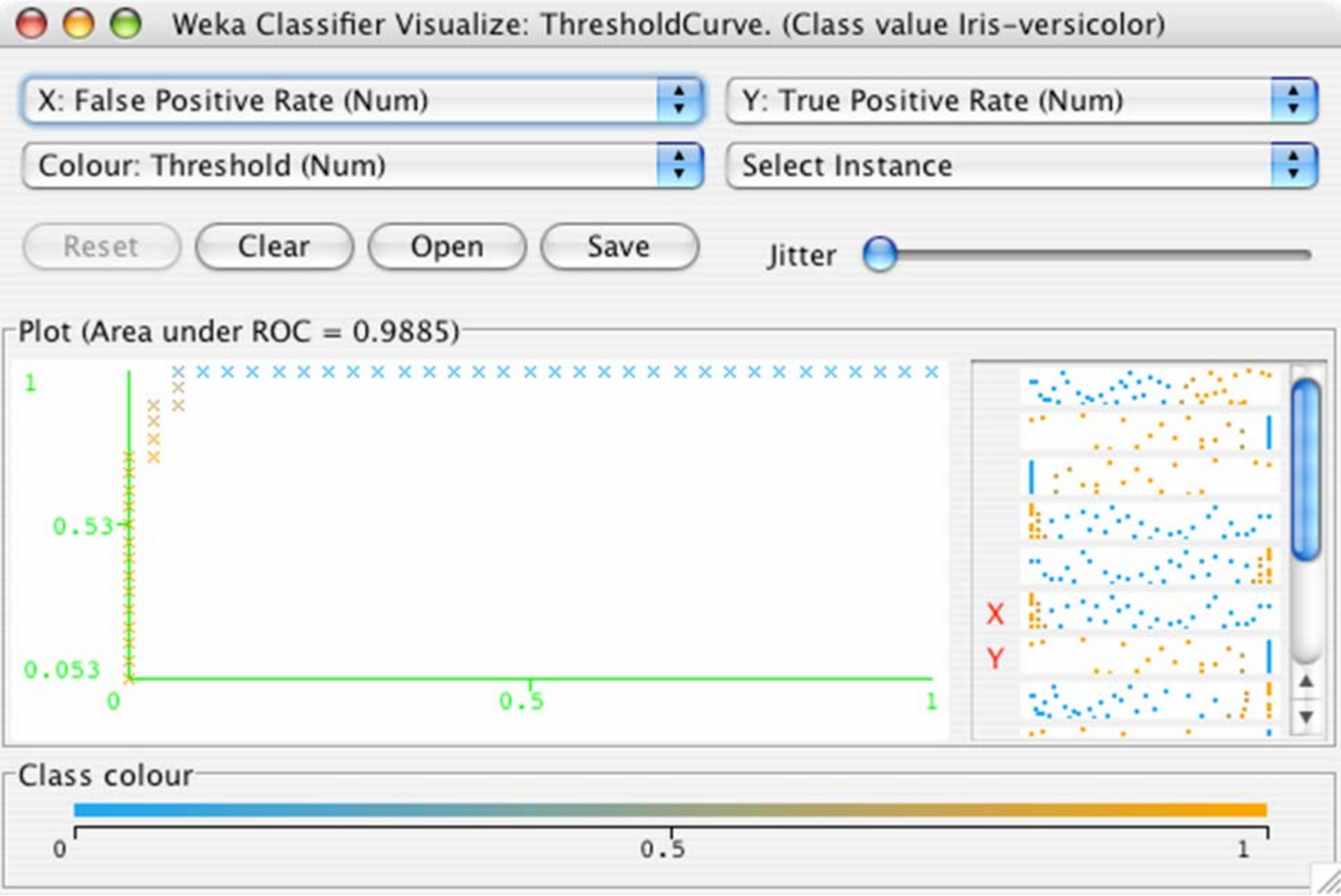
QuickTime?and a TIFF (LZW) decompressor are needed to see this picture.

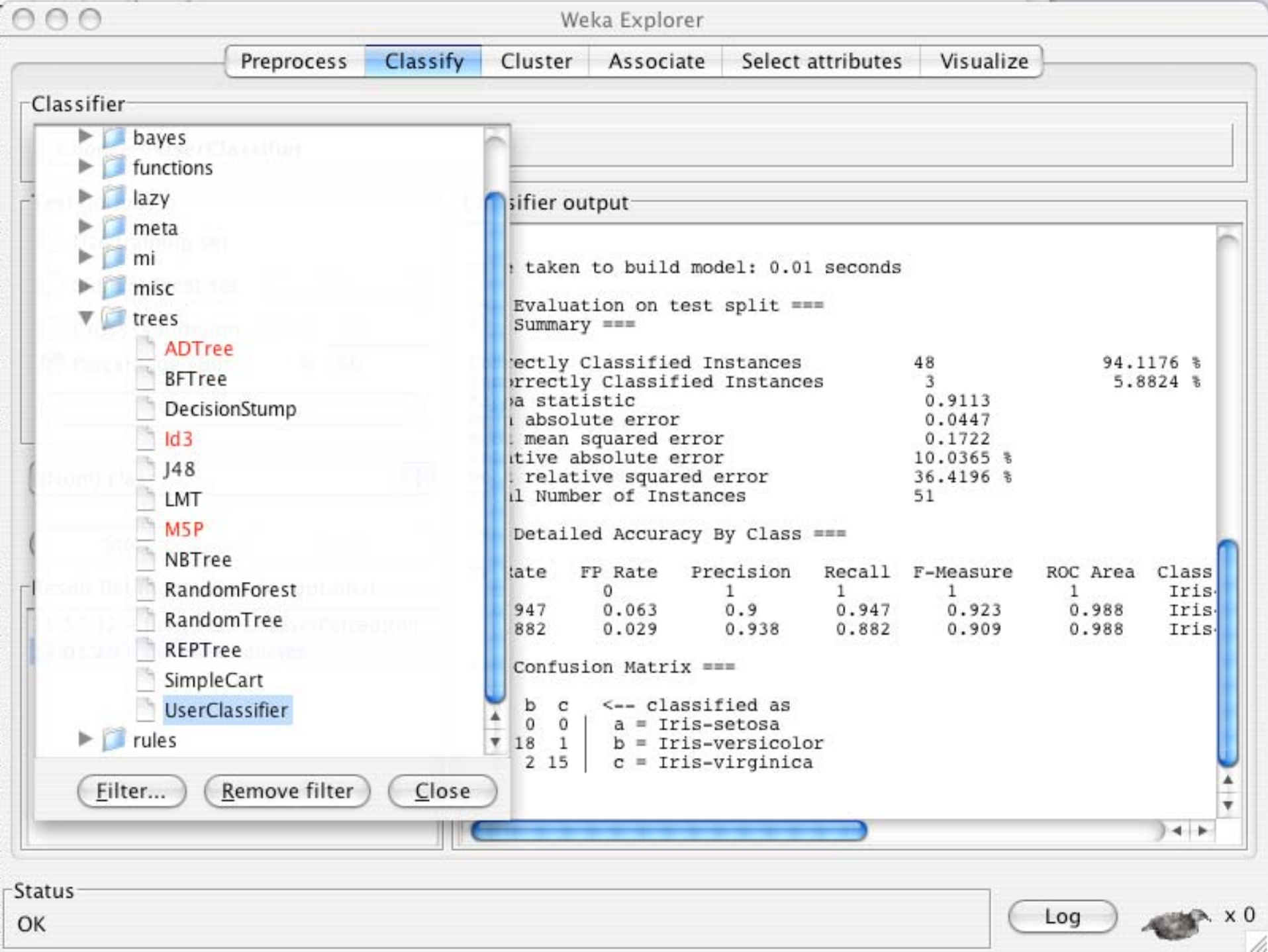


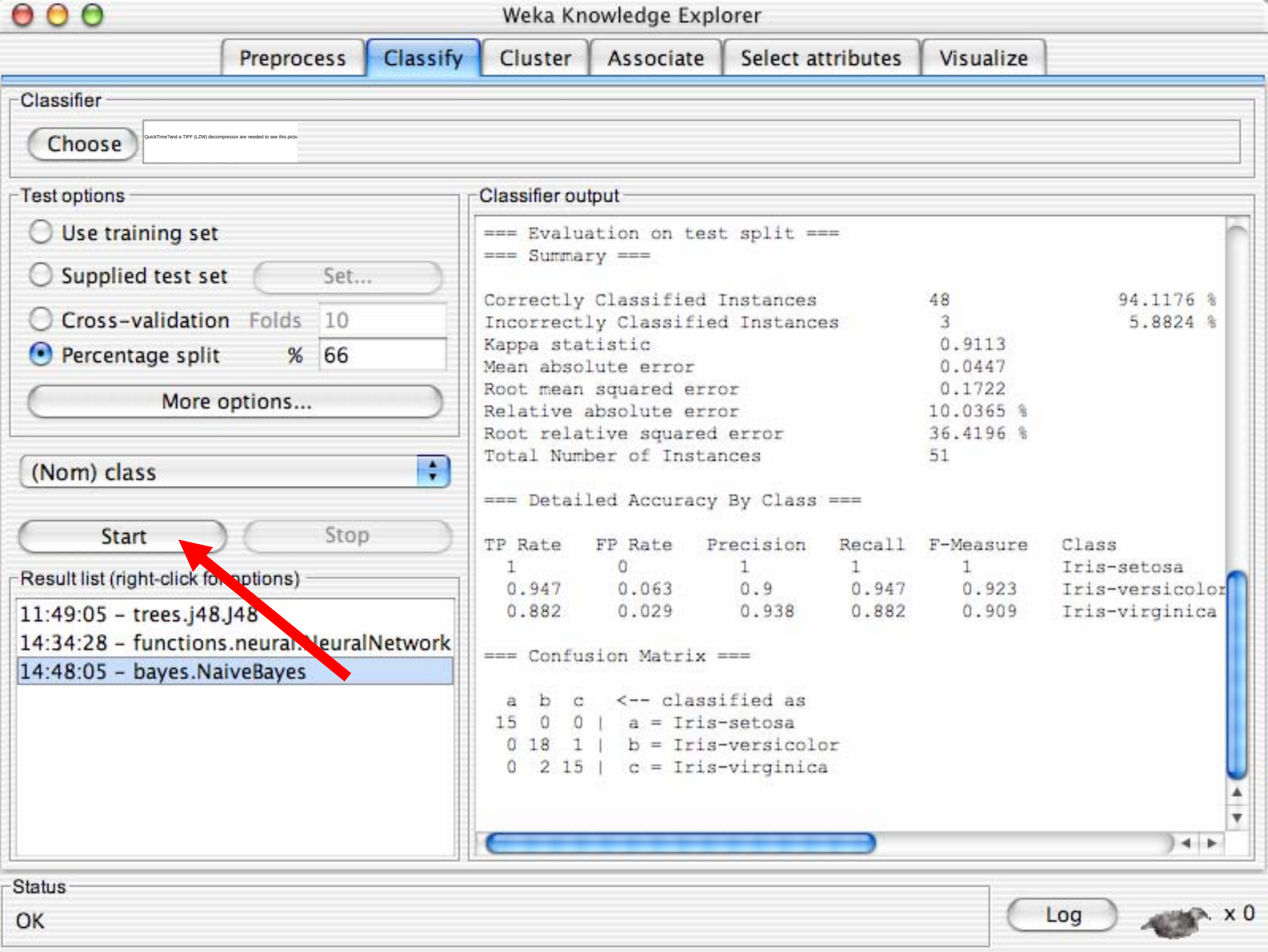
QuickTime?and a TIFF (LZW) decompressor are needed to see this picture.

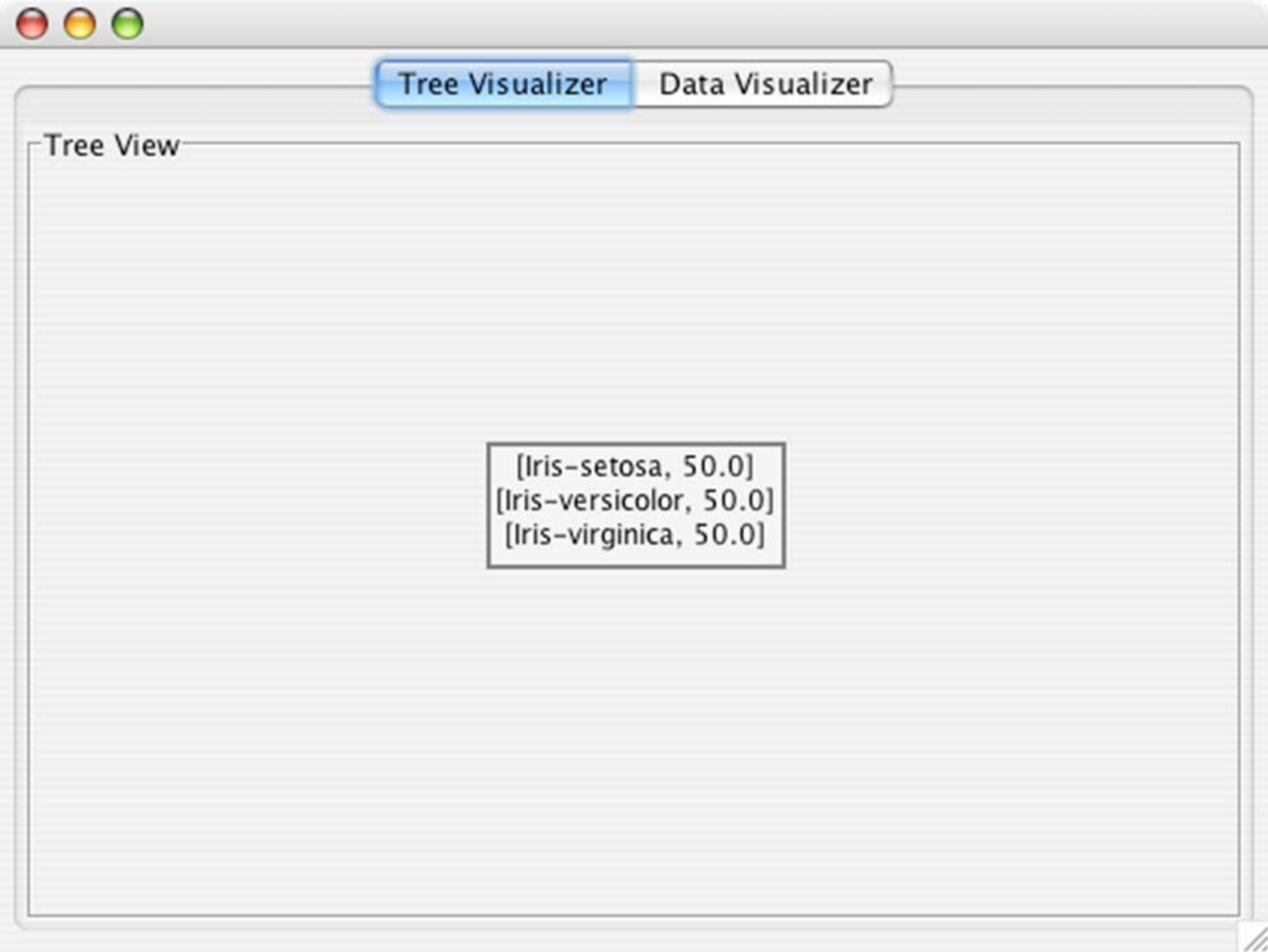

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
Classifier output
Time taken to build model: 0.01 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 48 94.1176 %
Incorrectly Classified Instances 3 5.8824 %
Kappa statistic 0.9113
Mean absolute error 0.0447
Root mean squared error 0.1722
Relative absolute error 10.0365 %
Root relative squared error 36.4196 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
0.947 0.063 0.9 0.947 0.923 0.988 Iris
0.882 0.029 0.938 0.882 0.909 0.988 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
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...
Classifier output
Time taken to build model: 0.01 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 48 94.1176 %
Incorrectly Classified Instances 3 5.8824 %
Kappa statistic 0.9113
Mean absolute error 0.0447
Root mean squared error 0.1722
Relative absolute error 10.0365 %
Root relative squared error 36.4196 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 1 1 Iris
0.947 0.923 0.988 Iris
0.882 0.909 0.988 Iris
(Nom) class
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPercep
14:01:20 - bayes.NaiveBayes
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Iris-setosa
Iris-versicolor
Iris-virginica
Status
OK
Log
x 0


Weka Classifier Visualize: ThresholdCurve. (Class value Iris-versicolor)
X: False Positive Rate (Num)
Y: True Positive Rate (Num)
Colour: Threshold (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot (Area under ROC = 0.9885)
1
0.53
0.053
0
0.5
1
X
Y
Class colour
0
0.5
1


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
bayes
functions
lazy
meta
mi
misc
trees
ADTree
BFTree
DecisionStump
Id3
J48
LMT
M5P
NBTree
RandomForest
RandomTree
REPTree
SimpleCart
UserClassifier
rules
Filter...
Remove filter
Close
sifier output
taken to build model: 0.01 seconds
Evaluation on test split ===
Summary ===
rectly Classified Instances 48 94.1176 %
rrectly Classified Instances 3 5.8824 %
a statistic 0.9113
absolute error 0.0447
mean squared error 0.1722
tive absolute error 10.0365 %
relative squared error 36.4196 %
l Number of Instances 51
Detailed Accuracy By Class ===
ate FP Rate Precision Recall F-Measure ROC Area Class
0 1 1 1 1 Iris
947 0.063 0.9 0.947 0.923 0.988 Iris
882 0.029 0.938 0.882 0.909 0.988 Iris
Confusion Matrix ===
b c <-- classified as
0 0 | a = Iris-setosa
18 1 | b = Iris-versicolor
2 15 | c = Iris-virginica
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
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)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
Classifier output
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 48 94.1176 %
Incorrectly Classified Instances 3 5.8824 %
Kappa statistic 0.9113
Mean absolute error 0.0447
Root mean squared error 0.1722
Relative absolute error 10.0365 %
Root relative squared error 36.4196 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure Class
1 0 1 1 1 Iris-setosa
0.947 0.063 0.9 0.947 0.923 Iris-versicolor
0.882 0.029 0.938 0.882 0.909 Iris-virginica
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 18 1 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0


Tree Visualizer
Data Visualizer
Tree View
[Iris-setosa, 50.0]
[Iris-versicolor, 50.0]
[Iris-virginica, 50.0]

Tree Visualizer | Data Visualizer

X: petallength (Num)

Y: petalwidth (Num)

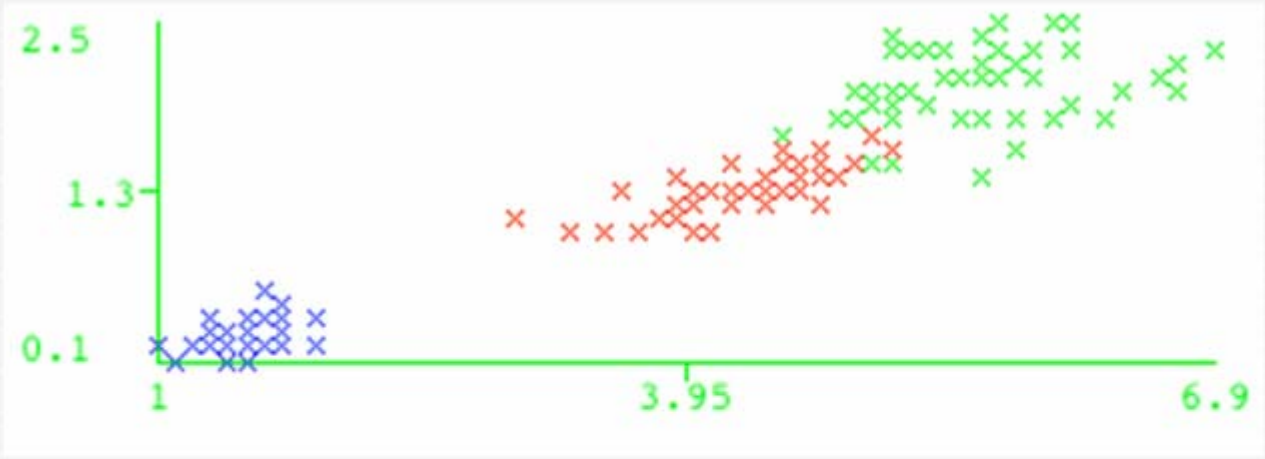
Colour: class (Nom)

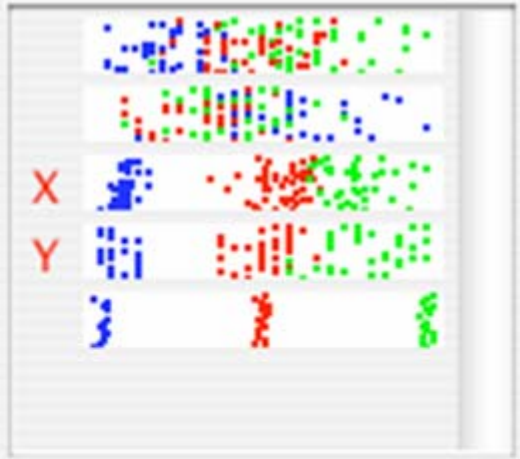
Select Instance

Submit | Clear | Open | Save

Jitter

Plot: iris





Class colour

Iris-setosa Iris-versicolor Iris-virginica

Tree Visualizer | Data Visualizer

X: petallength (Num)

Y: petalwidth (Num)

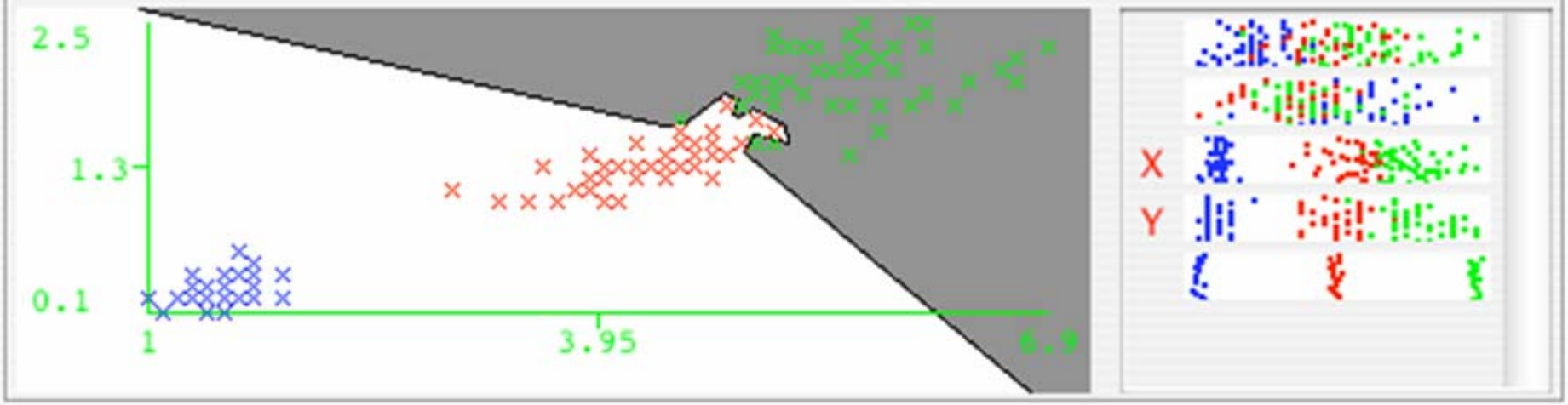
Colour: class (Nom)

Polyline

Submit | Clear | Open | Save

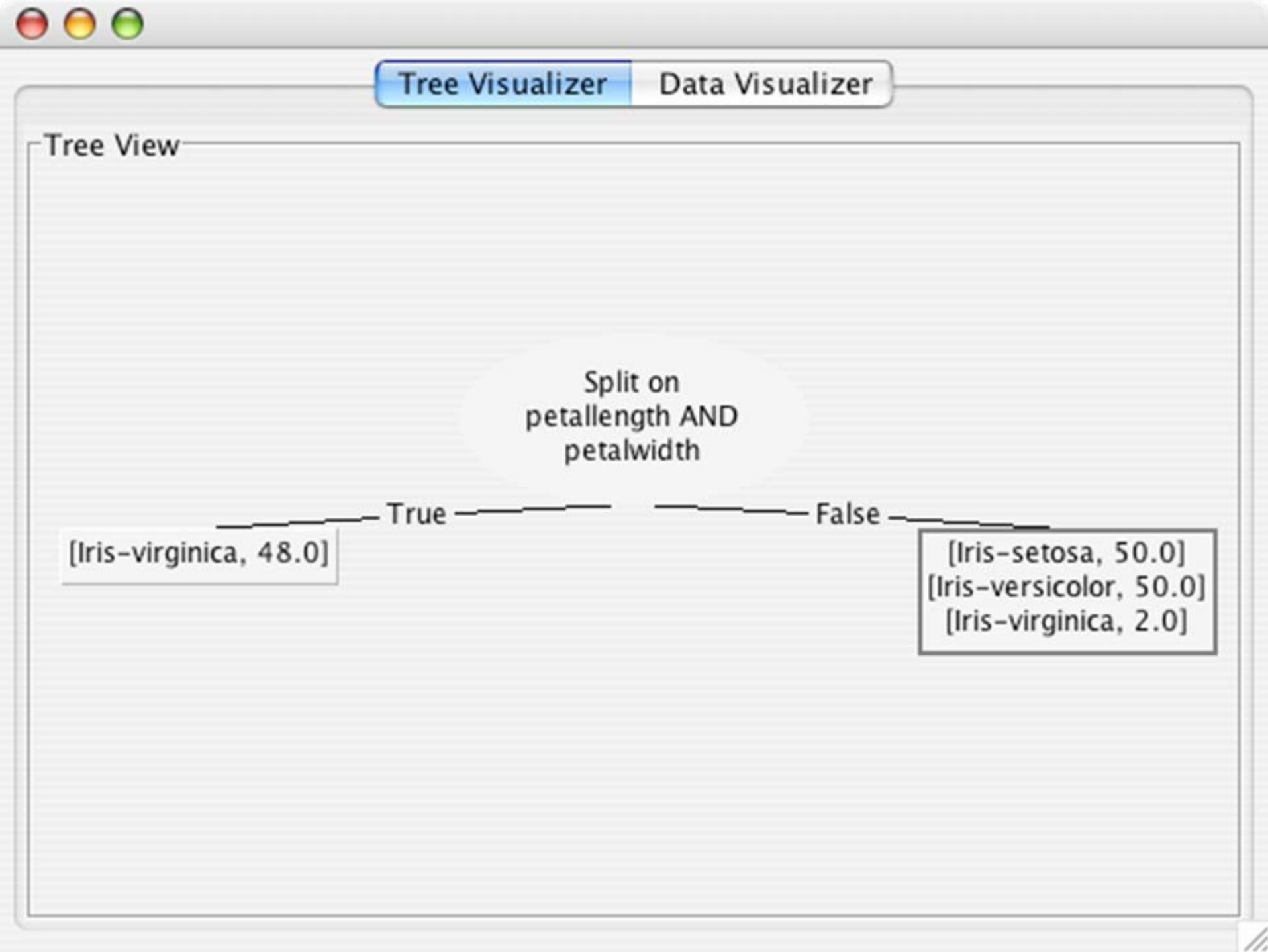
Jitter

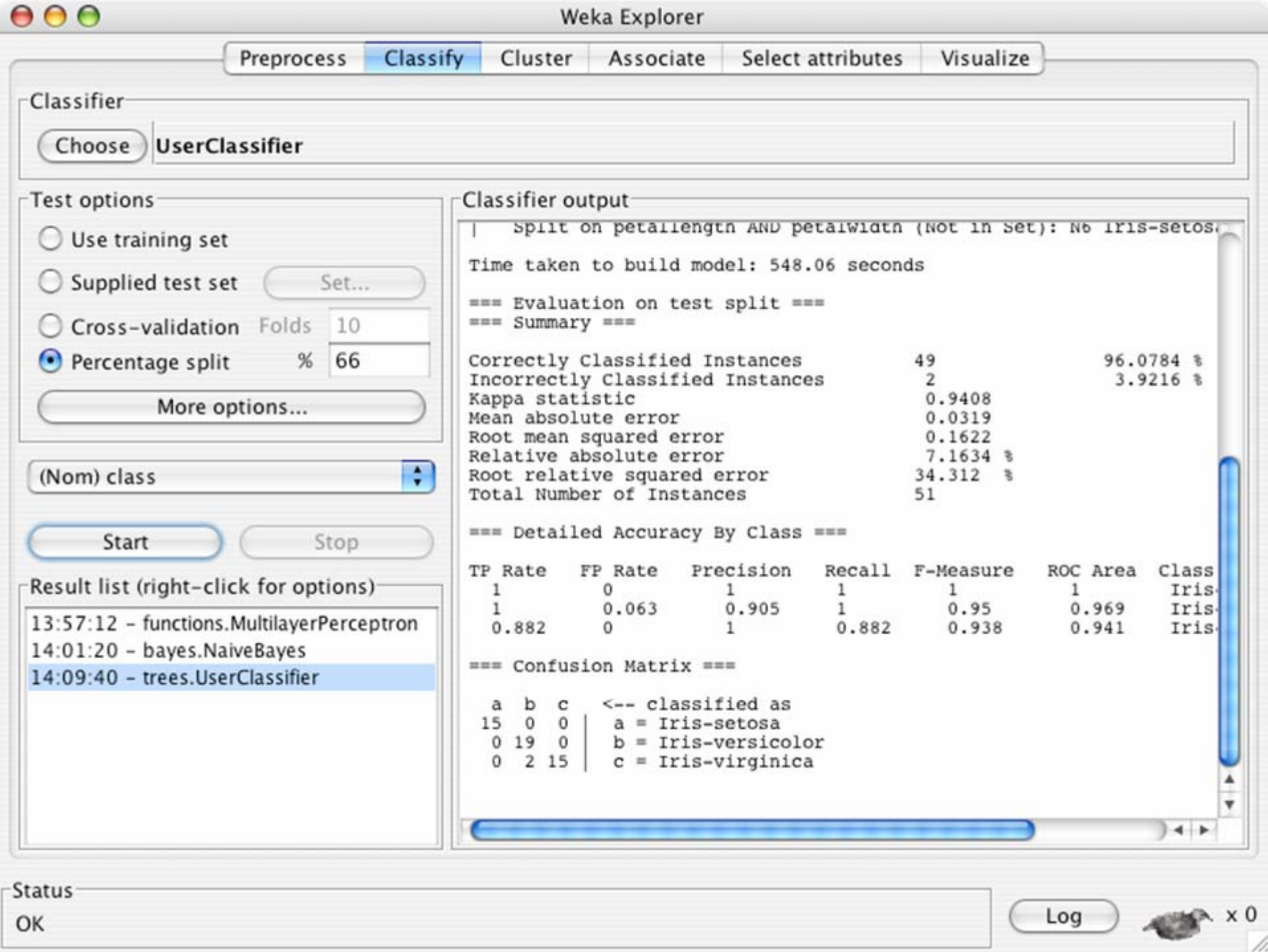
Plot: iris

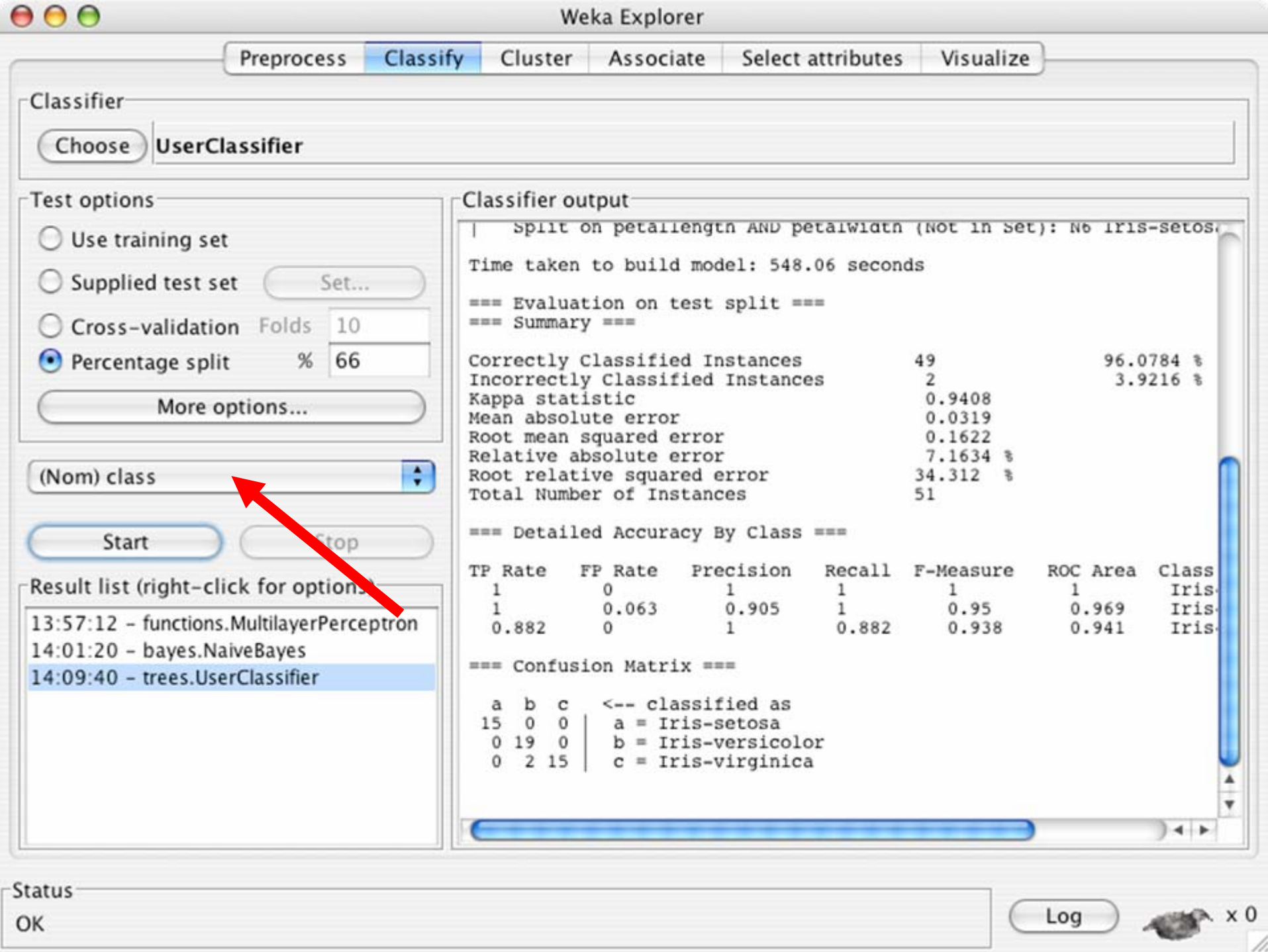


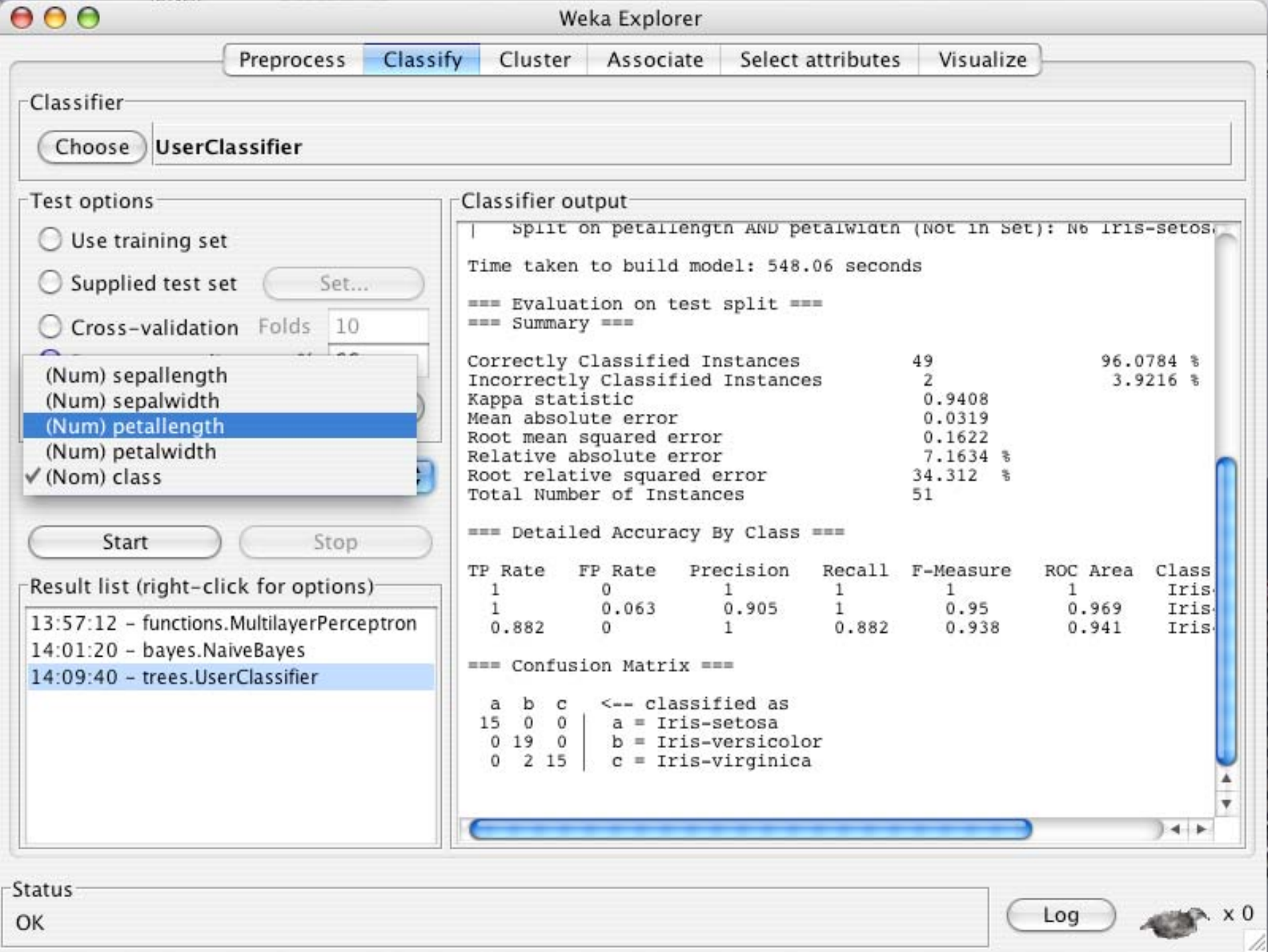
Class colour

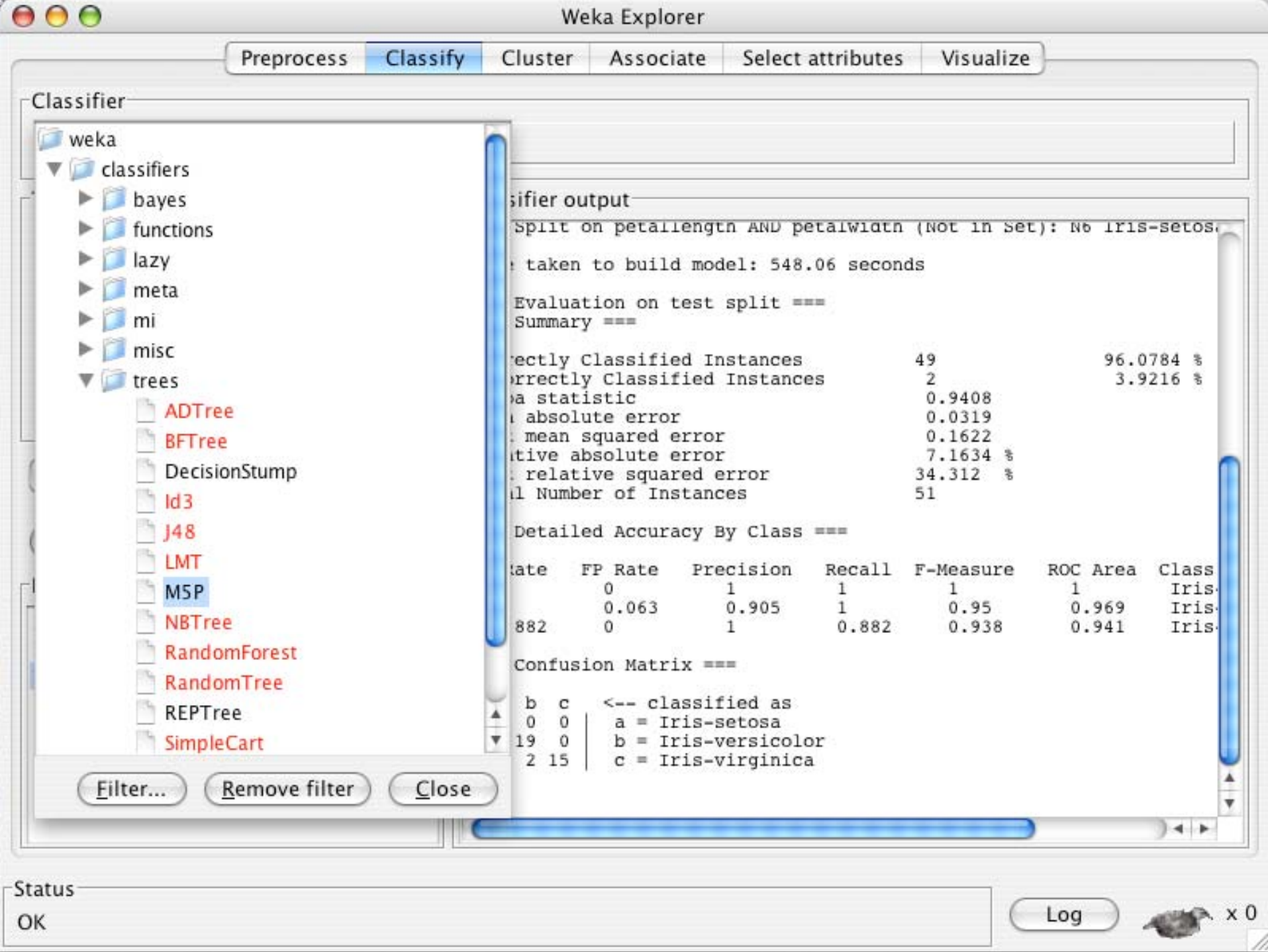
Iris-setosa Iris-versicolor Iris-virginica

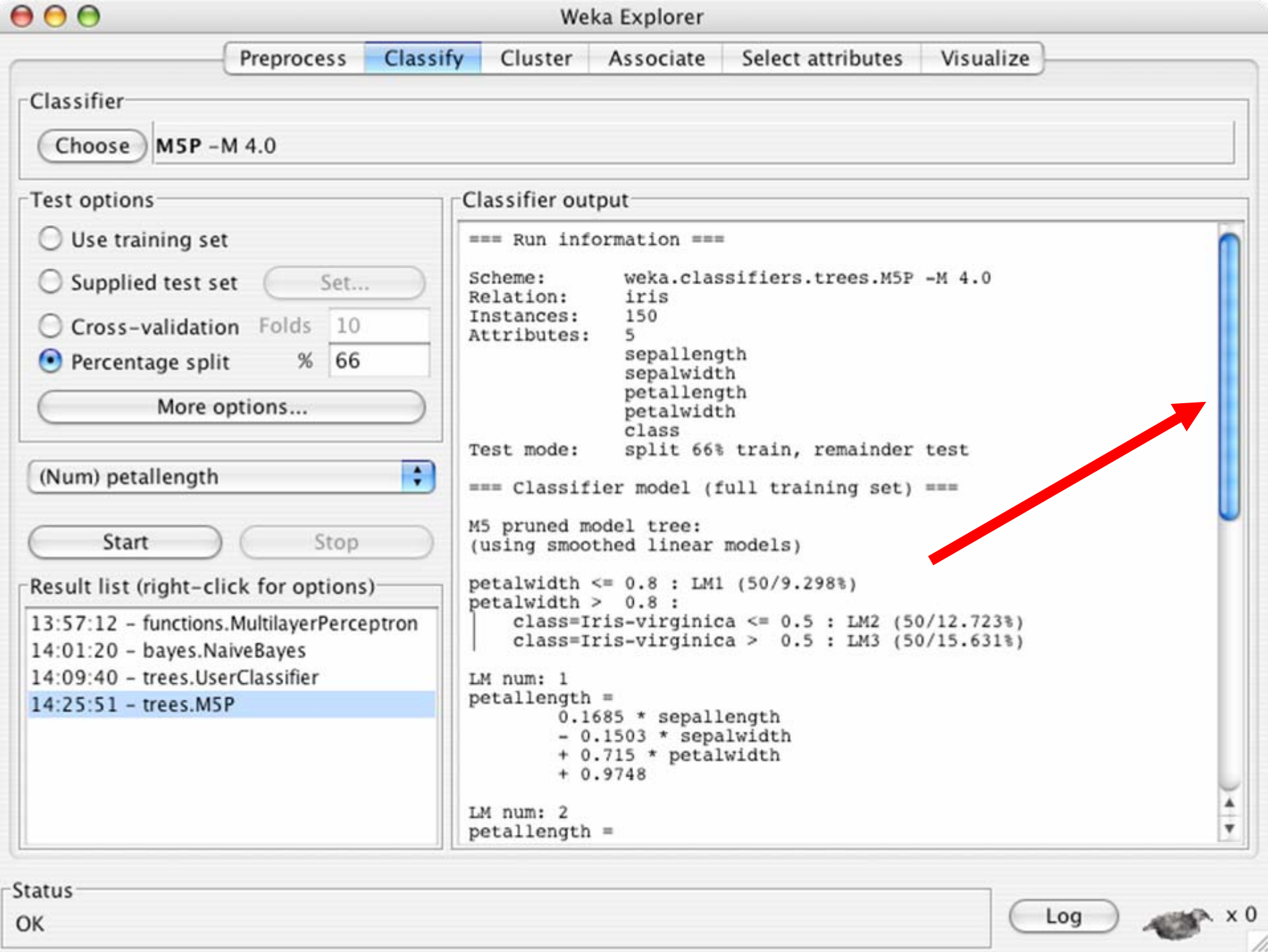

Tree Visualizer
Data Visualizer
Tree View
Split on
petallength AND
petalwidth
True
False
[Iris-virginica, 48.0]
[Iris-setosa, 50.0]
[Iris-versicolor, 50.0]
[Iris-virginica, 2.0]

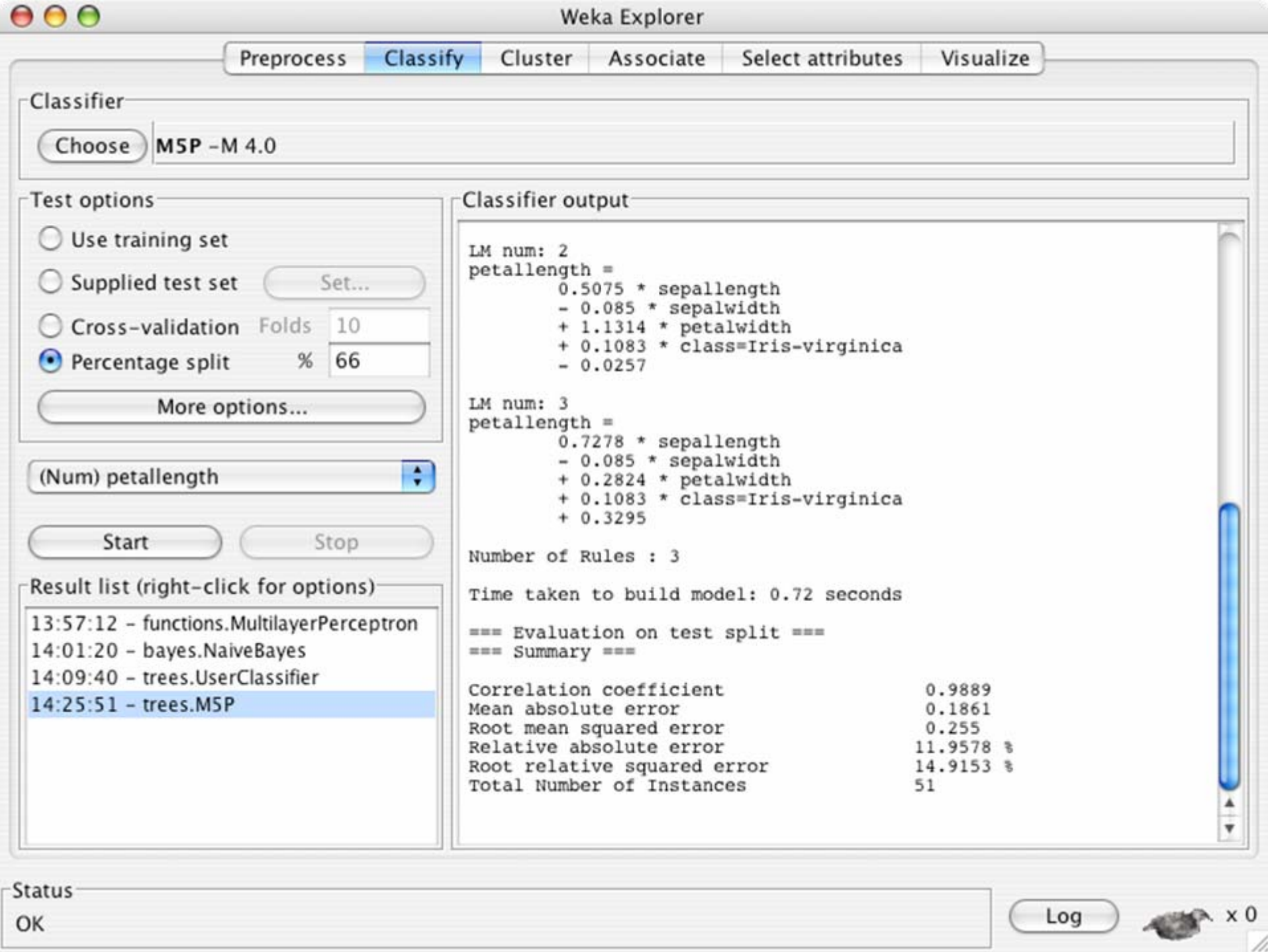

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
Classifier output
| Split on petallength AND petalwidth (Not in Set): N6 Iris-setos
Time taken to build model: 548.06 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0319
Root mean squared error 0.1622
Relative absolute error 7.1634 %
Root relative squared error 34.312 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.941 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

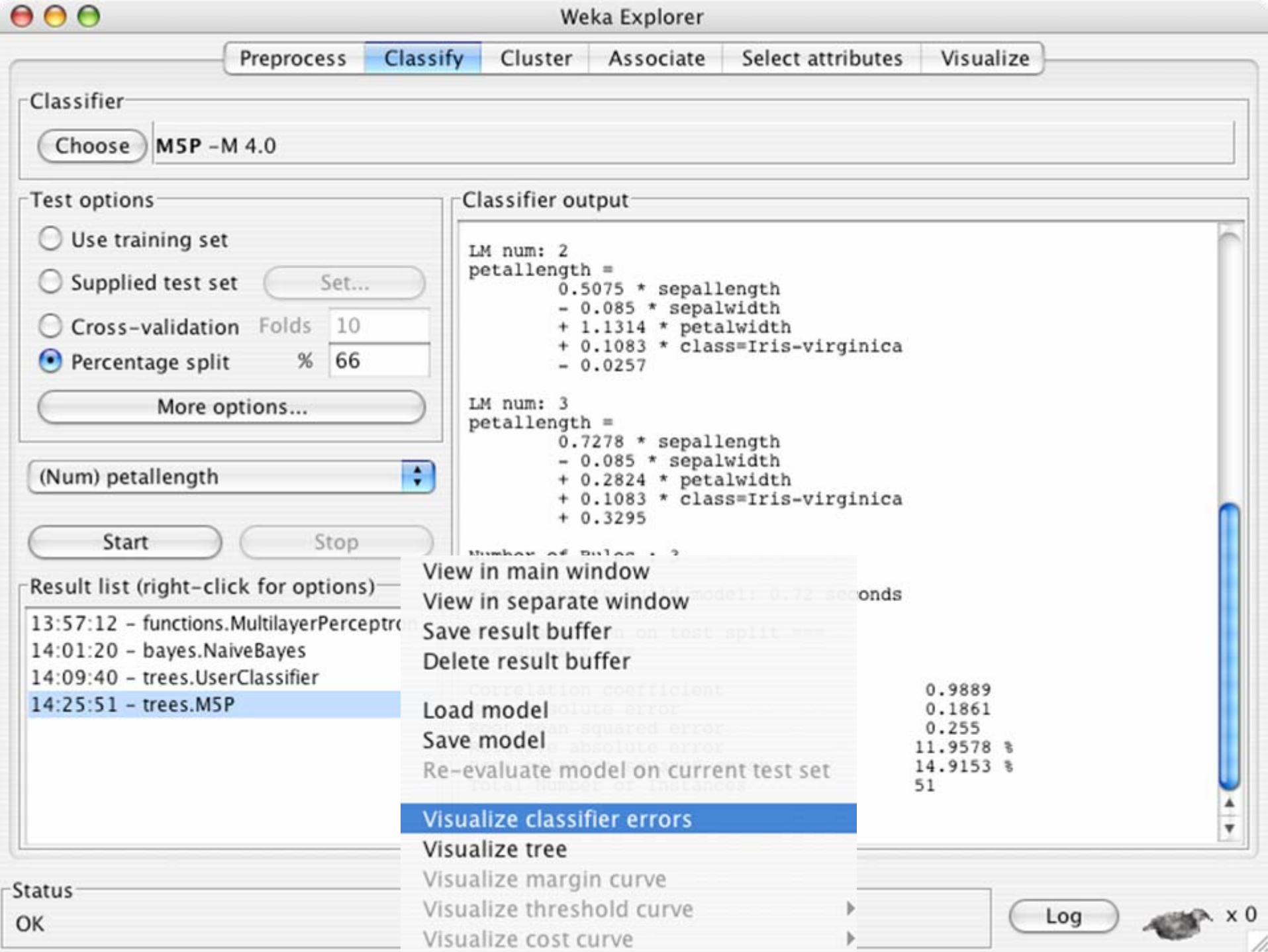

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
Classifier output
| Split on petallength AND petalwidth (Not in Set): N6 Iris-setos
Time taken to build model: 548.06 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0319
Root mean squared error 0.1622
Relative absolute error 7.1634 %
Root relative squared error 34.312 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.941 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

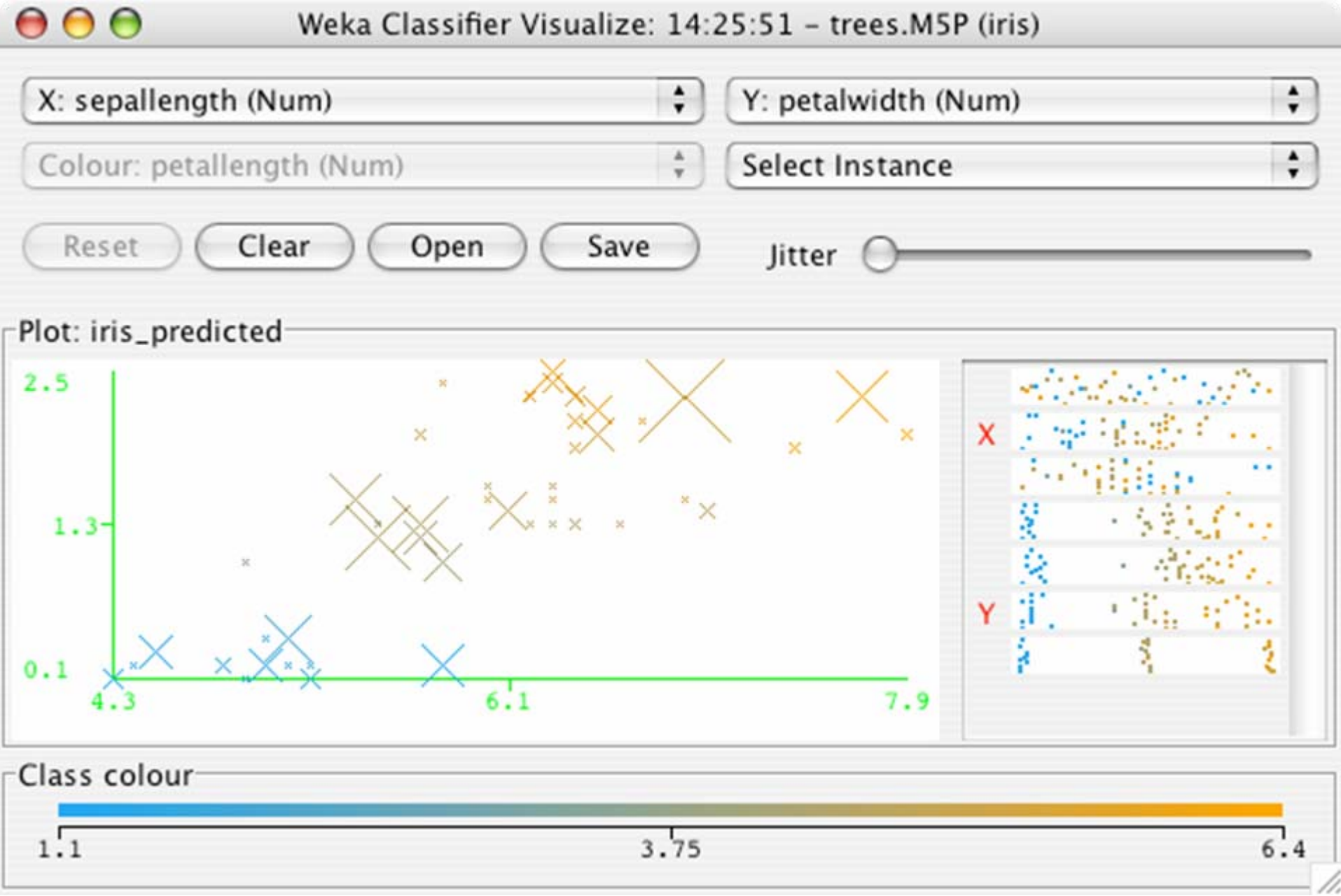

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
(Num) sepallength
(Num) sepalwidth
(Num) petallength
(Num) petalwidth
(Nom) class
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
Classifier output
| Split on petallength AND petalwidth (Not in Set): N6 Iris-setos
Time taken to build model: 548.06 seconds
=== Evaluation on test split ===
=== Summary ===
Correctly Classified Instances 49 96.0784 %
Incorrectly Classified Instances 2 3.9216 %
Kappa statistic 0.9408
Mean absolute error 0.0319
Root mean squared error 0.1622
Relative absolute error 7.1634 %
Root relative squared error 34.312 %
Total Number of Instances 51
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris
1 0.063 0.905 1 0.95 0.969 Iris
0.882 0 1 0.882 0.938 0.941 Iris
=== Confusion Matrix ===
a b c <-- classified as
15 0 0 | a = Iris-setosa
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

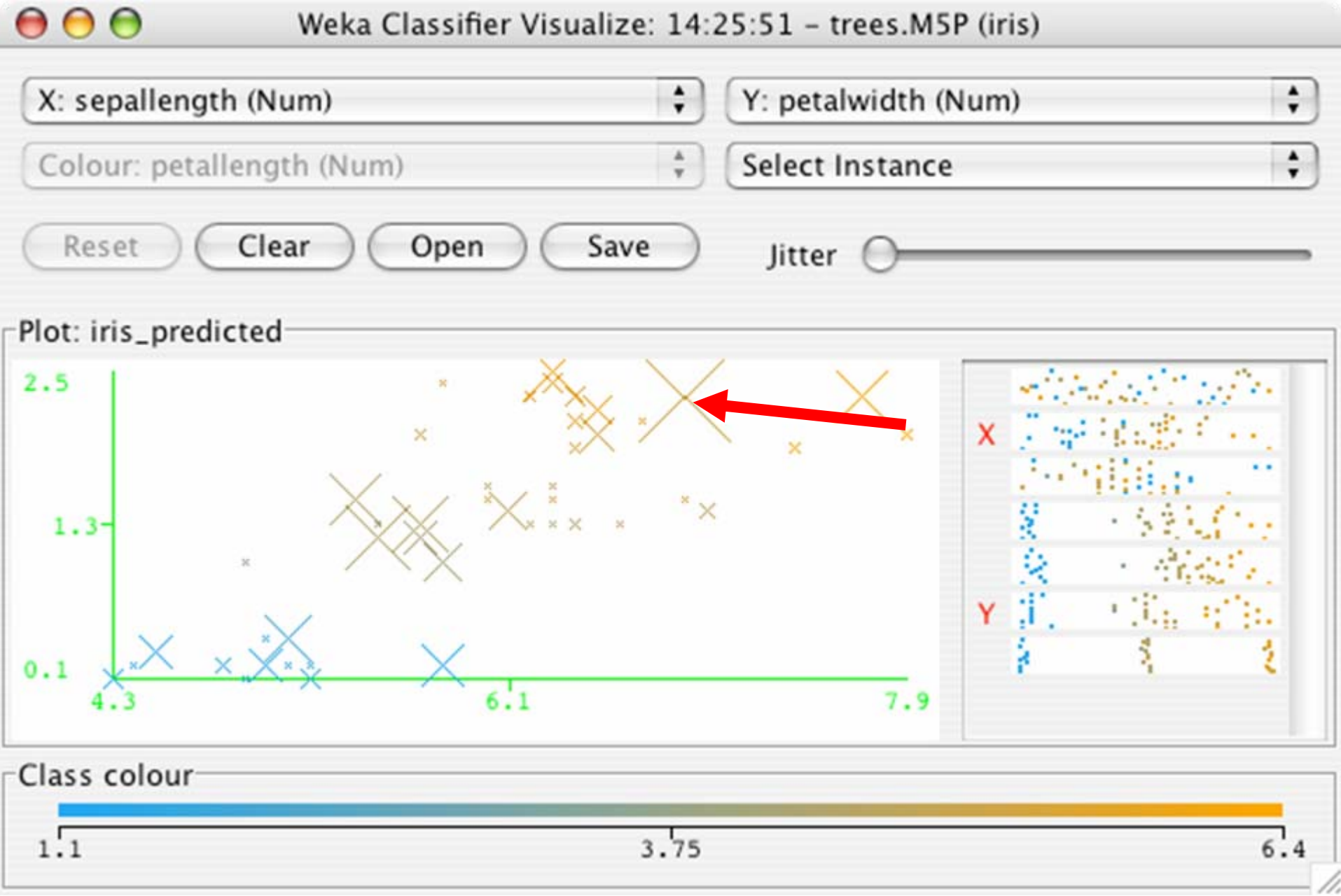

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
mi
misc
trees
ADTree
BFTree
DecisionStump
Id3
J48
LMT
M5P
NBTree
RandomForest
RandomTree
REPTree
SimpleCart
Filter...
Remove filter
Close
sifier output
Split on petallength AND petalwidth (Not In Set): N6 Iris-setos
taken to build model: 548.06 seconds
Evaluation on test split ===
Summary ===
ectly Classified Instances 49 96.0784 %
rrectly Classified Instances 2 3.9216 %
a statistic 0.9408
absolute error 0.0319
mean squared error 0.1622
tive absolute error 7.1634 %
relative squared error 34.312 %
l Number of Instances 51
Detailed Accuracy By Class ===
ate FP Rate Precision Recall F-Measure ROC Area Class
0 1 1 1 1 Iris
0.063 0.905 1 0.95 0.969 Iris
882 0 1 0.882 0.938 0.941 Iris
Confusion Matrix ===
b c <-- classified as
0 0 | a = Iris-setosa
19 0 | b = Iris-versicolor
2 15 | c = Iris-virginica
Status
OK
Log
x 0

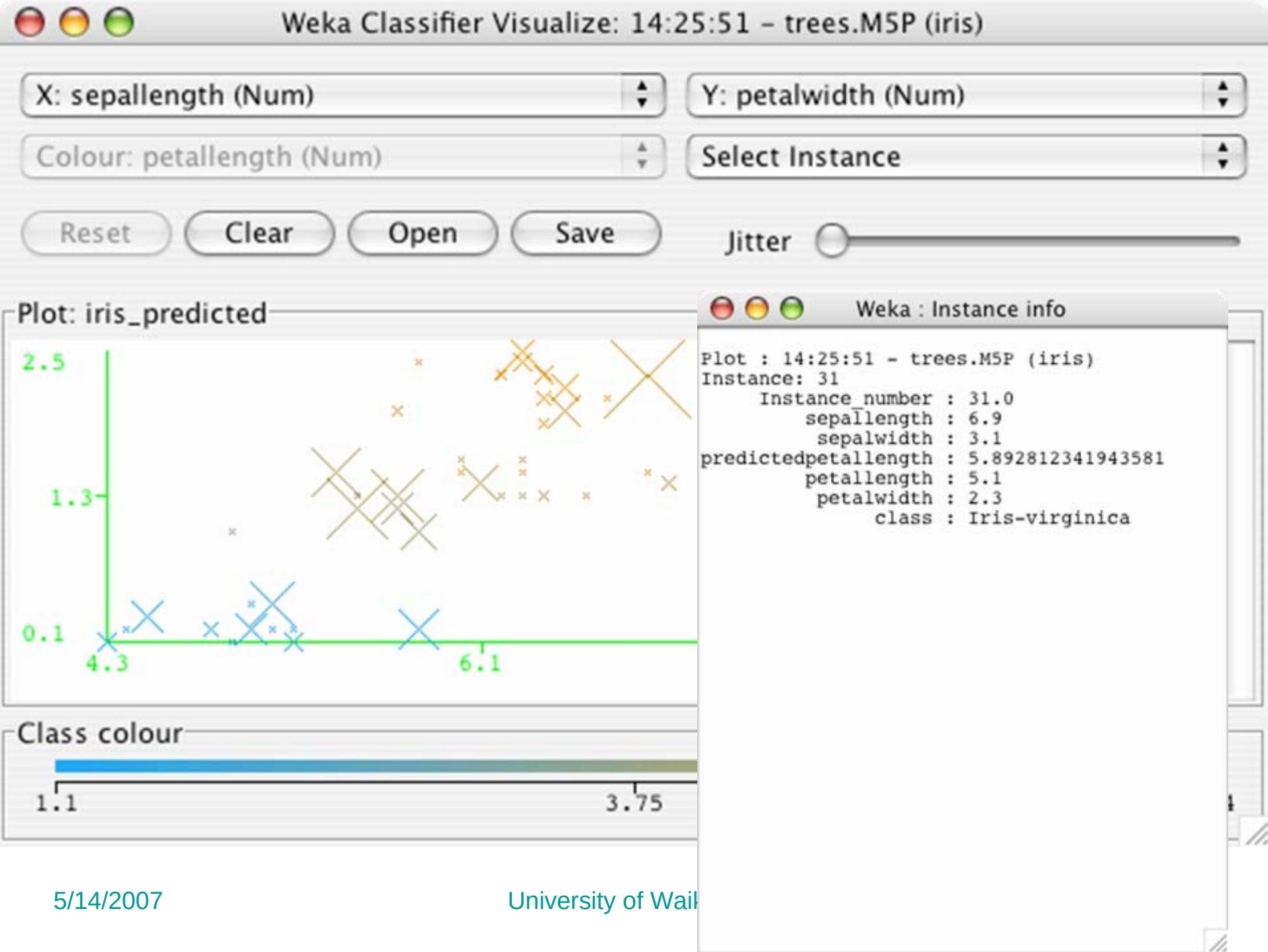

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P -M 4.0
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Num) petallength
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
14:25:51 - trees.M5P
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.M5P -M 4.0
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
class
Test mode: split 66% train, remainder test
=== Classifier model (full training set) ===
M5 pruned model tree:
(using smoothed linear models)
petalwidth <= 0.8 : LM1 (50/9.298%)
petalwidth > 0.8 :
| class=Iris-virginica <= 0.5 : LM2 (50/12.723%)
| class=Iris-virginica > 0.5 : LM3 (50/15.631%)
LM num: 1
petallength =
0.1685 * sepallength
- 0.1503 * sepalwidth
+ 0.715 * petalwidth
+ 0.9748
LM num: 2
petallength =
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P -M 4.0
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Num) petallength
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptron
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
14:25:51 - trees.M5P
Classifier output
LM num: 2
petallength =
0.5075 * sepallength
- 0.085 * sepalwidth
+ 1.1314 * petalwidth
+ 0.1083 * class=Iris-virginica
- 0.0257
LM num: 3
petallength =
0.7278 * sepallength
- 0.085 * sepalwidth
+ 0.2824 * petalwidth
+ 0.1083 * class=Iris-virginica
+ 0.3295
Number of Rules : 3
Time taken to build model: 0.72 seconds
=== Evaluation on test split ===
=== Summary ===
Correlation coefficient 0.9889
Mean absolute error 0.1861
Root mean squared error 0.255
Relative absolute error 11.9578 %
Root relative squared error 14.9153 %
Total Number of Instances 51
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P -M 4.0
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
LM num: 2
petallength =
0.5075 * sepallength
- 0.085 * sepalwidth
+ 1.1314 * petalwidth
+ 0.1083 * class=Iris-virginica
- 0.0257
LM num: 3
petallength =
0.7278 * sepallength
- 0.085 * sepalwidth
+ 0.2824 * petalwidth
+ 0.1083 * class=Iris-virginica
+ 0.3295
(Num) petallength
Start
Stop
Result list (right-click for options)
13:57:12 - functions.MultilayerPerceptr
14:01:20 - bayes.NaiveBayes
14:09:40 - trees.UserClassifier
14:25:51 - trees.M5P
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifier errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
0.9889
0.1861
0.255
11.9578 %
14.9153 %
51
Status
OK
Log
x 0

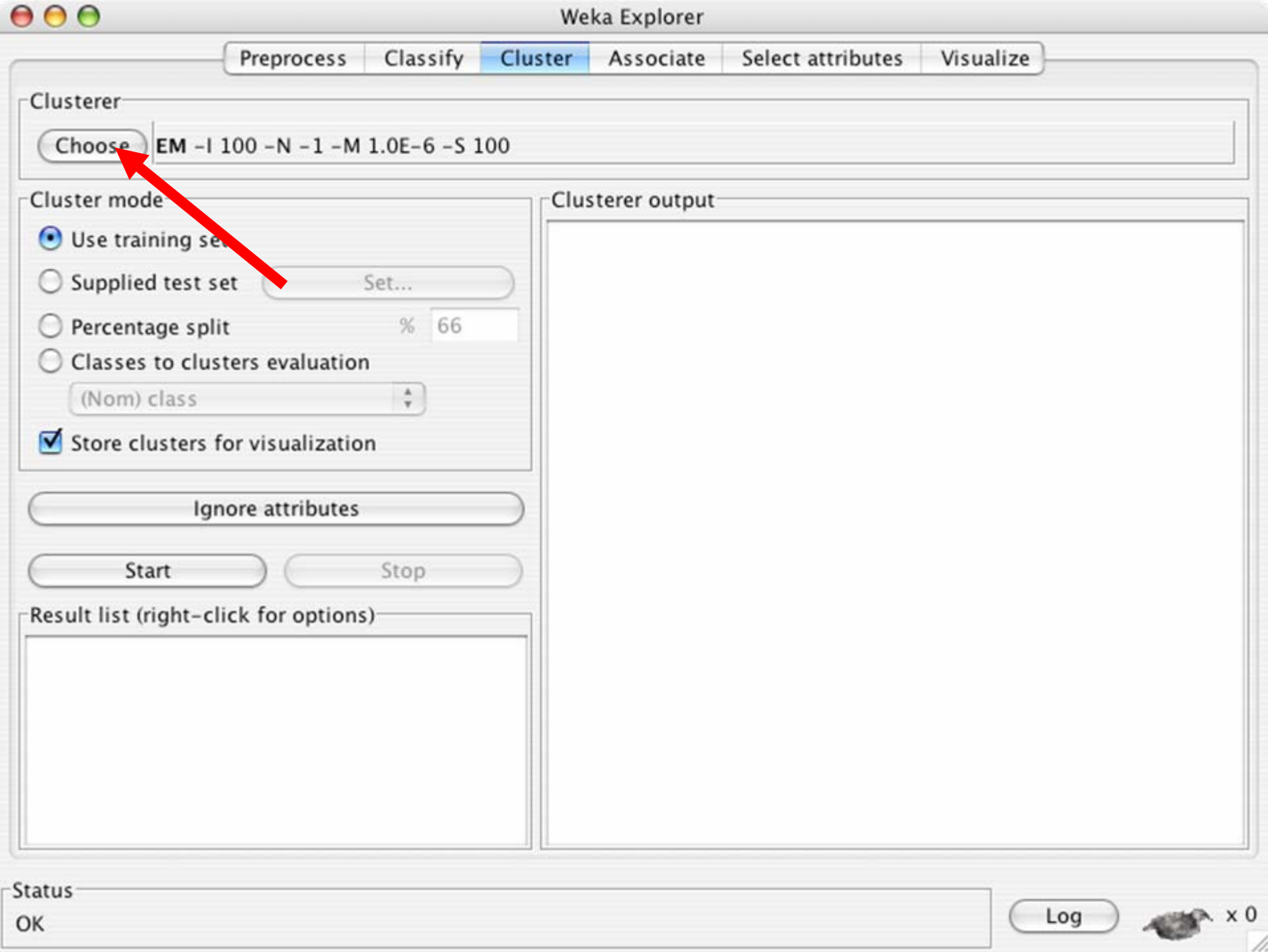

Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)
X: sepallength (Num)
Y: petalwidth (Num)
Colour: petallength (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_predicted
2.5
1.3
0.1
4.3
6.1
7.9
X
Y
Class colour
1.1
3.75
6.4

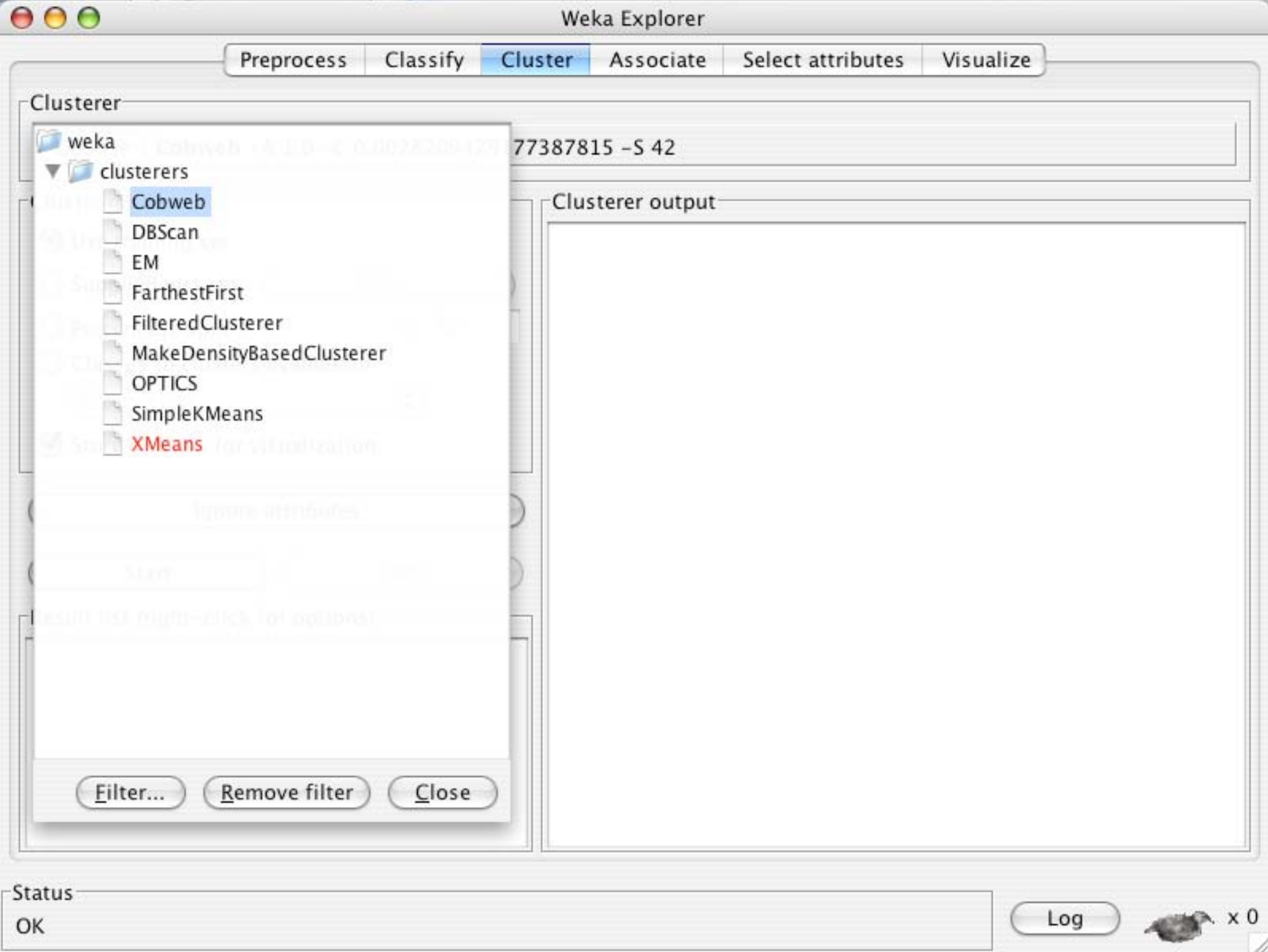

Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)
X: sepallength (Num)
Y: petalwidth (Num)
Colour: petallength (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_predicted
2.5
1.3
0.1
4.3
6.1
7.9
X
Y
Class colour
1.1
3.75
6.4

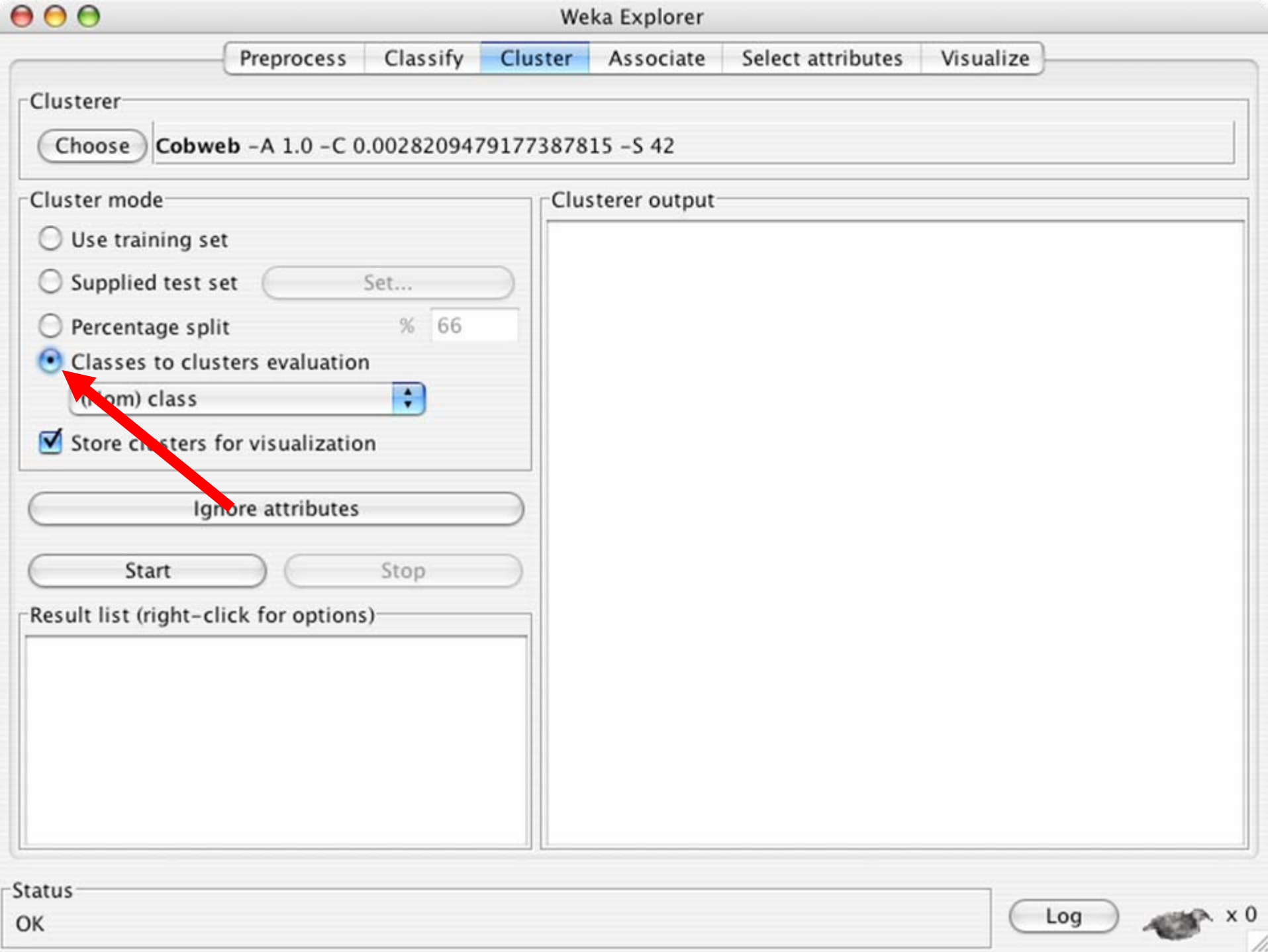

Weka Classifier Visualize: 14:25:51 - trees.M5P (iris)
X: sepallength (Num)
Y: petalwidth (Num)
Colour: petallength (Num)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_predicted
2.5
1.3
0.1
4.3
6.1
Class colour
1.1
3.75
Weka : Instance info
Plot : 14:25:51 - trees.M5P (iris)
Instance: 31
Instance_number : 31.0
sepallength : 6.9
sepalwidth : 3.1
predictedpetallength : 5.892812341943581
petallength : 5.1
petalwidth : 2.3
class : Iris-virginica

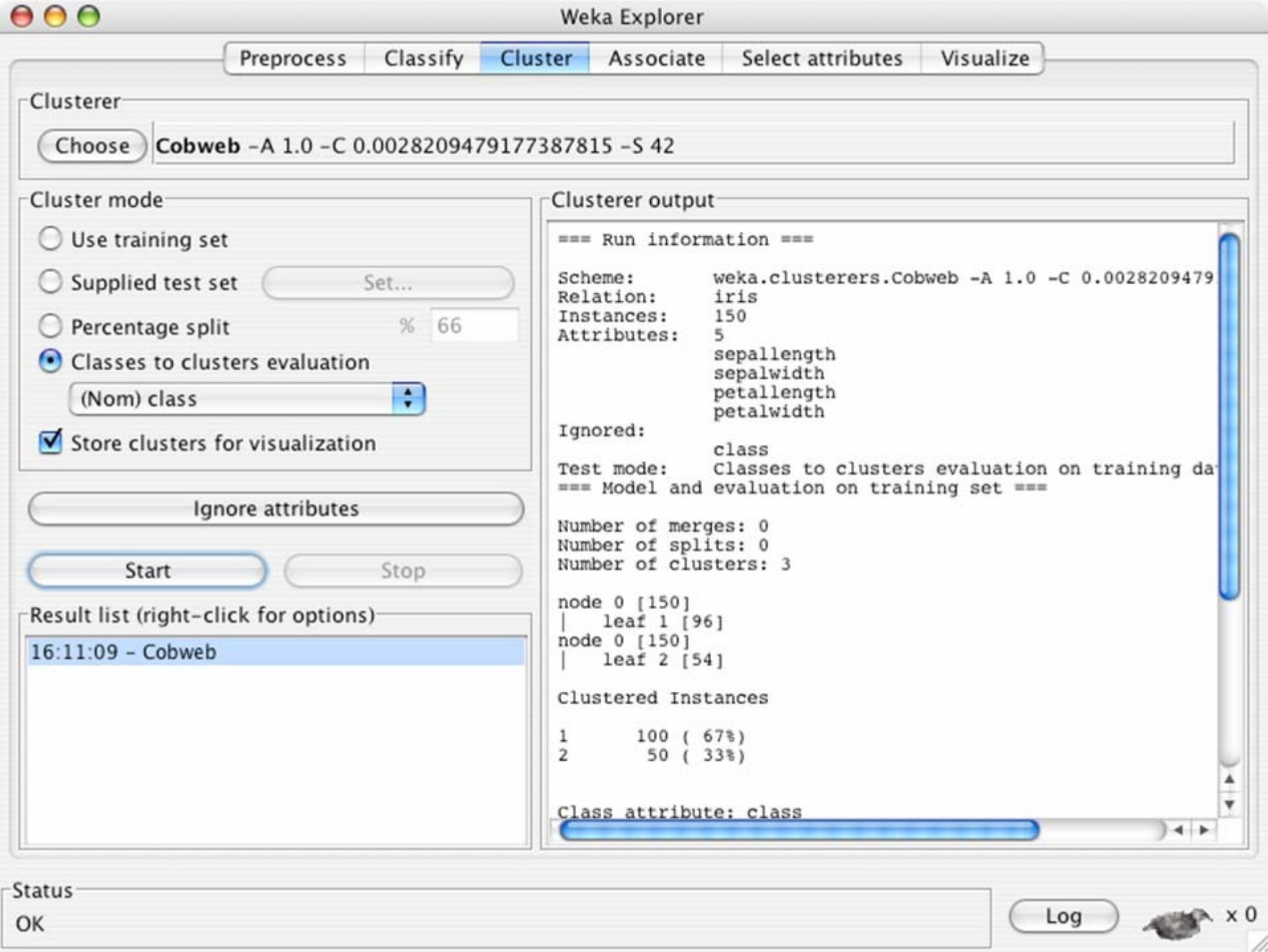
Explorer: clustering data

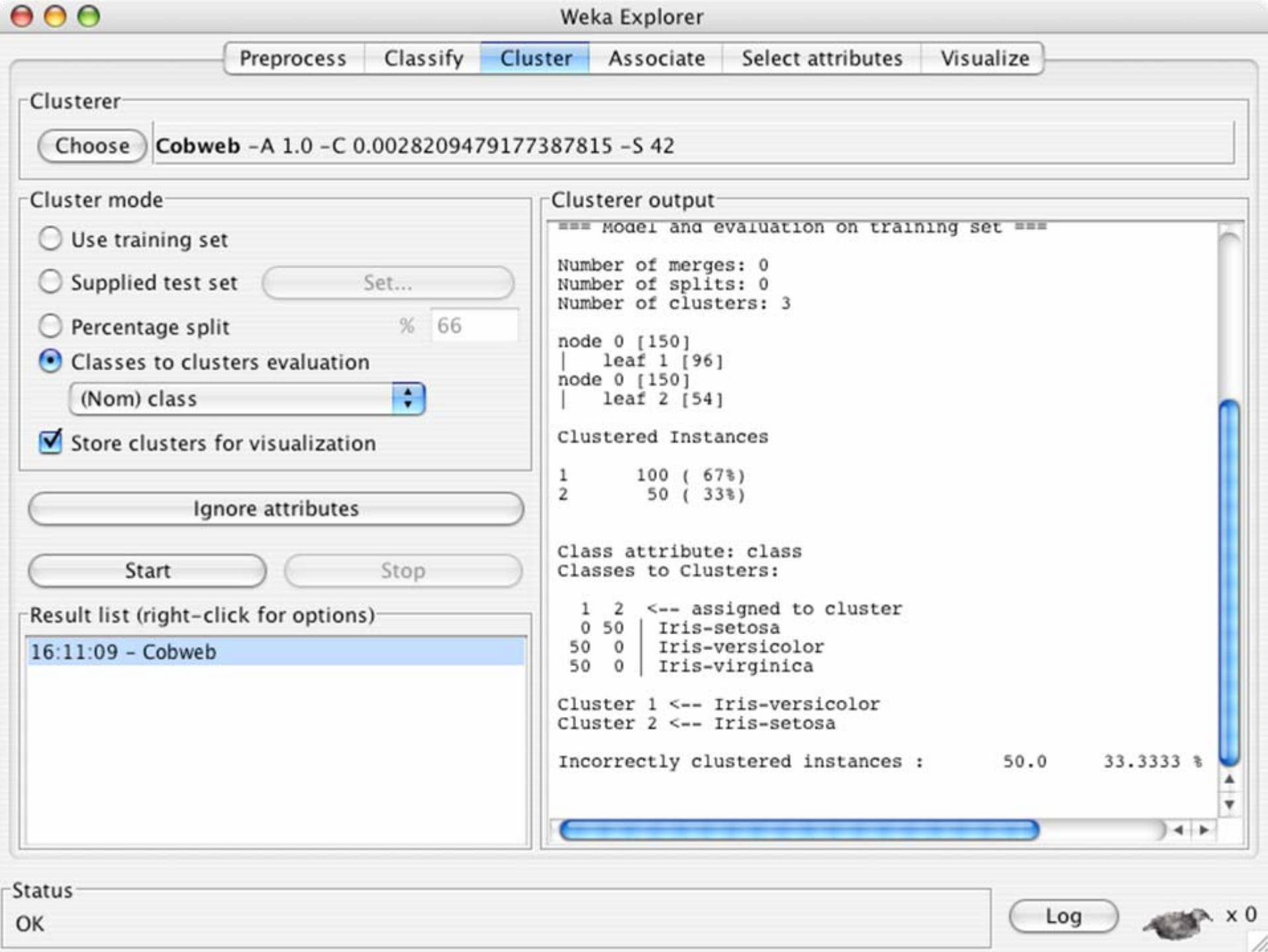
- WEKA contains “clusterers” for finding groups of similar instances in a dataset
- Some implemented schemes are:
  - *k*-Means, EM, Cobweb, *X*-means, FarthestFirst
- Clusters can be visualized and compared to “true” clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a probability distribution

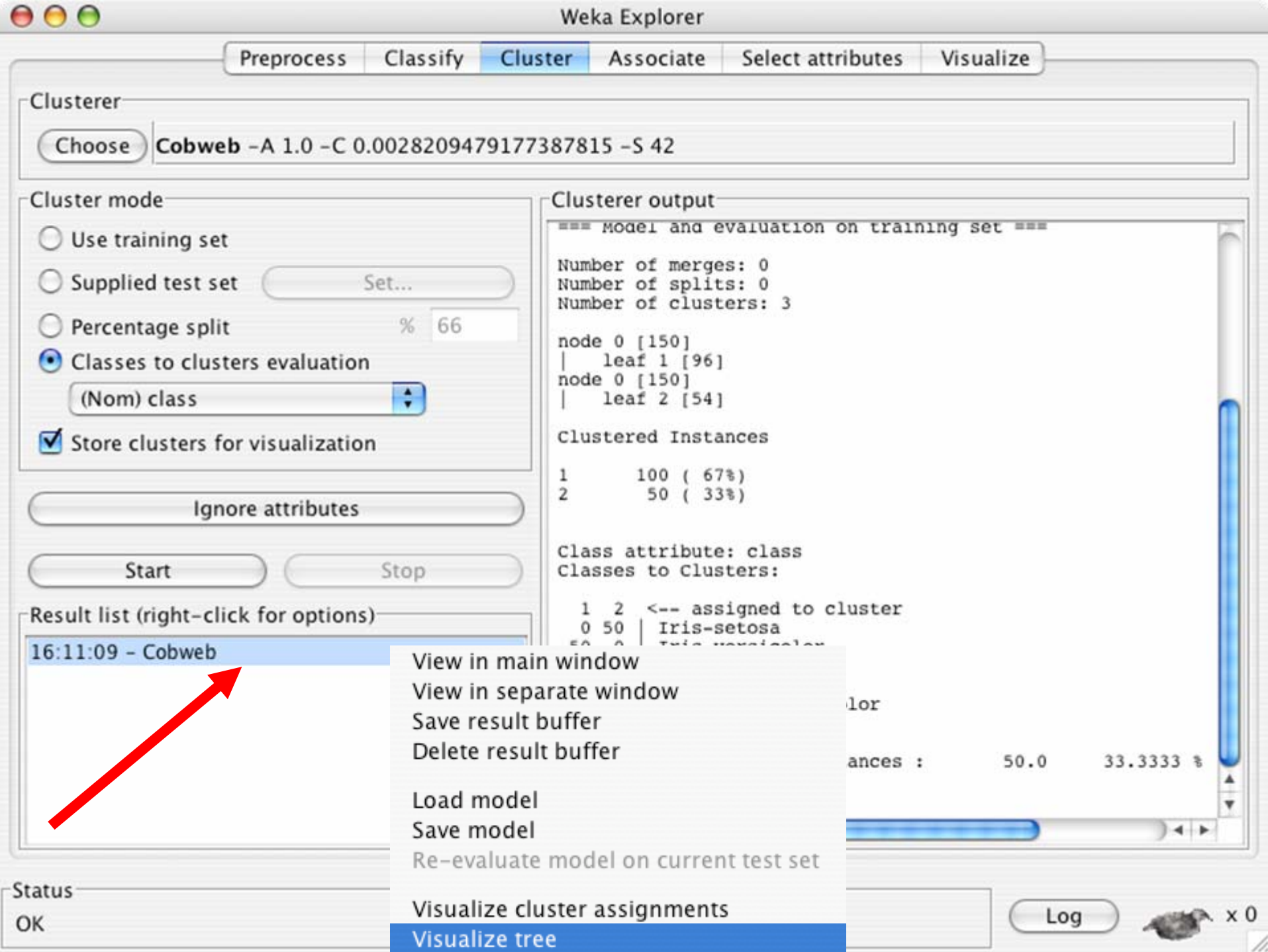

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -M 1.0E-6 -S 100
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
Clusterer output
Status
OK
Log
x 0

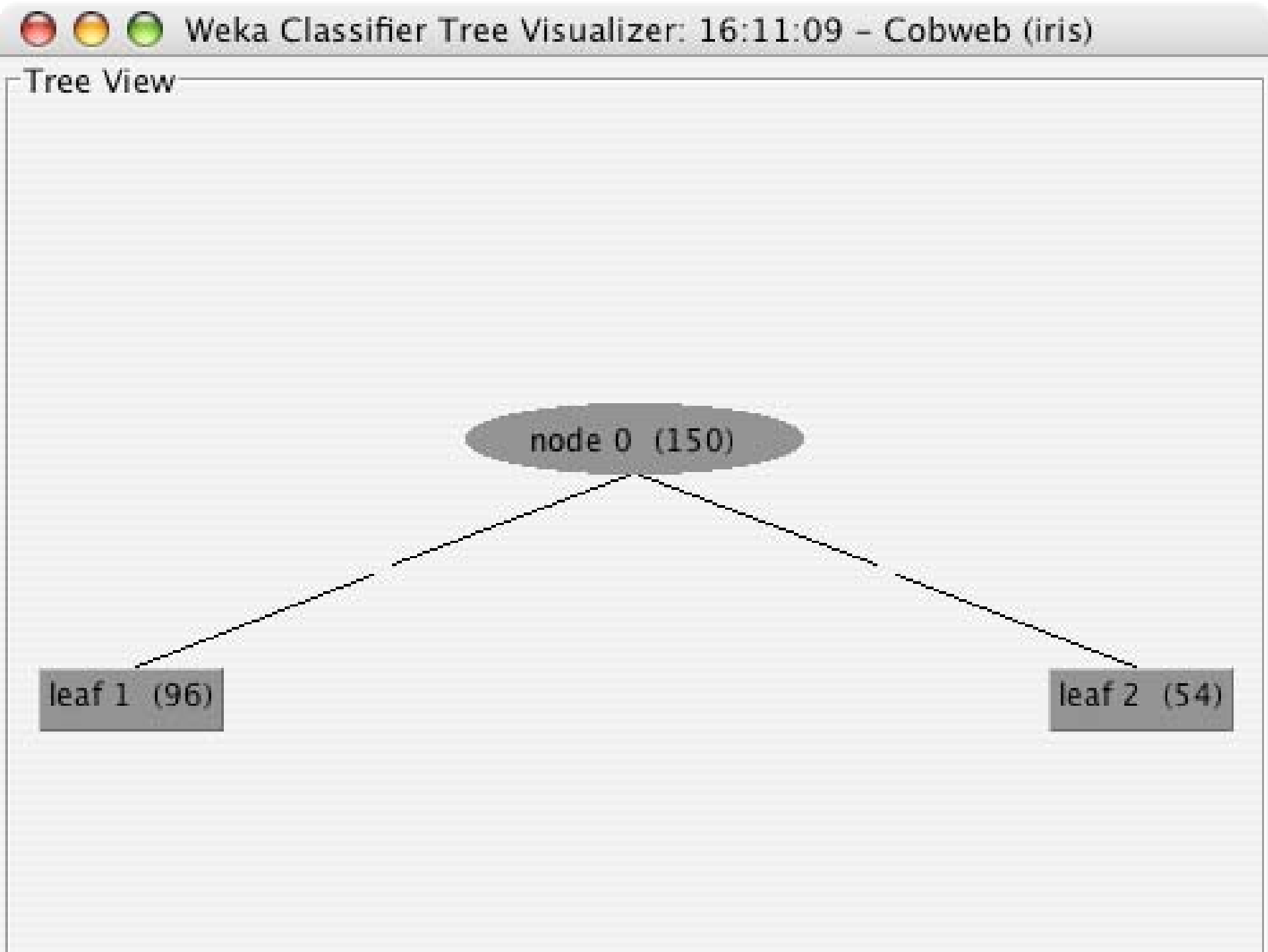

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
weka
clusterers
Cobweb
DBScan
EM
FarthestFirst
FilteredClusterer
MakeDensityBasedClusterer
OPTICS
SimpleKMeans
XMeans
77387815 -S 42
Clusterer output
Filter...
Remove filter
Close
Status
OK
Log
x 0

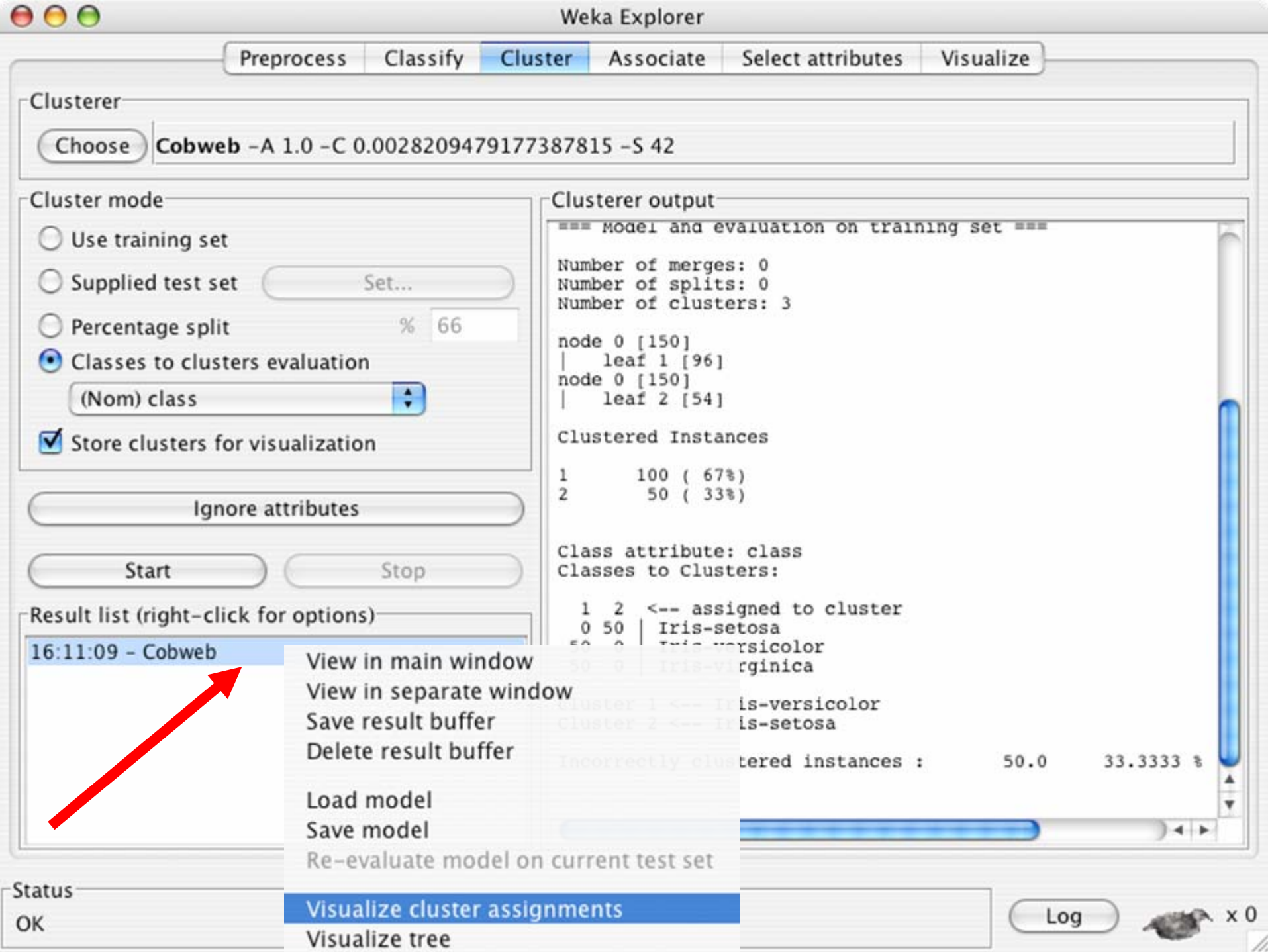

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
Clusterer output
Status
OK
Log
x 0

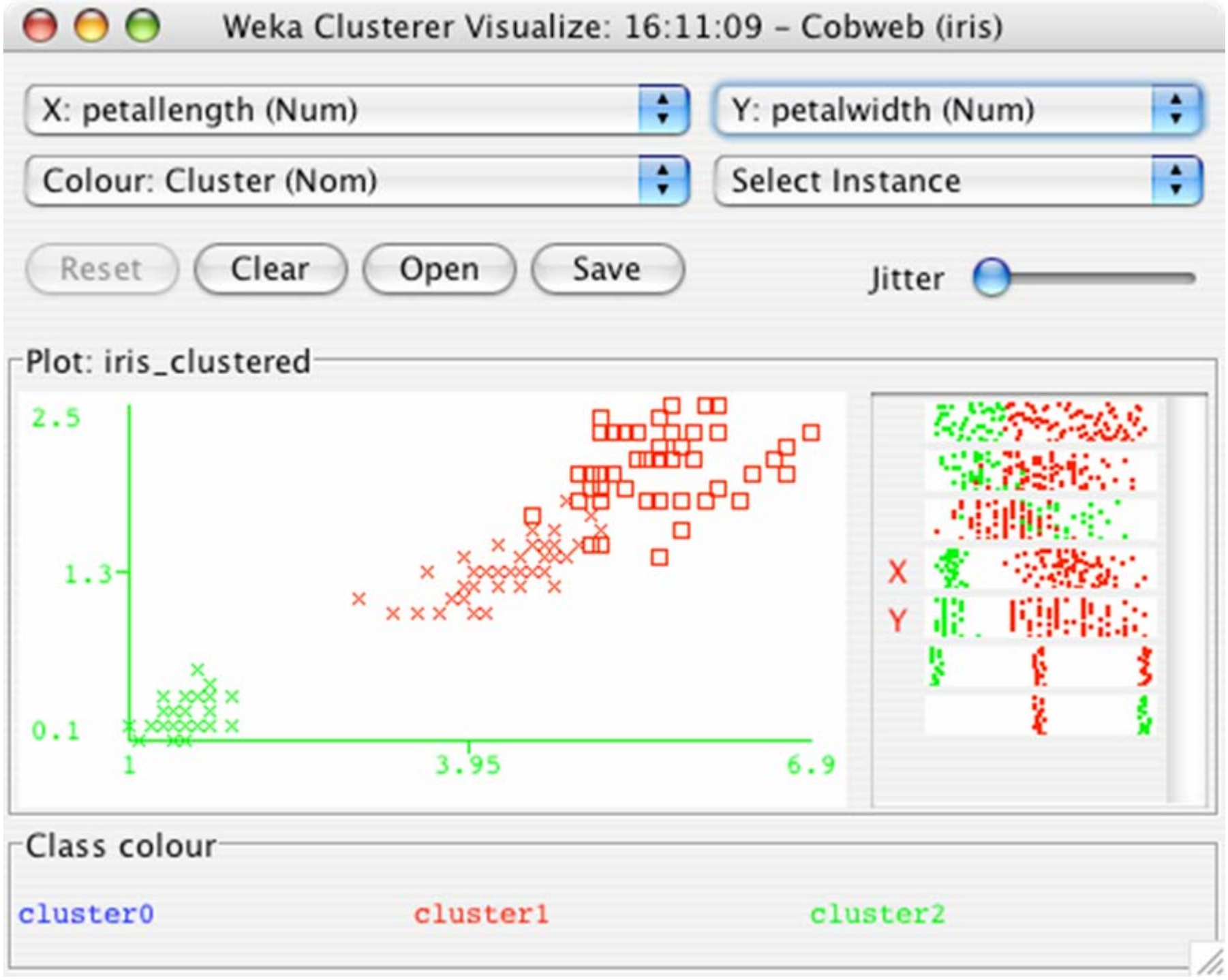

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:11:09 - Cobweb
Clusterer output
=== Run information ===
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.0028209479
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
Ignored:
class
Test mode: Classes to clusters evaluation on training da
=== Model and evaluation on training set ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
node 0 [150]
| leaf 1 [96]
node 0 [150]
| leaf 2 [54]
Clustered Instances
1 100 (67%)
2 50 (33%)
Class attribute: class
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
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)
16:11:09 - Cobweb
Clusterer output
=== Model and evaluation on training set ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
node 0 [150]
| leaf 1 [96]
node 0 [150]
| leaf 2 [54]
Clustered Instances
1 100 (67%)
2 50 (33%)
Class attribute: class
Classes to Clusters:
1 2 <-- assigned to cluster
0 50 | Iris-setosa
50 0 | Iris-versicolor
50 0 | Iris-virginica
Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa
Incorrectly clustered instances : 50.0 33.3333 %
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
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)
16:11:09 - Cobweb
Clusterer output
=== Model and evaluation on training set ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
node 0 [150]
| leaf 1 [96]
node 0 [150]
| leaf 2 [54]
Clustered Instances
1 100 (67%)
2 50 (33%)
Class attribute: class
Classes to Clusters:
1 2 <-- assigned to cluster
0 50 | Iris-setosa
lor
ances : 50.0 33.3333 %
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize cluster assignments
Visualize tree
Status
OK
Log
x 0

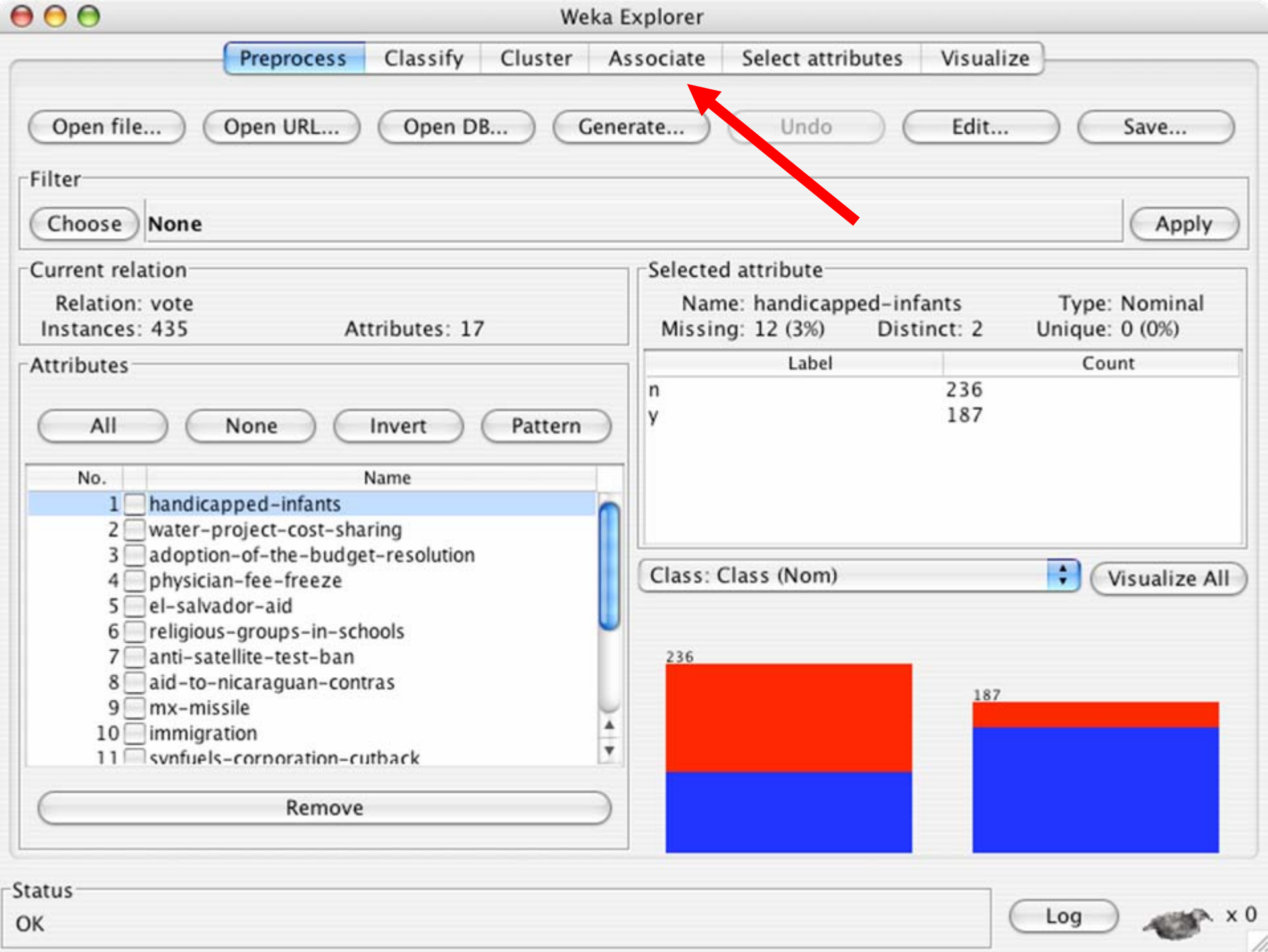

Weka Classifier Tree Visualizer: 16:11:09 - Cobweb (iris)
Tree View
node 0 (150)
leaf 1 (96)
leaf 2 (54)

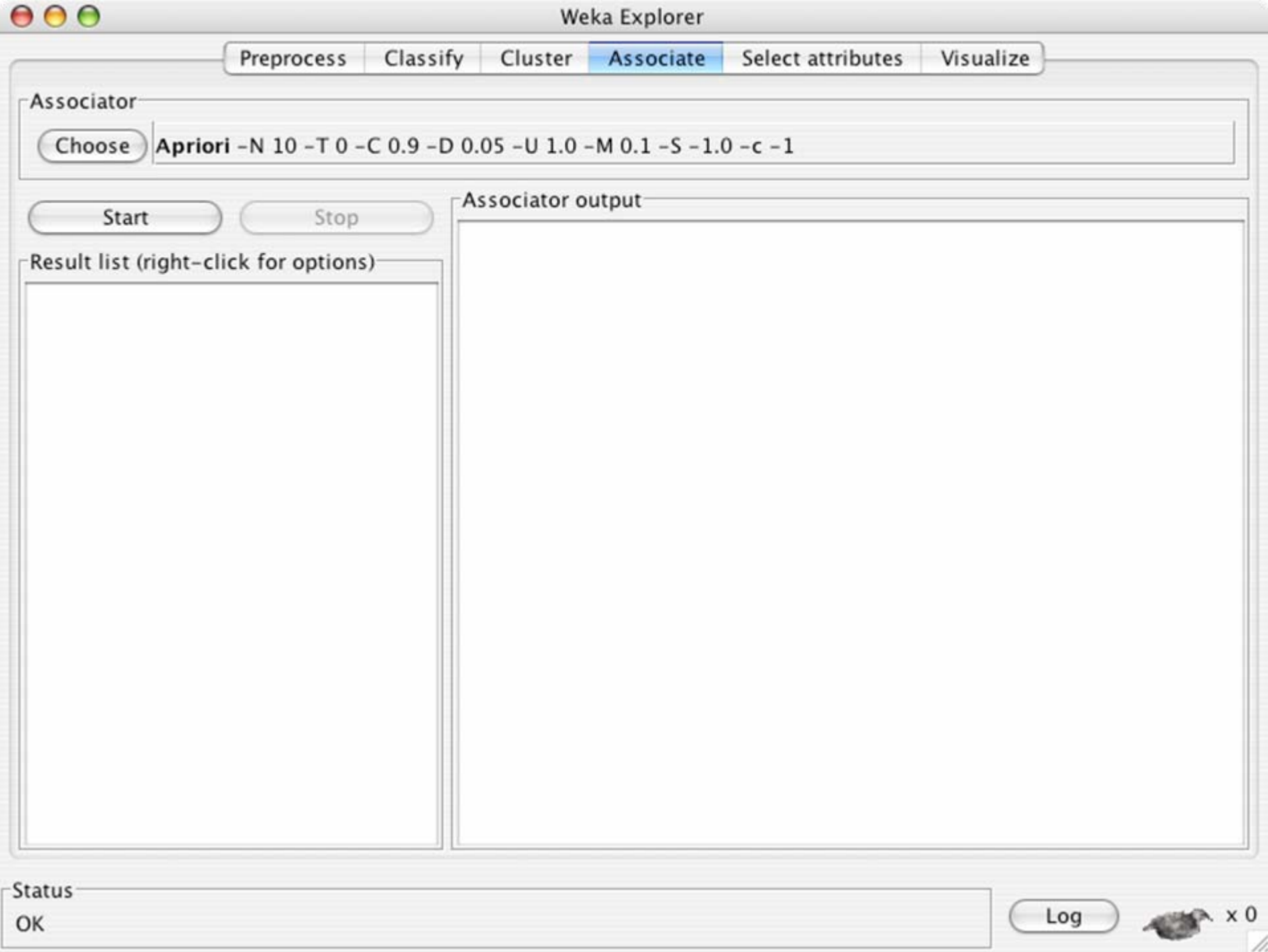

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:11:09 - Cobweb
Clusterer output
=== Model and evaluation on training set ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
node 0 [150]
| leaf 1 [96]
node 0 [150]
| leaf 2 [54]
Clustered Instances
1 100 (67%)
2 50 (33%)
Class attribute: class
Classes to Clusters:
1 2 <-- assigned to cluster
0 50 | Iris-setosa
Iris-versicolor
Iris-virginica
Cluster 1 <-- Iris-versicolor
Cluster 2 <-- Iris-setosa
Incorrectly clustered instances : 50.0 33.3333 %
View in main window
View in separate window
Save result buffer
Delete result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize cluster assignments
Visualize tree
Status
OK
Log
x 0

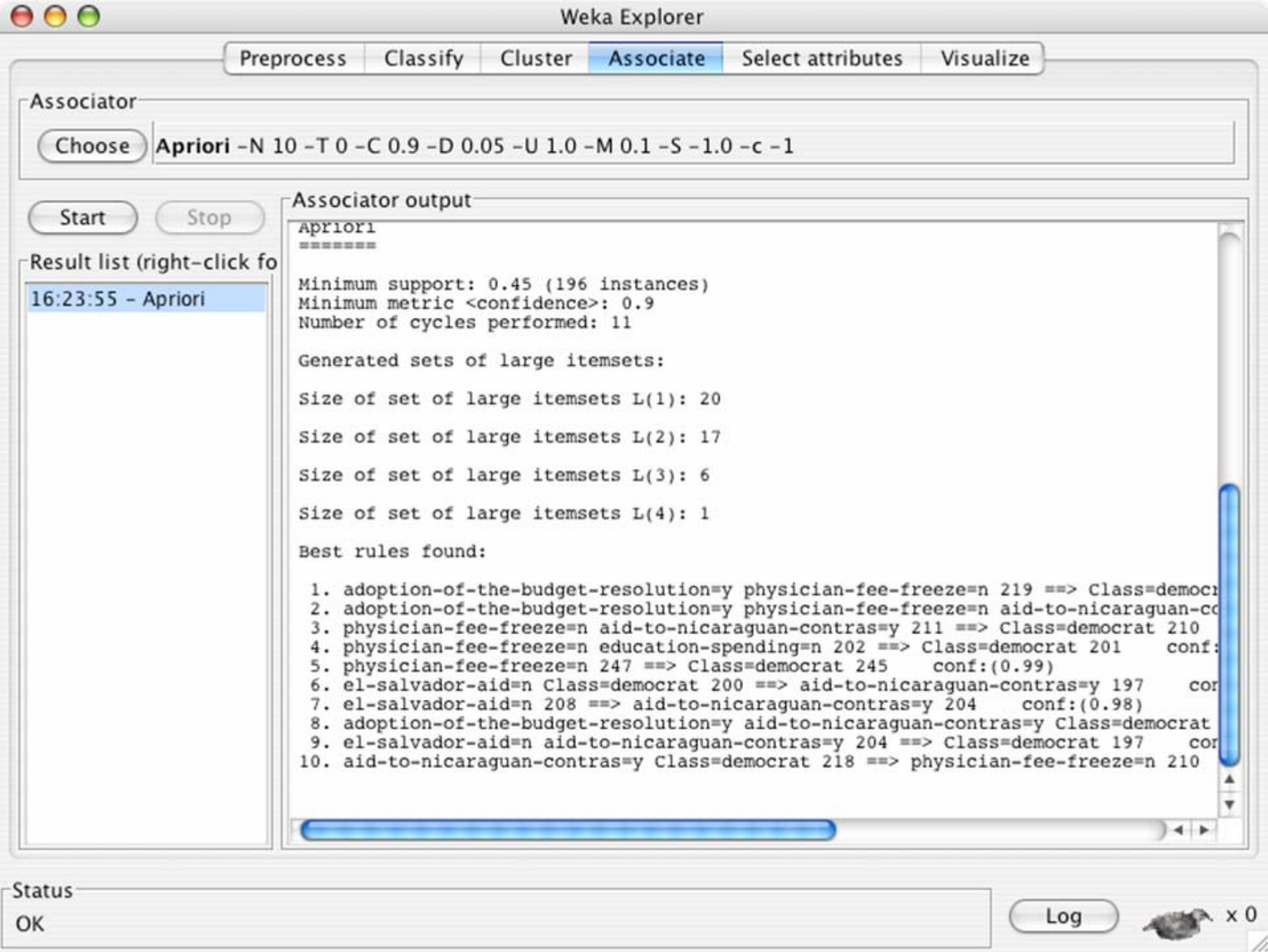

Weka Clusterer Visualize: 16:11:09 - Cobweb (iris)
X: petallength (Num)
Y: petalwidth (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Open
Save
Jitter
Plot: iris_clustered
2.5
1.3
0.1
1
3.95
6.9
X
Y
Class colour
cluster0
cluster1
cluster2

Explorer: finding associations

- WEKA contains the Apriori algorithm (among others) for learning association rules
  - Works only with discrete data
- Can identify statistical dependencies between groups of attributes:
  - milk, butter $\Rightarrow$ bread, eggs (with confidence 0.9 and support 2000)
- Apriori can compute all rules that have a given minimum support and exceed a given confidence

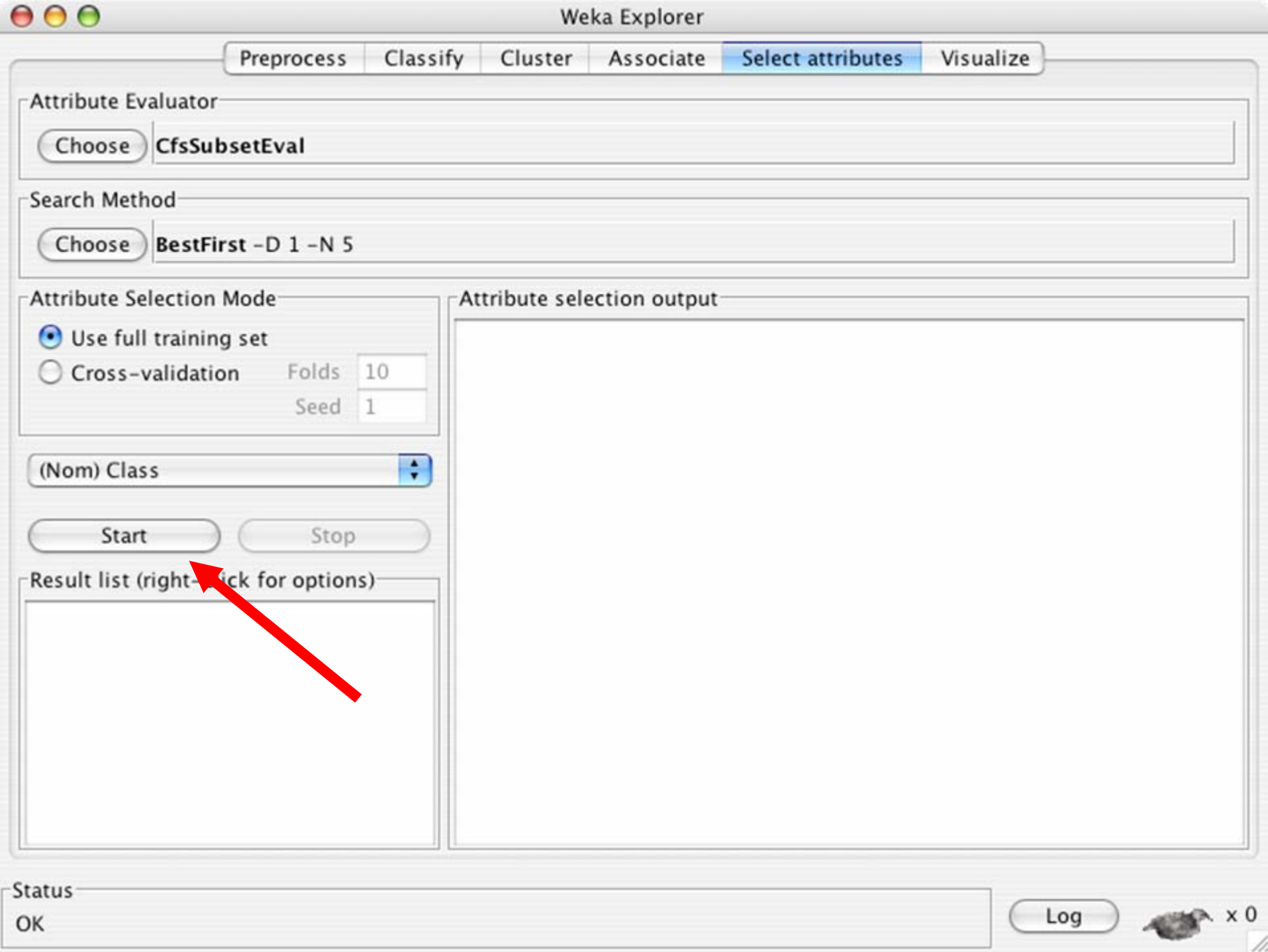

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Generate...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Current relation
Relation: vote
Instances: 435
Attributes: 17
Attributes
All
None
Invert
Pattern
No.
Name
1 handicapped-infants
2 water-project-cost-sharing
3 adoption-of-the-budget-resolution
4 physician-fee-freeze
5 el-salvador-aid
6 religious-groups-in-schools
7 anti-satellite-test-ban
8 aid-to-nicaraguan-contras
9 mx-missile
10 immigration
11 synfuels-corporation-cutback
Remove
Selected attribute
Name: handicapped-infants
Type: Nominal
Missing: 12 (3%)
Distinct: 2
Unique: 0 (0%)
Label
Count
n 236
y 187
Class: Class (Nom)
Visualize All
236
187
Status
OK
Log
x 0

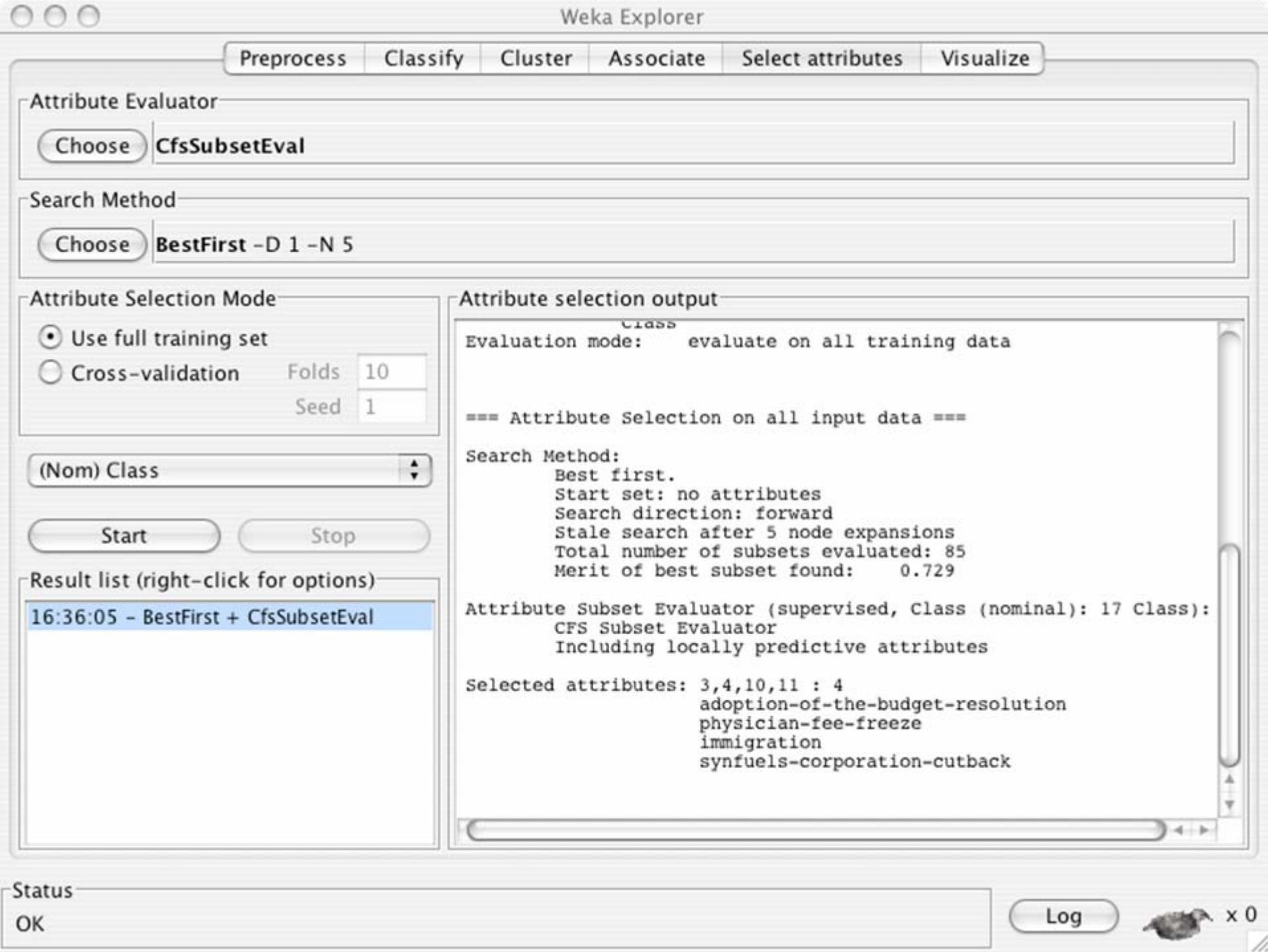

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0

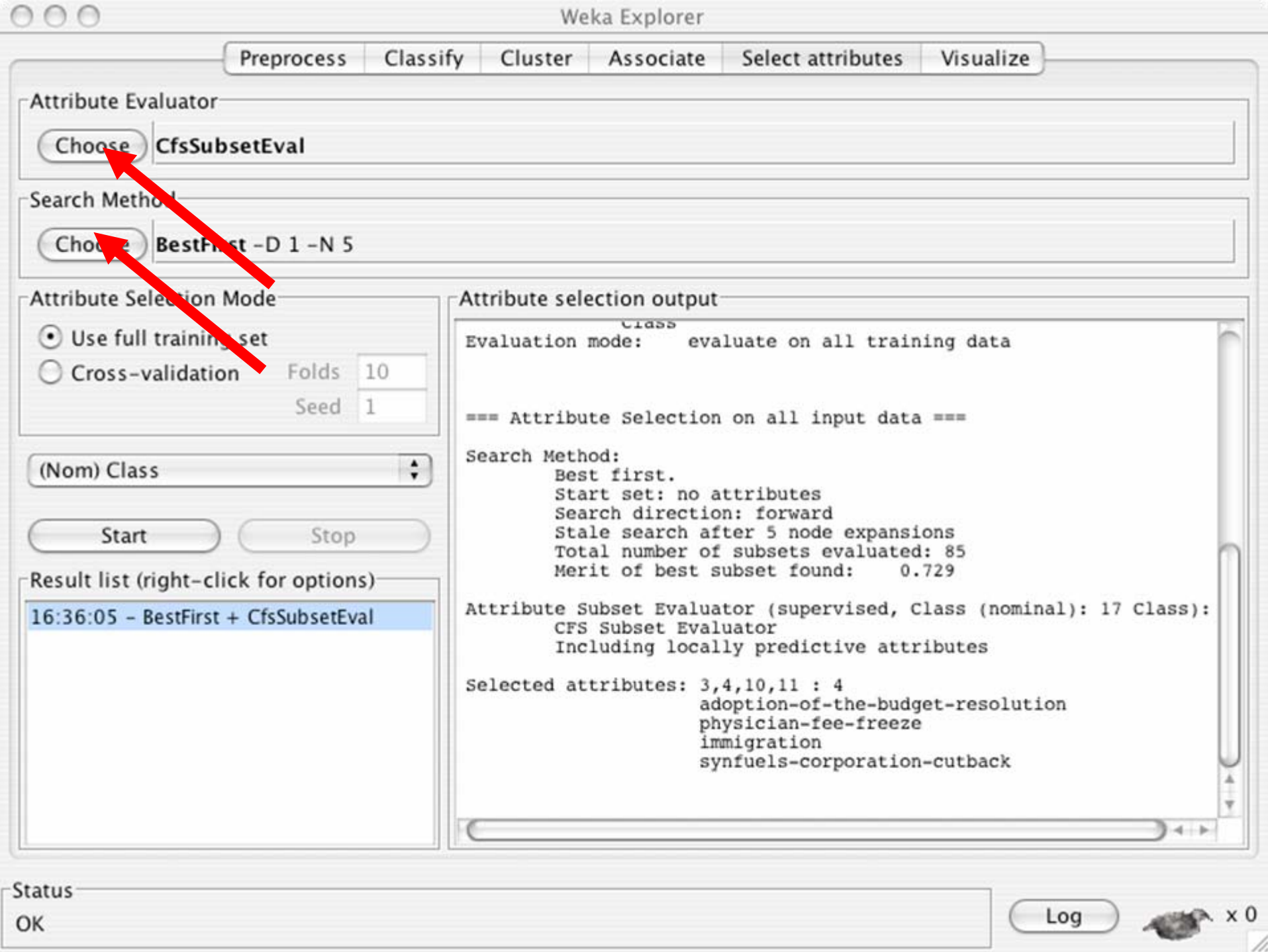

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0 -c -1
Start
Stop
Result list (right-click fo
16:23:55 - Apriori
Associator output
Apriori
=======
Minimum support: 0.45 (196 instances)
Minimum metric <confidence>: 0.9
Number of cycles performed: 11
Generated sets of large itemsets:
Size of set of large itemsets L(1): 20
Size of set of large itemsets L(2): 17
Size of set of large itemsets L(3): 6
Size of set of large itemsets L(4): 1
Best rules found:
1. adoption-of-the-budget-resolution=y physician-fee-freeze=n 219 ==> Class=democr
2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-co
3. physician-fee-freeze=n aid-to-nicaraguan-contras=y 211 ==> Class=democrat 210
4. physician-fee-freeze=n education-spending=n 202 ==> Class=democrat 201 conf:
5. physician-fee-freeze=n 247 ==> Class=democrat 245 conf:(0.99)
6. el-salvador-aid=n Class=democrat 200 ==> aid-to-nicaraguan-contras=y 197 cor
7. el-salvador-aid=n 208 ==> aid-to-nicaraguan-contras=y 204 conf:(0.98)
8. adoption-of-the-budget-resolution=y aid-to-nicaraguan-contras=y Class=democrat
9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197 cor
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210
Status
OK
Log
x 0

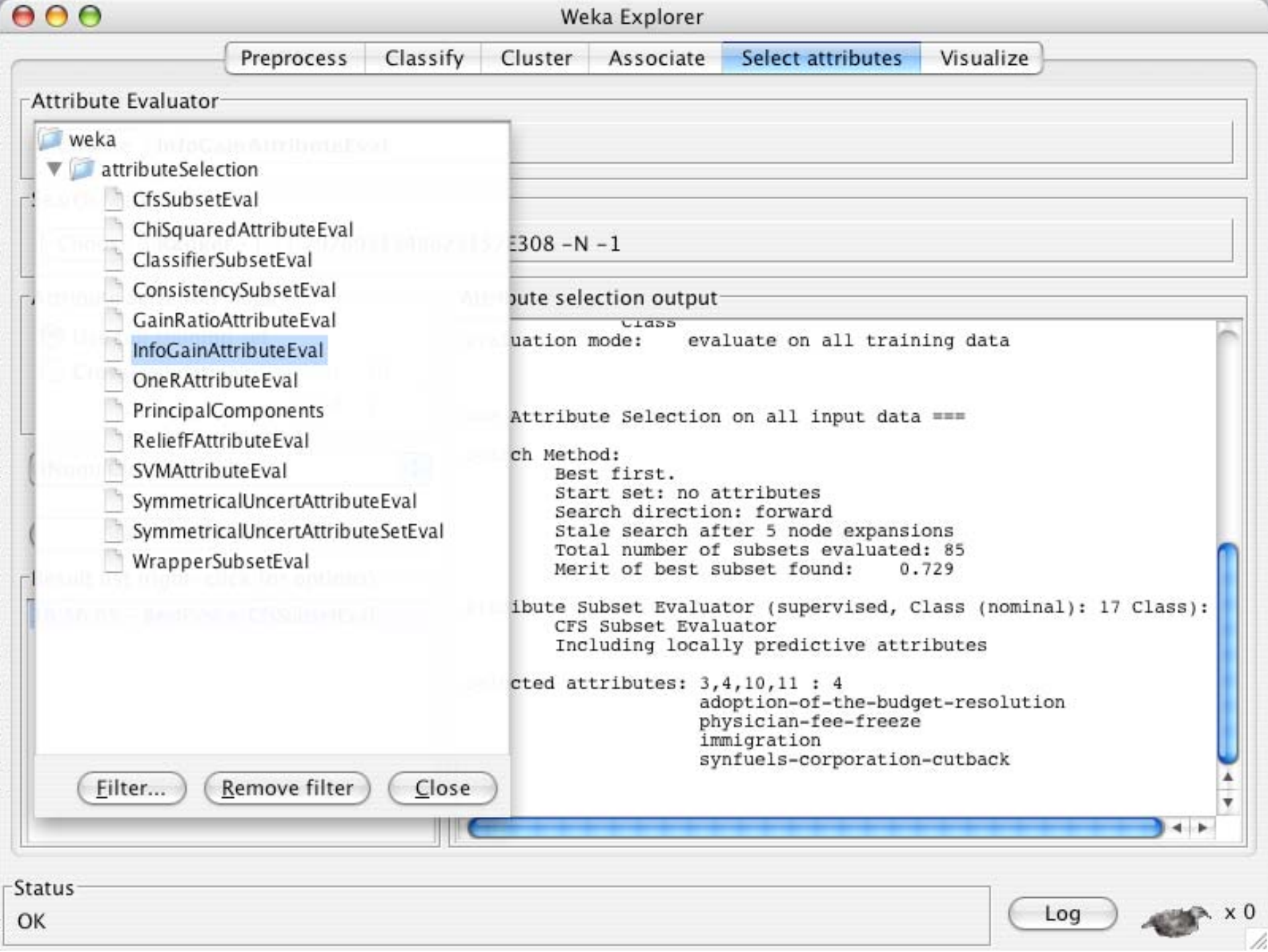
Explorer: attribute selection

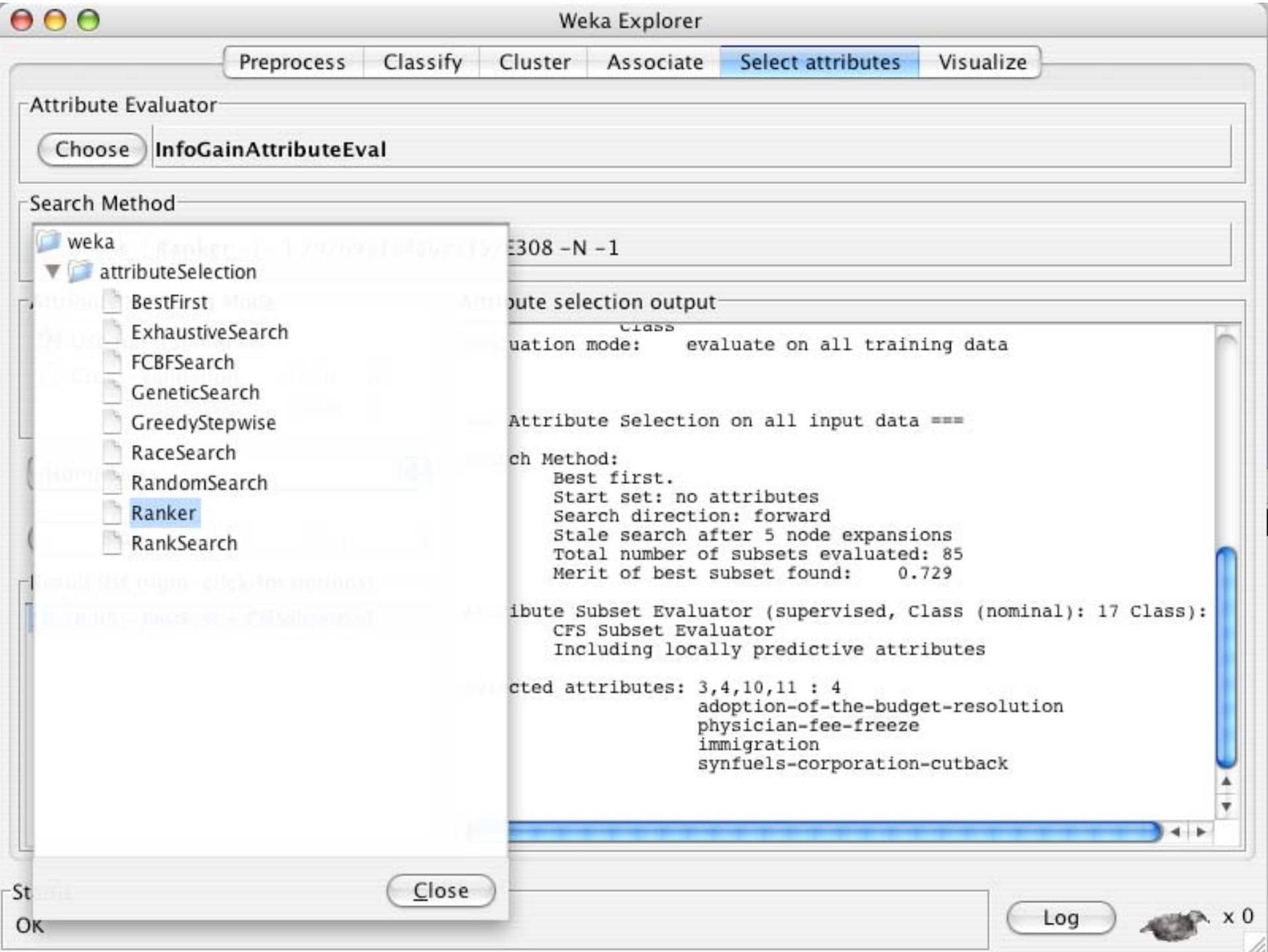
- Panel that can be used to investigate which (subsets of) attributes are the most predictive ones
- Attribute selection methods contain two parts:
  - A search method: best-first, forward selection, random, exhaustive, genetic algorithm, ranking
  - An evaluation method: correlation-based, wrapper, information gain, chi-squared, …
- Very flexible: WEKA allows (almost) arbitrary combinations of these two

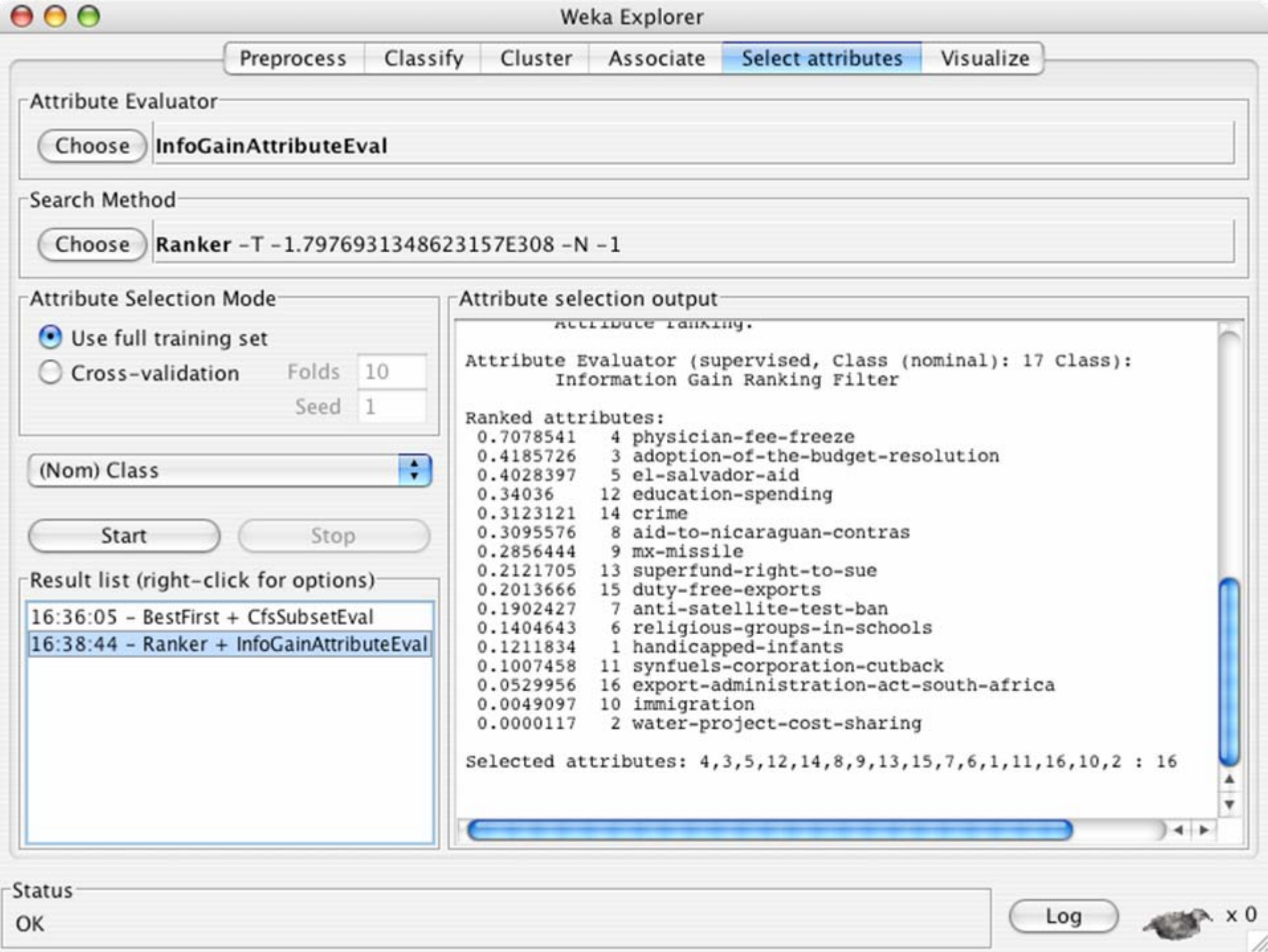

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
CfsSubsetEval
Search Method
Choose
BestFirst -D 1 -N 5
Attribute Selection Mode
Use full training set
Cross-validation
Folds
10
Seed
1
Attribute selection output
(Nom) Class
Start
Stop
Result list (right-click for options)
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
CfsSubsetEval
Search Method
Choose
BestFirst -D 1 -N 5
Attribute Selection Mode
Use full training set
Cross-validation
Folds
10
Seed
1
(Nom) Class
Start
Stop
Result list (right-click for options)
16:36:05 - BestFirst + CfsSubsetEval
Attribute selection output
Evaluation mode: evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 85
Merit of best subset found: 0.729
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Including locally predictive attributes
Selected attributes: 3,4,10,11 : 4
adoption-of-the-budget-resolution
physician-fee-freeze
immigration
synfuels-corporation-cutback
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
CfsSubsetEval
Search Method
Choose
BestFirst -D 1 -N 5
Attribute Selection Mode
Use full training set
Cross-validation
Folds
10
Seed
1
(Nom) Class
Start
Stop
Result list (right-click for options)
16:36:05 - BestFirst + CfsSubsetEval
Attribute selection output
Evaluation mode: evaluate on all training data
=== Attribute Selection on all input data ===
Search Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 85
Merit of best subset found: 0.729
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Including locally predictive attributes
Selected attributes: 3,4,10,11 : 4
adoption-of-the-budget-resolution
physician-fee-freeze
immigration
synfuels-corporation-cutback
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
weka
attributeSelection
CfsSubsetEval
ChiSquaredAttributeEval
ClassifierSubsetEval
ConsistencySubsetEval
GainRatioAttributeEval
InfoGainAttributeEval
OneRAttributeEval
PrincipalComponents
ReliefFAttributeEval
SVMAttributeEval
SymmetricalUncertAttributeEval
SymmetricalUncertAttributeSetEval
WrapperSubsetEval
Filter...
Remove filter
Close
E308 -N -1
bute selection output
uation mode: evaluate on all training data
Attribute Selection on all input data ===
ch Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 85
Merit of best subset found: 0.729
ibute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Including locally predictive attributes
cted attributes: 3,4,10,11 : 4
adoption-of-the-budget-resolution
physician-fee-freeze
immigration
synfuels-corporation-cutback
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
InfoGainAttributeEval
Search Method
weka
attributeSelection
BestFirst
ExhaustiveSearch
FCBFSearch
GeneticSearch
GreedyStepwise
RaceSearch
RandomSearch
Ranker
RankSearch
Close
308 -N -1
ute selection output
uation mode: evaluate on all training data
Attribute Selection on all input data ===
ch Method:
Best first.
Start set: no attributes
Search direction: forward
Stale search after 5 node expansions
Total number of subsets evaluated: 85
Merit of best subset found: 0.729
ibute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Including locally predictive attributes
cted attributes: 3,4,10,11 : 4
adoption-of-the-budget-resolution
physician-fee-freeze
immigration
synfuels-corporation-cutback
Ok
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
InfoGainAttributeEval
Search Method
Choose
Ranker -T -1.7976931348623157E308 -N -1
Attribute Selection Mode
Use full training set
Cross-validation
Folds
10
Seed
1
(Nom) Class
Start
Stop
Result list (right-click for options)
16:36:05 - BestFirst + CfsSubsetEval
16:38:44 - Ranker + InfoGainAttributeEval
Attribute selection output
Attribute ranking.
Attribute Evaluator (supervised, Class (nominal): 17 Class):
Information Gain Ranking Filter
Ranked attributes:
0.7078541 4 physician-fee-freeze
0.4185726 3 adoption-of-the-budget-resolution
0.4028397 5 el-salvador-aid
0.34036 12 education-spending
0.3123121 14 crime
0.3095576 8 aid-to-nicaraguan-contras
0.2856444 9 mx-missile
0.2121705 13 superfund-right-to-sue
0.2013666 15 duty-free-exports
0.1902427 7 anti-satellite-test-ban
0.1404643 6 religious-groups-in-schools
0.1211834 1 handicapped-infants
0.1007458 11 synfuels-corporation-cutback
0.0529956 16 export-administration-act-south-africa
0.0049097 10 immigration
0.0000117 2 water-project-cost-sharing
Selected attributes: 4,3,5,12,14,8,9,13,15,7,6,1,11,16,10,2 : 16
Status
OK
Log
x 0

Explorer: data visualization

- Visualization very useful in practice: e.g. helps to determine difficulty of the learning problem
- WEKA can visualize single attributes (1-d) and pairs of attributes (2-d)
  - To do: rotating 3-d visualizations (Xgobi-style)
- Color-coded class values
- “Jitter” option to deal with nominal attributes (and to detect “hidden” data points)
- “Zoom-in” function

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: Glass
Instances: 214 Attributes: 10

Attributes

All | None | Invert | Pattern

No.		Name
1	☐	RI
2	☐	Na
3	☐	Mg
4	☐	Al
5	☐	Si
6	☐	K
7	☐	Ca
8	☐	Ba
9	☐	Fe
10	☐	Type

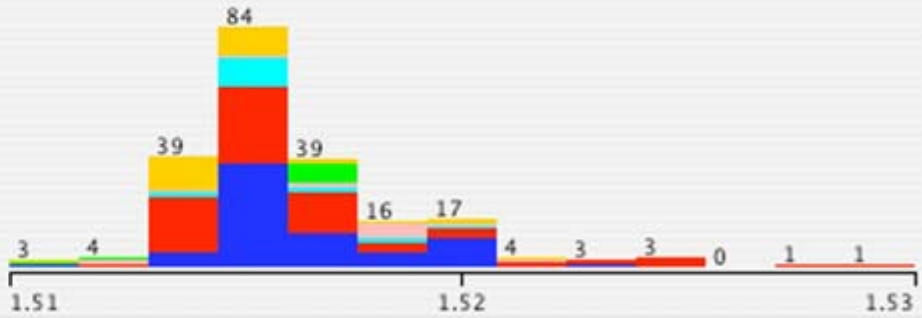
Remove

Selected attribute

Name: RI Type: Numeric
Missing: 0 (0%) Distinct: 178 Unique: 145 (68%)

Statistic	Value
Minimum	1.511
Maximum	1.534
Mean	1.518
StdDev	0.003

Class: Type (Nom) | Visualize All



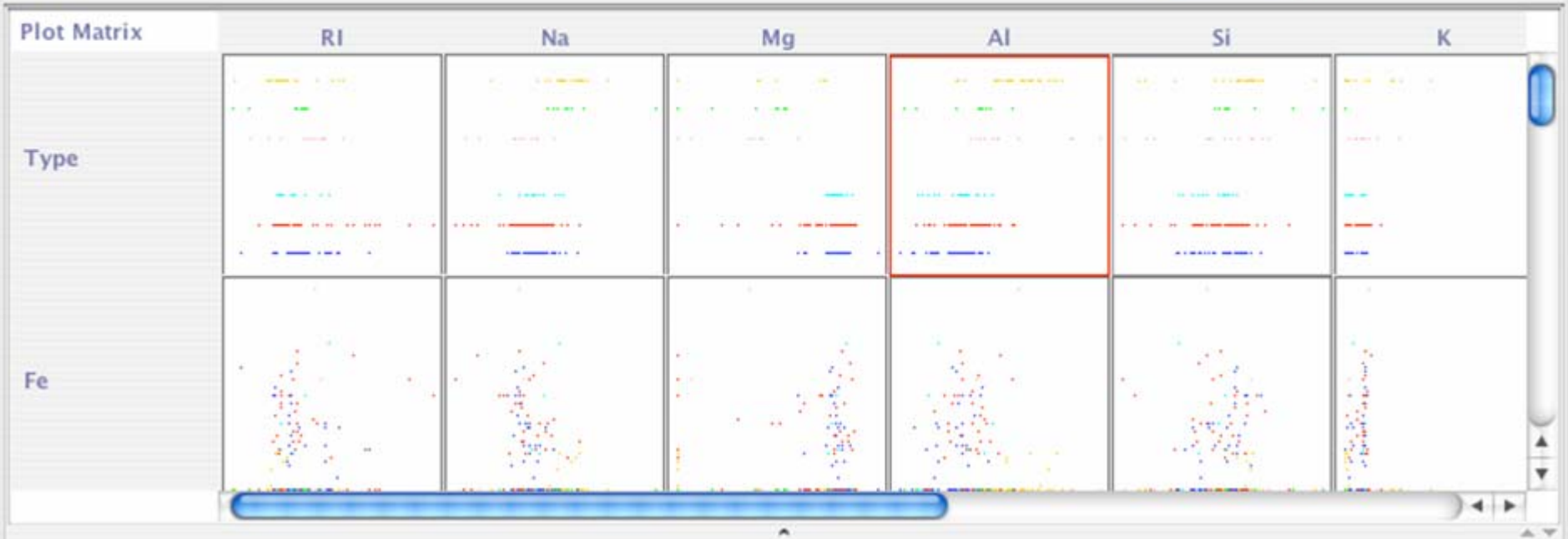
Status

OK

Log | x 0

Weka Explorer

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



PlotSize: [100]

PointSize: [1]

Jitter:

Update

Select Attributes

Colour: Type (Nom)

SubSample % : 100

Class Colour

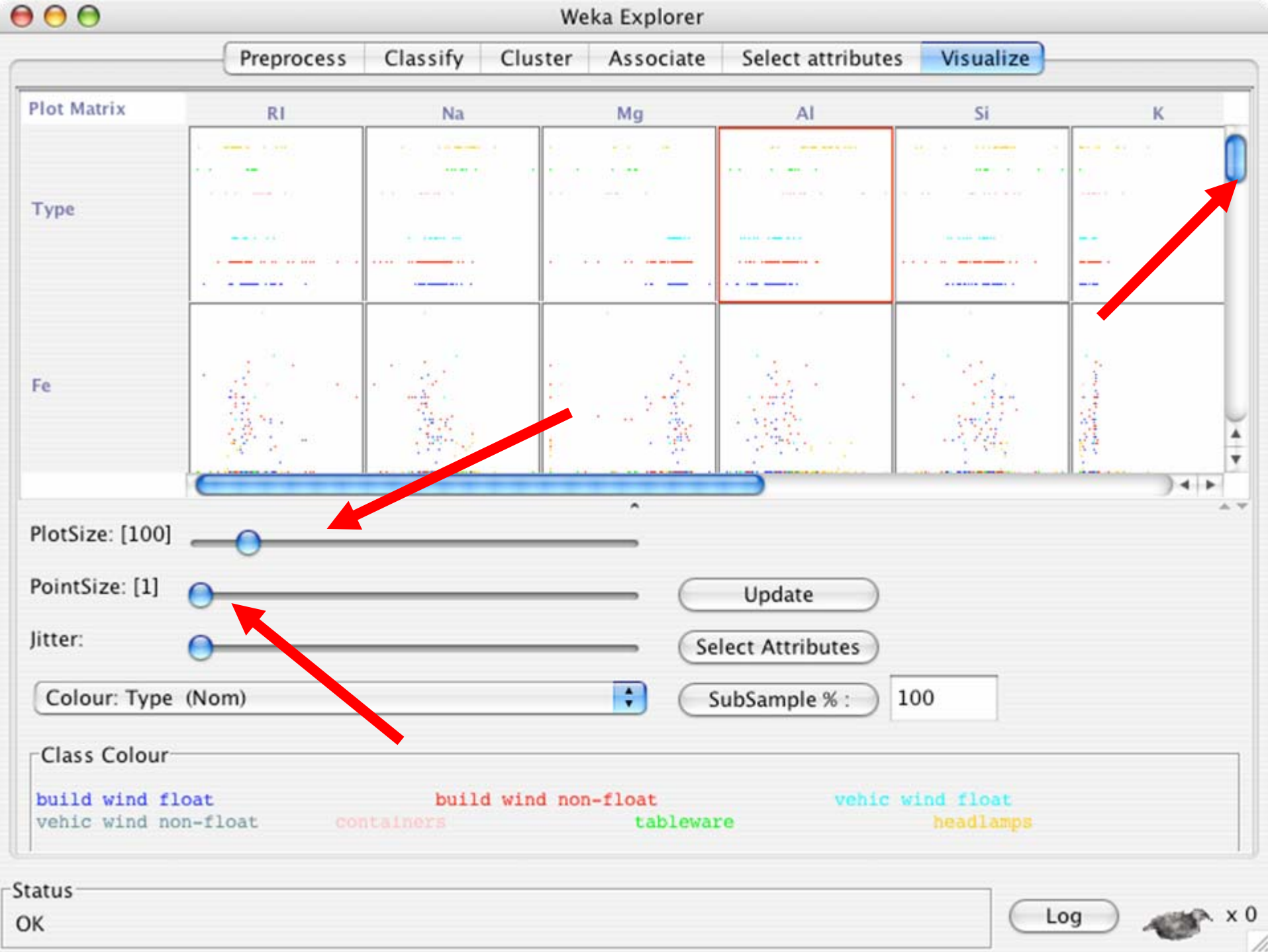
build wind float build wind non-float vehic wind float
vehic wind non-float containers tableware headlamps

Status

OK

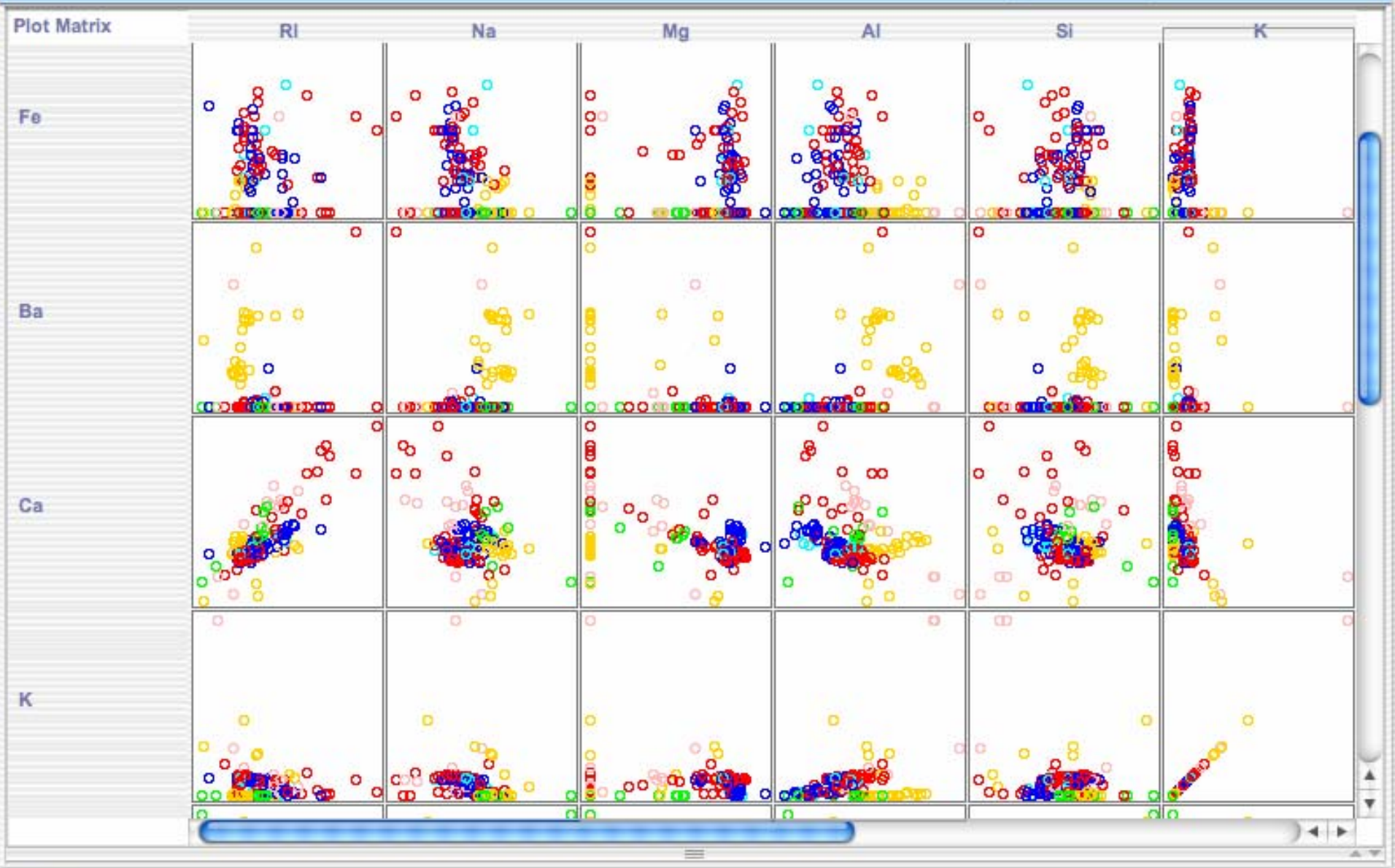
Log

x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
RI
Na
Mg
Al
Si
K
Type
Fe
PlotSize: [100]
PointSize: [1]
Jitter:
Update
Select Attributes
Colour: Type (Nom)
SubSample % :
100
Class Colour
build wind float
build wind non-float
vehic wind float
vehic wind non-float
containers
tableware
headlamps
Status
OK
Log
x 0

Weka Knowledge Explorer

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



Status

OK

Log

x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
RI
Na
Mg
Al
Si
K
Ca
Fe
Ba
Ca
K
Si
X: Al Y: Ca (click to enlarge)
Status
OK
Log
x 0

Weka Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

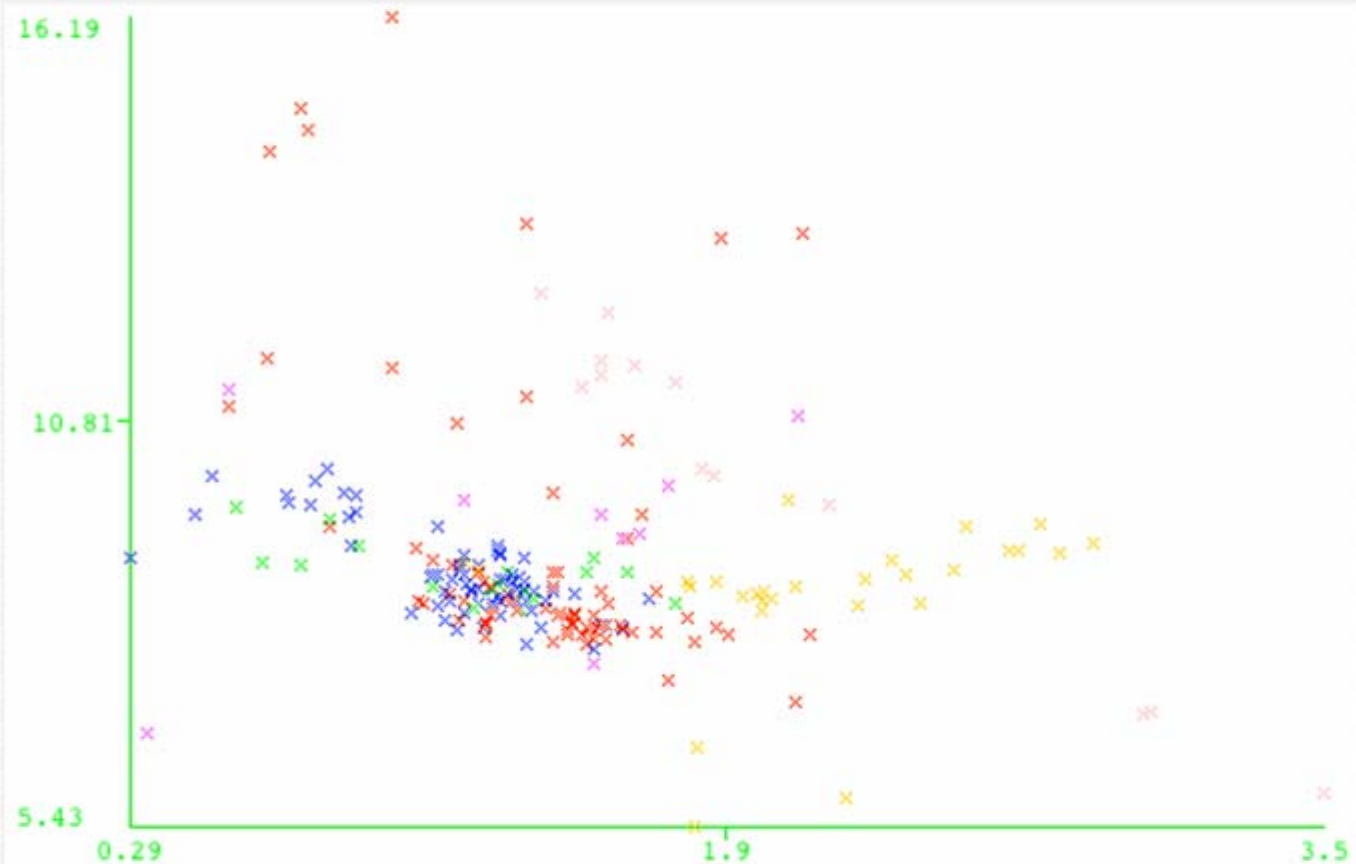
Colour: Type (Nom)

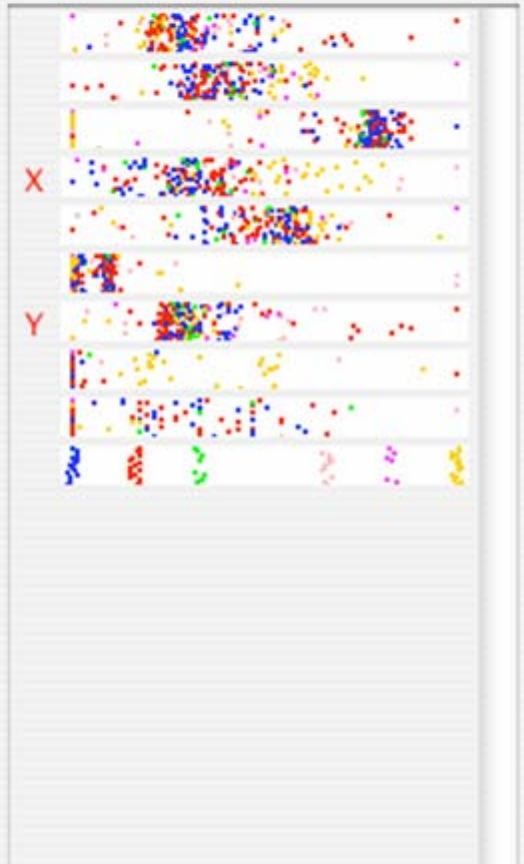
Select Instance

Reset | Clear | Open | Save

Jitter

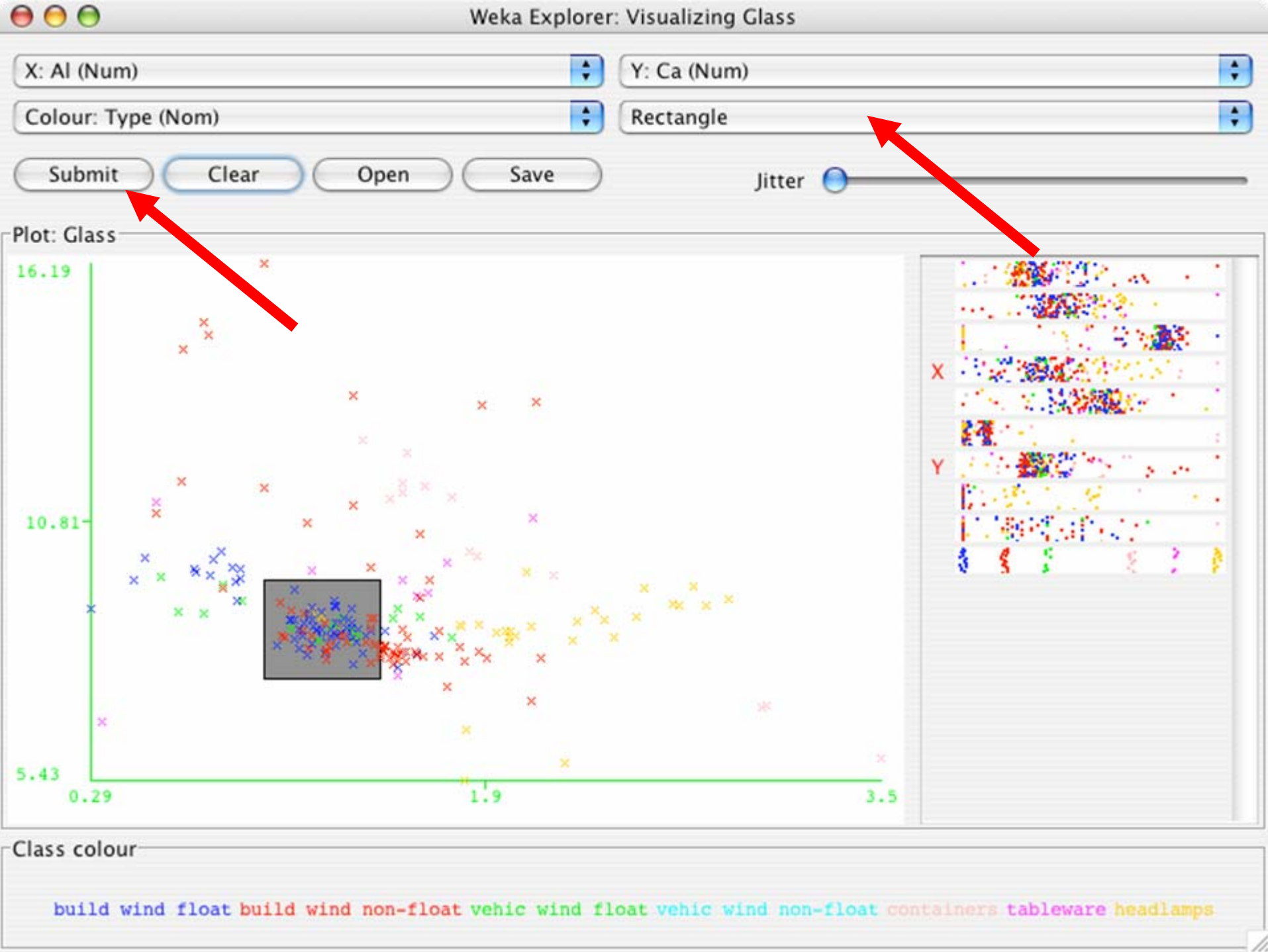
Plot: Glass





Class colour

build wind float
build wind non-float
vehic wind float
vehic wind non-float
containers
tableware
headlamps


Weka Explorer: Visualizing Glass
X: Al (Num)
Y: Ca (Num)
Colour: Type (Nom)
Rectangle
Submit
Clear
Open
Save
Jitter
Plot: Glass
16.19
10.81
5.43
0.29
1.9
3.5
X
Y
Class colour
build wind float
build wind non-float
vehic wind float
vehic wind non-float
containers
tableware
headlamps

Weka Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

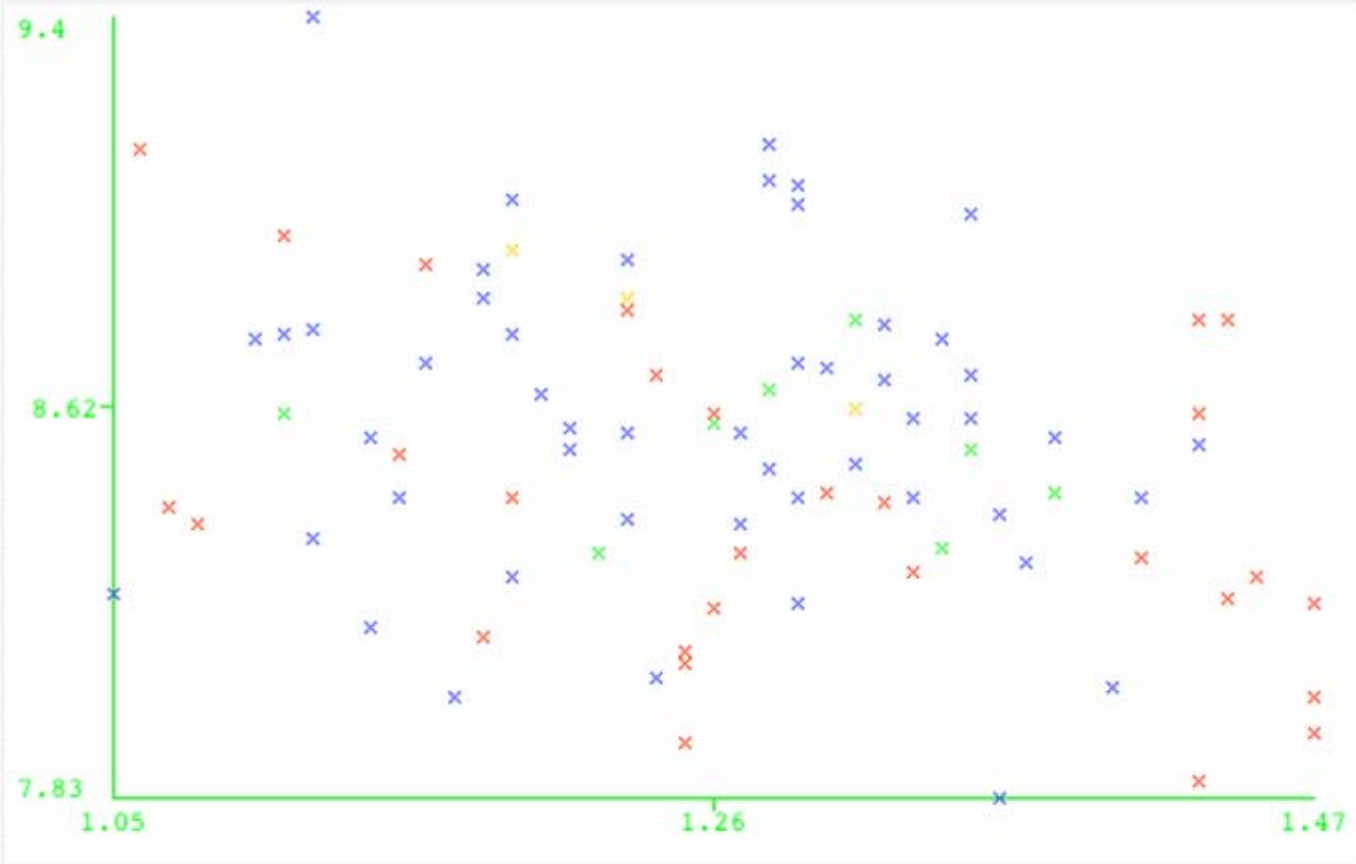
Colour: Type (Nom)

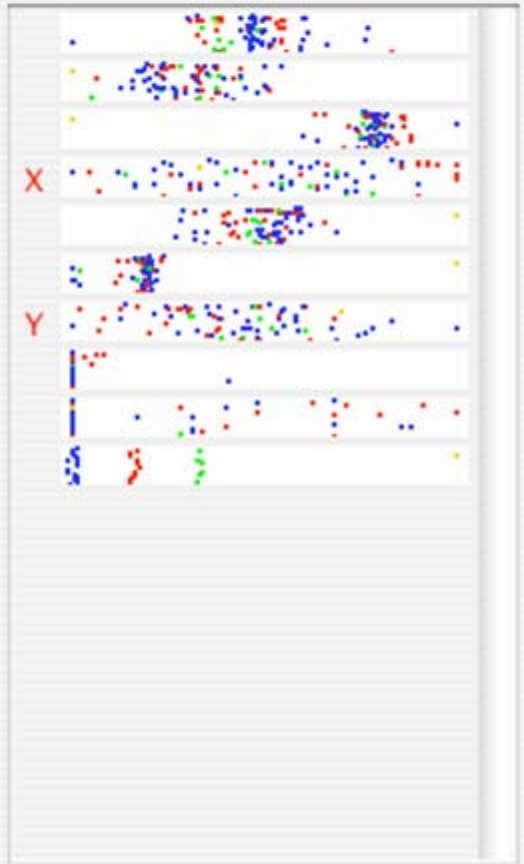
Rectangle

Reset | Clear | Open | Save

Jitter

Plot: Glass

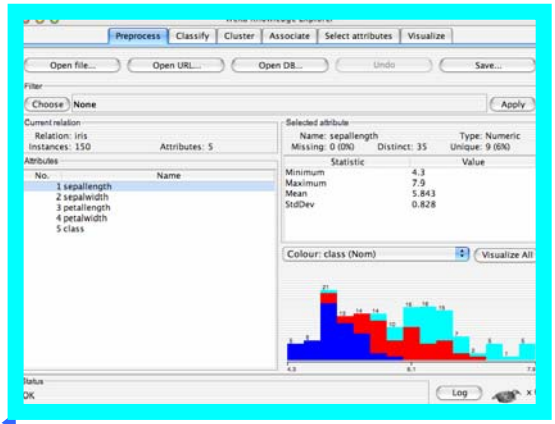


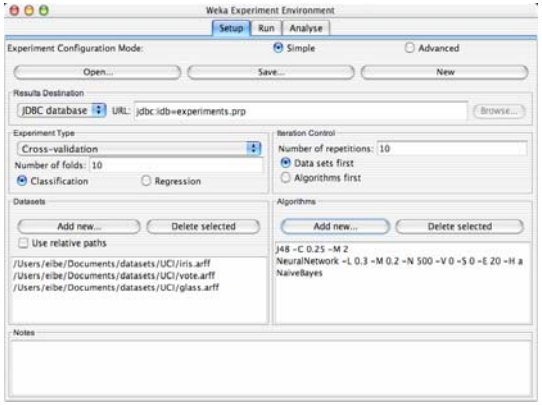


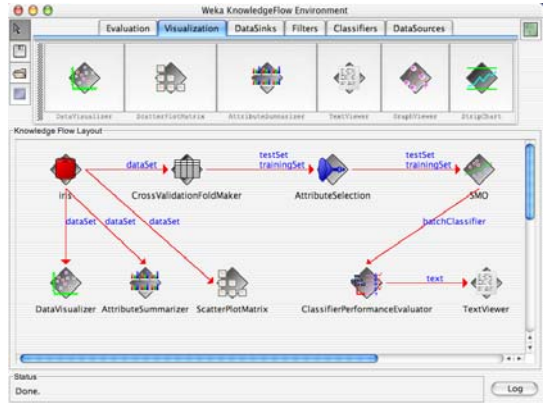
Class colour

build wind float
build wind non-float
vehic wind float
vehic wind non-float
containers
tableware
headlamps

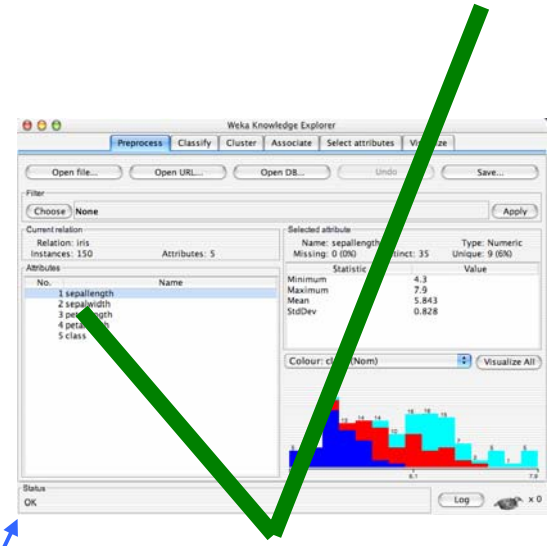

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log

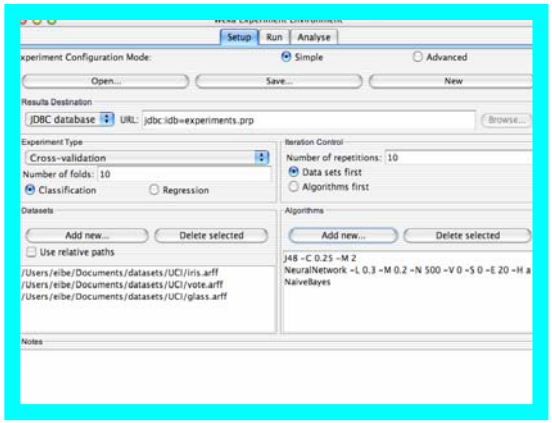


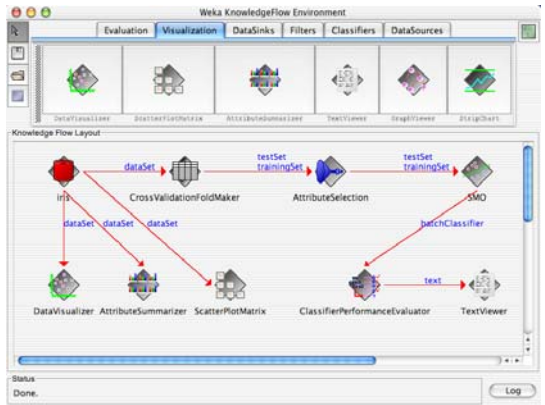





Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log

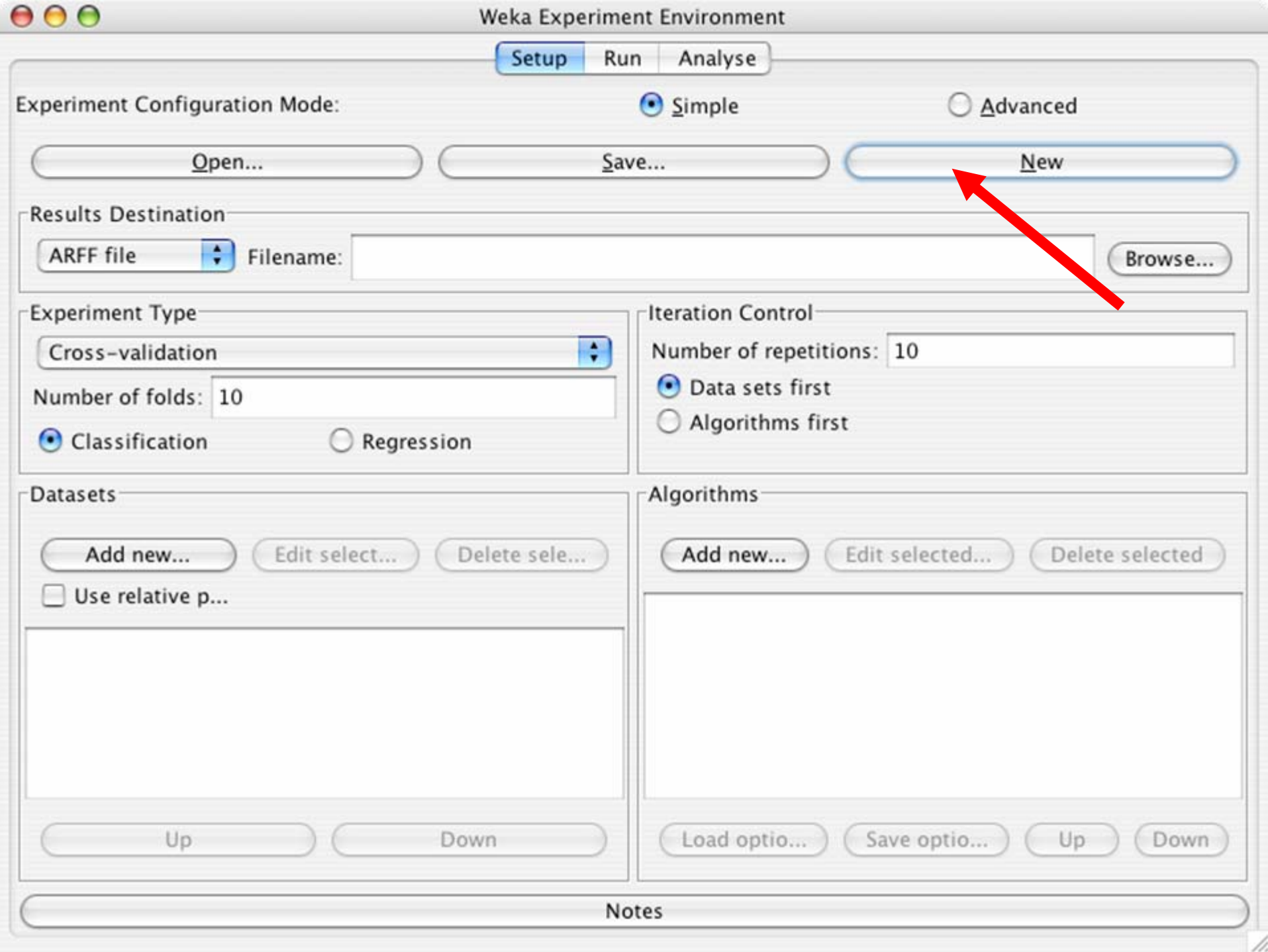

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Open file...
Open URL...
Open DB...
Save...
Choose
None
Apply
Visualize All
Status
OK
Log

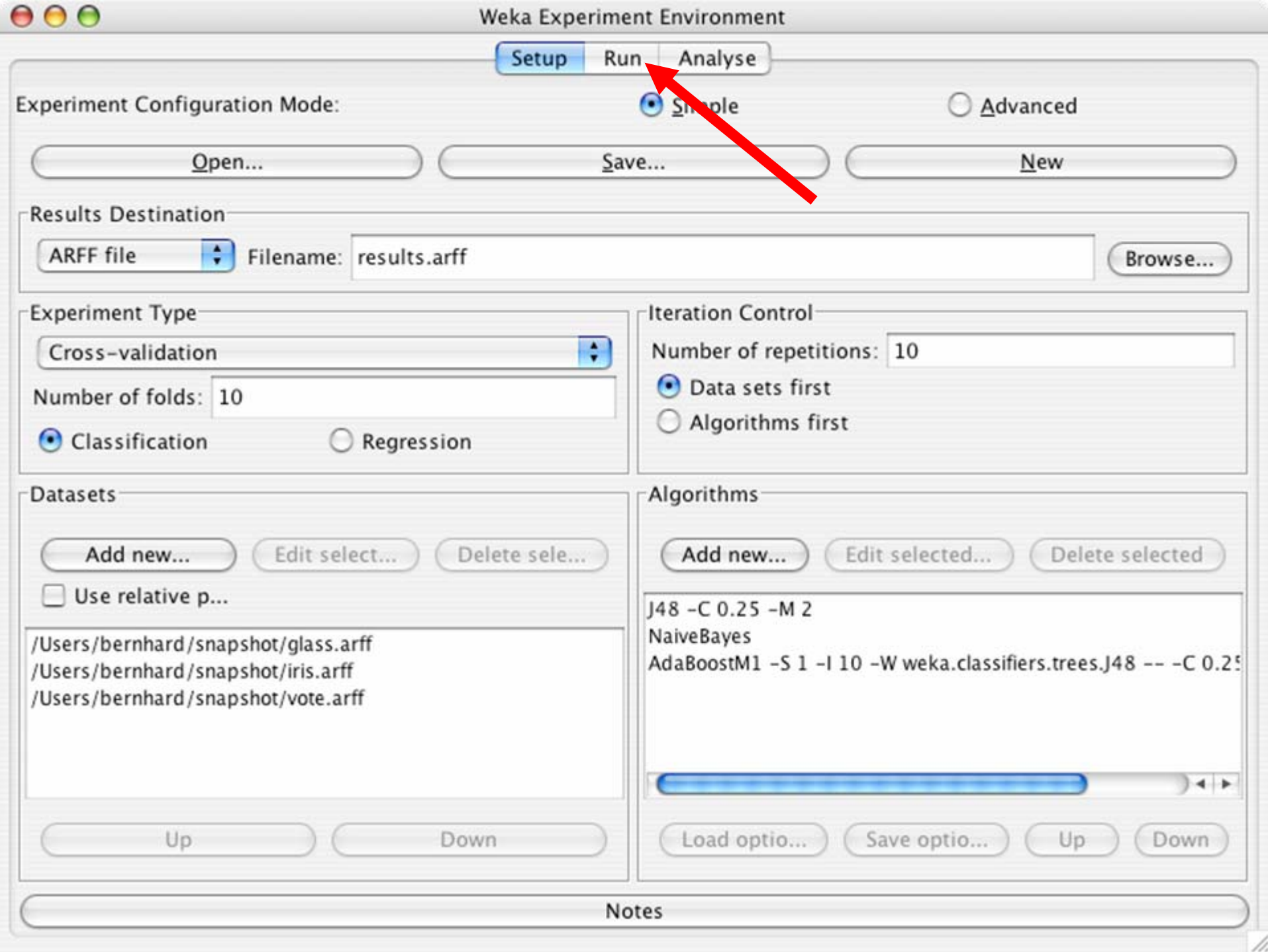

Setup
Run
Analyse
Simple
Advanced
Open...
Save...
New
JDBC database
jdbc:idb=experiments.prp
Cross-validation
Number of folds: 10
Classification
Regression
Data sets first
Algorithms first
Add new...
Delete selected
Use relative paths
/Users/eibe/Documents/datasets/UCI/iris.arff
/Users/eibe/Documents/datasets/UCI/vote.arff
/Users/eibe/Documents/datasets/UCI/glass.arff
NaiveBayes

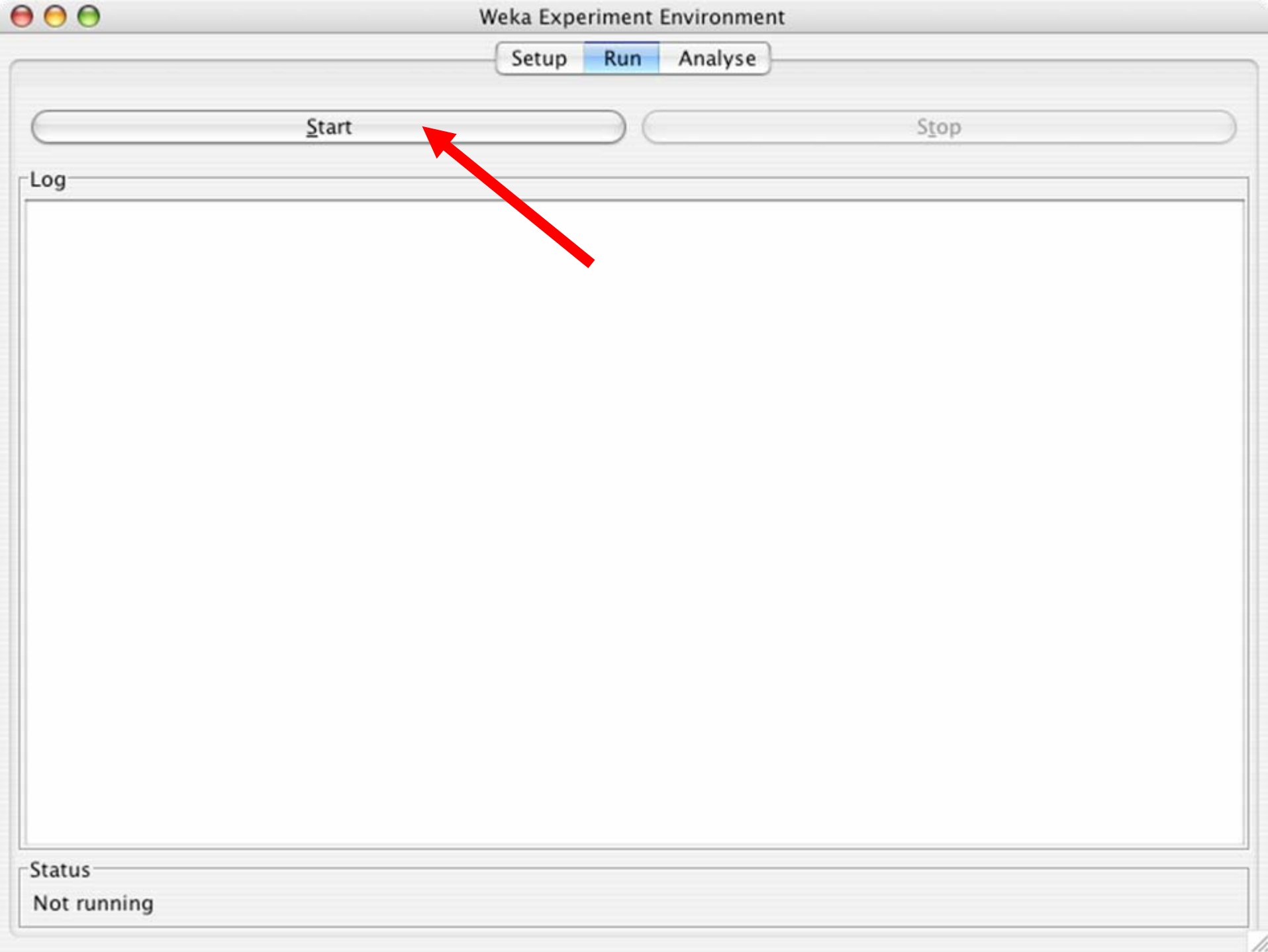

Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources
Knowledge Flow Layout
dataSet
testSet
trainingSet
CrossValidationFoldMaker
AttributeSelection
SMO
batchClassifier
DataVisualizer
AttributeSummarizer
ScatterPlotMatrix
ClassifierPerformanceEvaluator
TextViewer
text
Status
Done.
Log

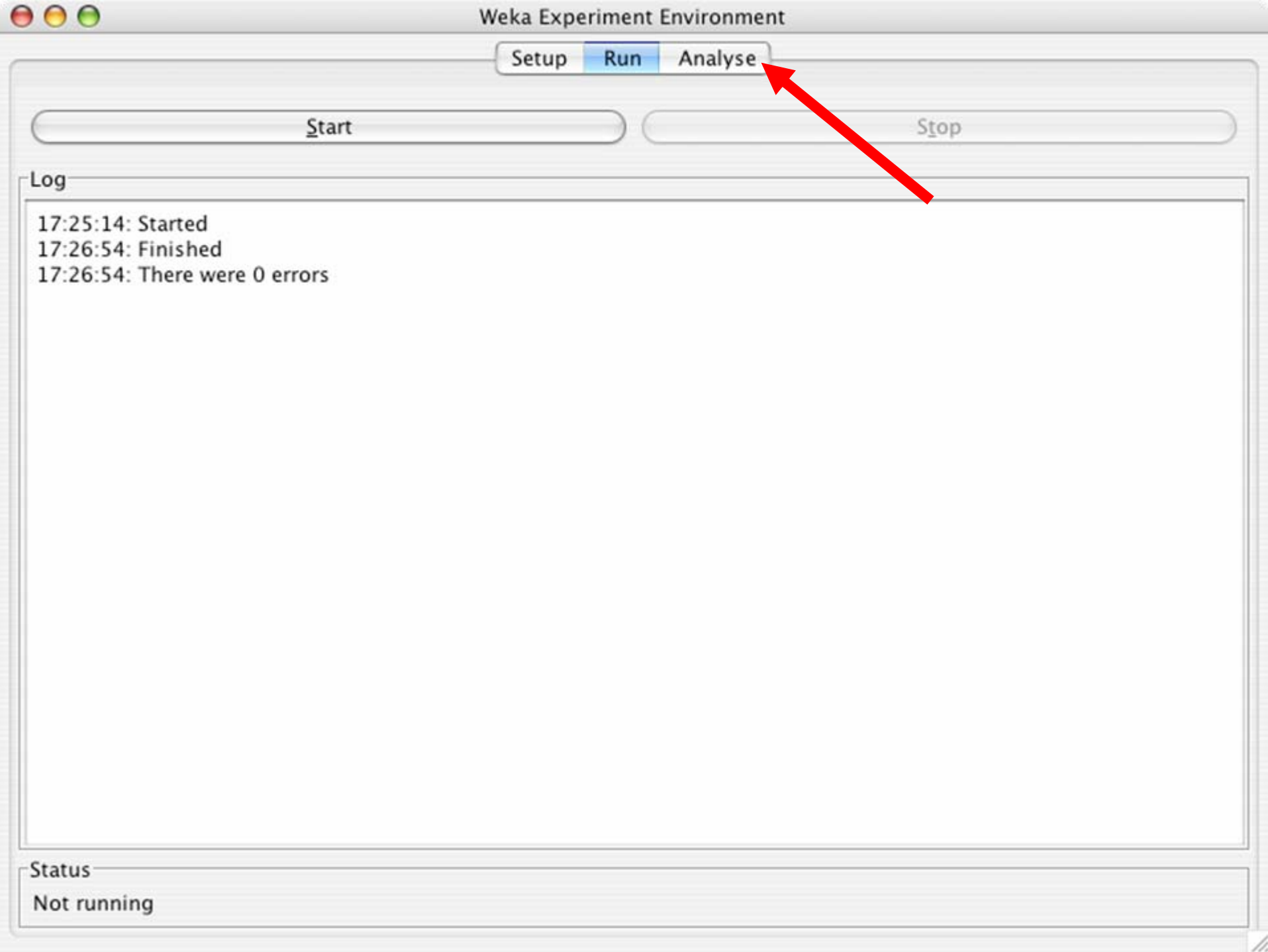
Performing experiments

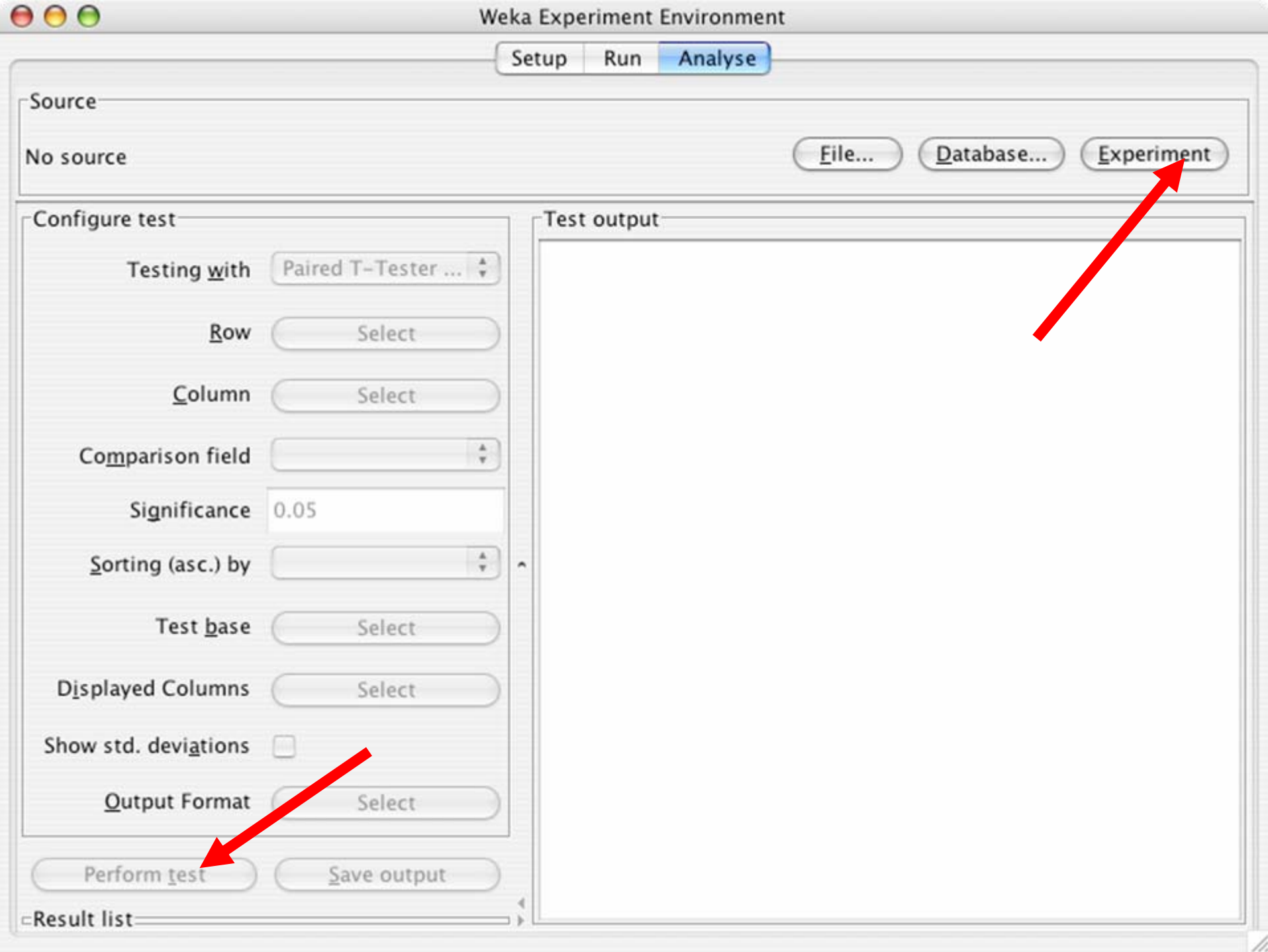
- Experimenter makes it easy to compare the performance of different learning schemes
- For classification and regression problems
- Results can be written into file or database
- Evaluation options: cross-validation, learning curve, hold-out
- Can also iterate over different parameter settings
- Significance-testing built in!

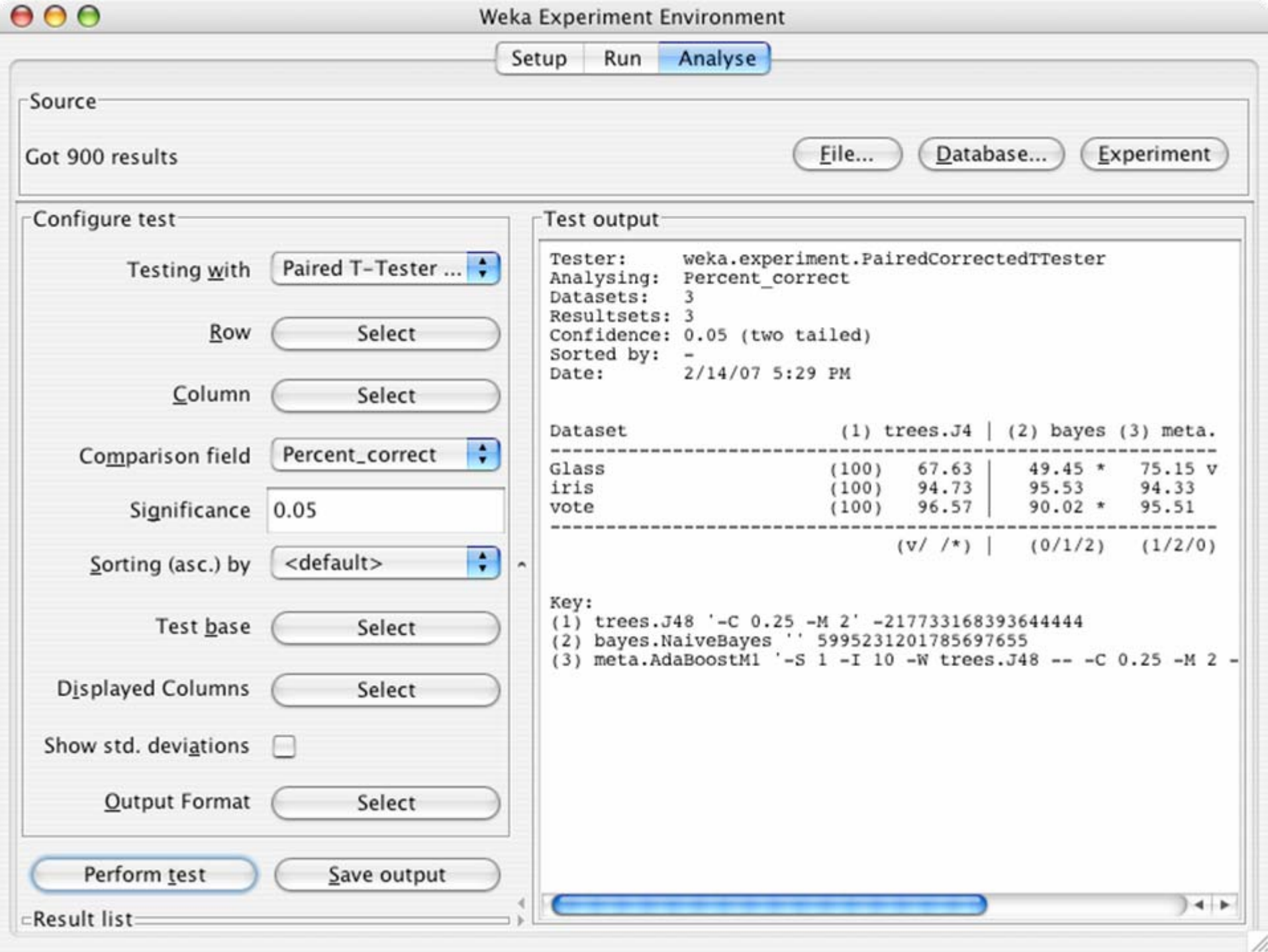

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
ARFF file
Filename:
Browse...
Experiment Type
Cross-validation
Number of folds: 10
Classification
Regression
Iteration Control
Number of repetitions: 10
Data sets first
Algorithms first
Datasets
Add new...
Edit select...
Delete sele...
Use relative p...
Up
Down
Algorithms
Add new...
Edit selected...
Delete selected
Load optio...
Save optio...
Up
Down
Notes

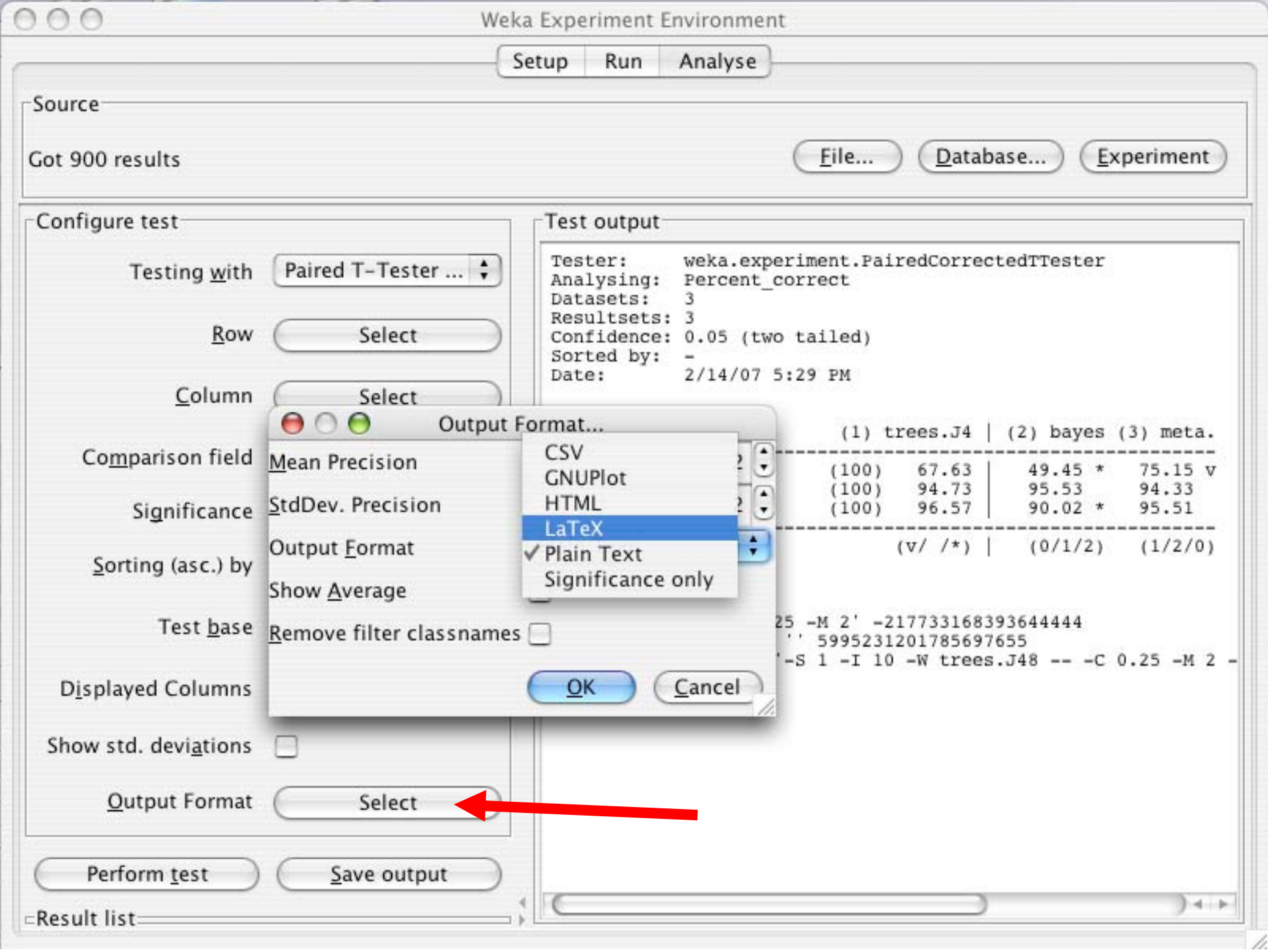

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
ARFF file
Filename:
results.arff
Browse...
Experiment Type
Cross-validation
Number of folds:
10
Classification
Regression
Iteration Control
Number of repetitions:
10
Data sets first
Algorithms first
Datasets
Add new...
Edit select...
Delete sele...
Use relative p...
/Users/bernhard/snapshot/glass.arff
/Users/bernhard/snapshot/iris.arff
/Users/bernhard/snapshot/vote.arff
Up
Down
Algorithms
Add new...
Edit selected...
Delete selected
J48 -C 0.25 -M 2
NaiveBayes
AdaBoostM1 -S 1 -I 10 -W weka.classifiers.trees.J48 -- -C 0.25
Load optio...
Save optio...
Up
Down
Notes

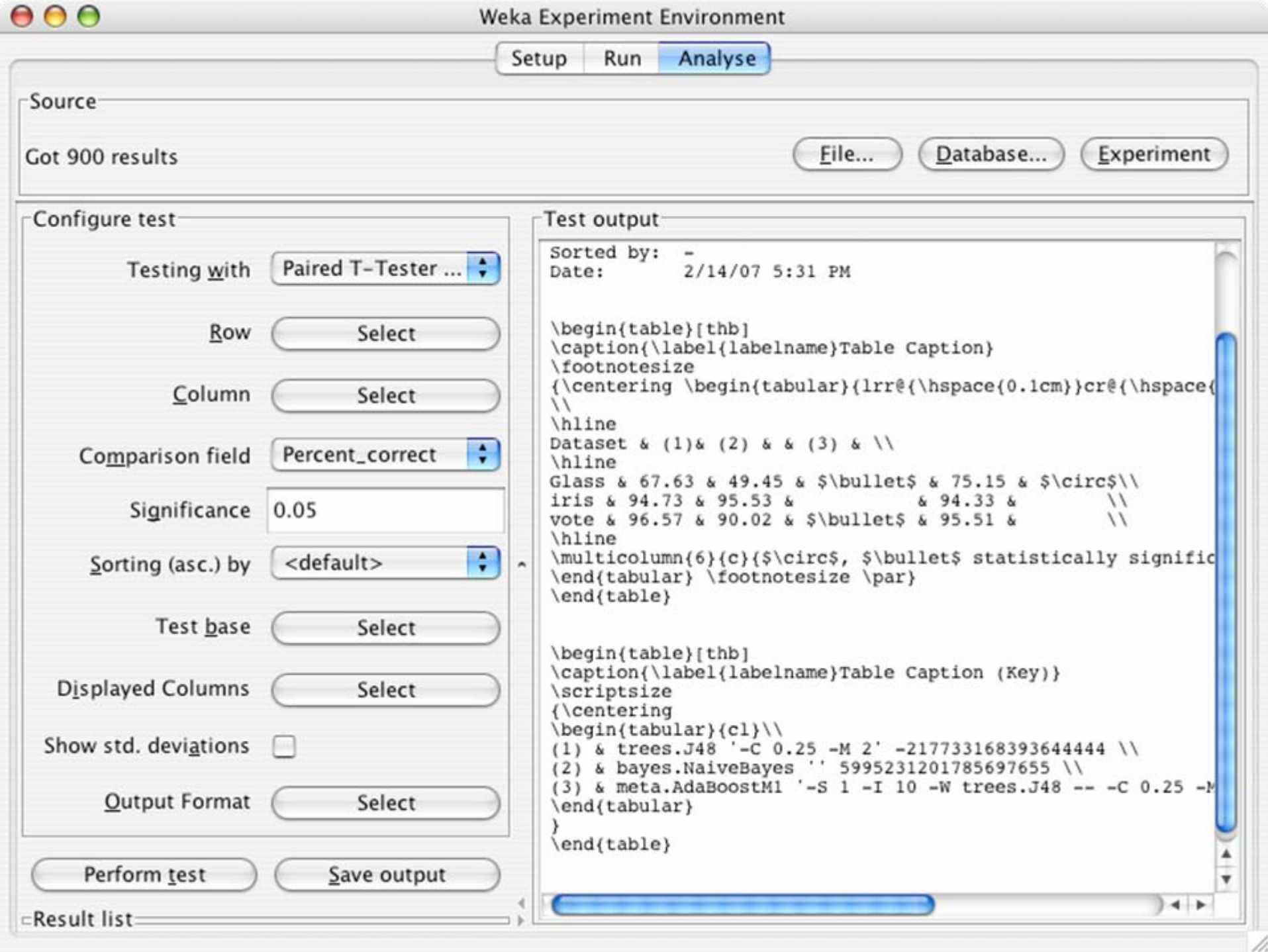

Weka Experiment Environment
Setup
Run
Analyse
Start
Stop
Log
Status
Not running


Weka Experiment Environment
Setup
Run
Analyse
Start
Stop
Log
17:25:14: Started
17:26:54: Finished
17:26:54: There were 0 errors
Status
Not running

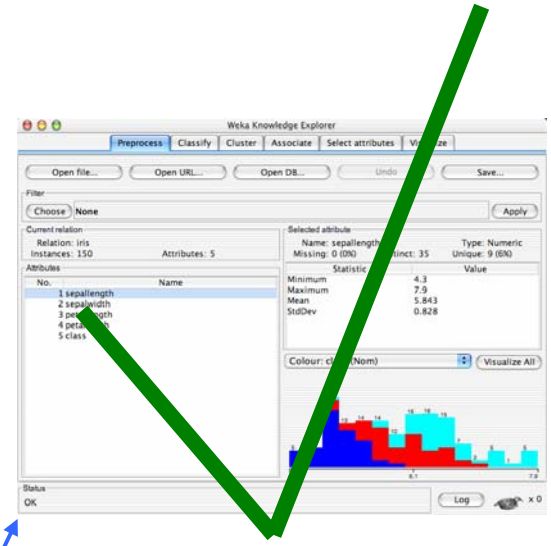

Weka Experiment Environment
Setup
Run
Analyse
Source
No source
File...
Database...
Experiment
Configure test
Testing with
Paired T-Tester ...
Row
Select
Column
Select
Comparison field
Significance
0.05
Sorting (asc.) by
Test base
Select
Displayed Columns
Select
Show std. deviations
Output Format
Select
Perform test
Save output
Result list
Test output

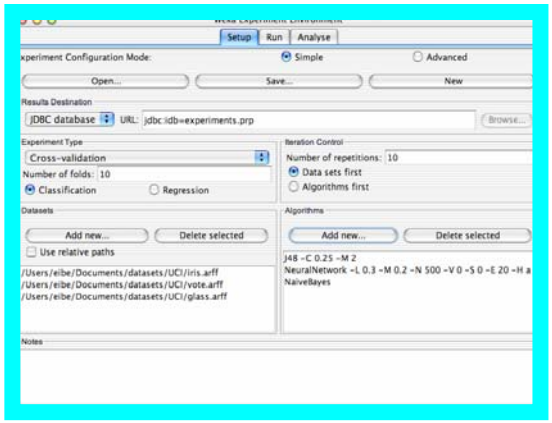

Weka Experiment Environment
Setup
Run
Analyse
Source
Got 900 results
File...
Database...
Experiment
Configure test
Testing with
Paired T-Tester ...
Row
Select
Column
Select
Comparison field
Percent_correct
Significance
0.05
Sorting (asc.) by
<default>
Test base
Select
Displayed Columns
Select
Show std. deviations
Output Format
Select
Perform test
Save output
Result list
Test output
Tester: weka.experiment.PairedCorrectedTTester
Analysing: Percent_correct
Datasets: 3
Resultsets: 3
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 2/14/07 5:29 PM
Dataset (1) trees.J4 | (2) bayes (3) meta.
Glass (100) 67.63 | 49.45 * 75.15 v
iris (100) 94.73 | 95.53 94.33
vote (100) 96.57 | 90.02 * 95.51
(v/ /*) | (0/1/2) (1/2/0)
Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644444
(2) bayes.NaiveBayes '' 5995231201785697655
(3) meta.AdaBoostM1 '-S 1 -I 10 -W trees.J48 -- -C 0.25 -M 2 -

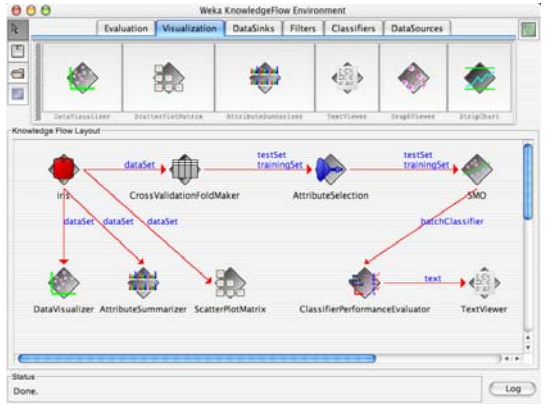

Weka Experiment Environment
Setup
Run
Analyse
Source
Got 900 results
File...
Database...
Experiment
Configure test
Testing with
Paired T-Tester ...
Row
Select
Column
Select
Comparison field
Significance
Sorting (asc.) by
Test base
Displayed Columns
Show std. deviations
Output Format
Select
Perform test
Save output
Result list
Test output
Tester: weka.experiment.PairedCorrectedTTester
Analysing: Percent_correct
Datasets: 3
Resultsets: 3
Confidence: 0.05 (two tailed)
Sorted by: -
Date: 2/14/07 5:29 PM
(1) trees.J4 | (2) bayes (3) meta.
(100) 67.63 | 49.45 * 75.15 v
(100) 94.73 | 95.53 94.33
(100) 96.57 | 90.02 * 95.51
(v/ /*) | (0/1/2) (1/2/0)
25 -M 2' -217733168393644444
'' 5995231201785697655
'-S 1 -I 10 -W trees.J48 -- -C 0.25 -M 2 -
Output Format...
Mean Precision
StdDev. Precision
Output Format
Show Average
Remove filter classnames
CSV
GNUPlot
HTML
LaTeX
Plain Text
Significance only
OK
Cancel


Weka Experiment Environment
Setup
Run
Analyse
Source
Got 900 results
File...
Database...
Experiment
Configure test
Testing with
Paired T-Tester ...
Row
Select
Column
Select
Comparison field
Percent_correct
Significance
0.05
Sorting (asc.) by
<default>
Test base
Select
Displayed Columns
Select
Show std. deviations
Output Format
Select
Perform test
Save output
Result list
Test output
Sorted by: -
Date: 2/14/07 5:31 PM
\begin{table}[thb]
\caption{\label{labelname}Table Caption}
\footnotesize
{\centering \begin{tabular}{lrr@{\hspace{0.1cm}}cr@{\hspace{
\\
\hline
Dataset & (1)& (2) & & (3) & \\
\hline
Glass & 67.63 & 49.45 & $\bullet$ & 75.15 & $\circ$\\
iris & 94.73 & 95.53 & & 94.33 & \\
vote & 96.57 & 90.02 & $\bullet$ & 95.51 & \\
\hline
\multicolumn{6}{c}{$\circ$, $\bullet$ statistically signific
\end{tabular} \footnotesize \par}
\end{table}
\begin{table}[thb]
\caption{\label{labelname}Table Caption (Key)}
\scriptsize
{\centering
\begin{tabular}{cl}\\
(1) & trees.J48 '-C 0.25 -M 2' -217733168393644444 \\
(2) & bayes.NaiveBayes '' 5995231201785697655 \\
(3) & meta.AdaBoostM1 '-S 1 -I 10 -W trees.J48 -- -C 0.25 -M
\end{tabular}
}
\end{table}

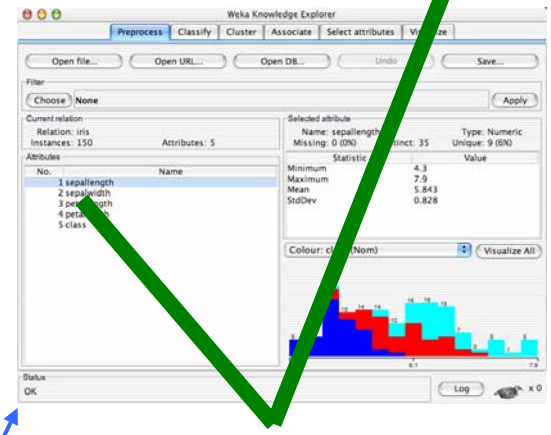

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log

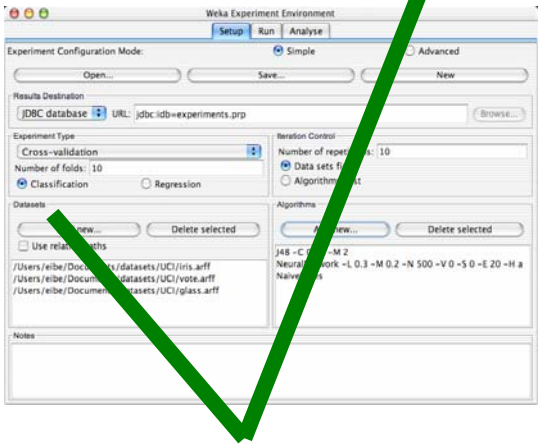

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Open file...
Open URL...
Open DB...
Save...
Choose
None
Apply
Visualize All
Status
OK
Log

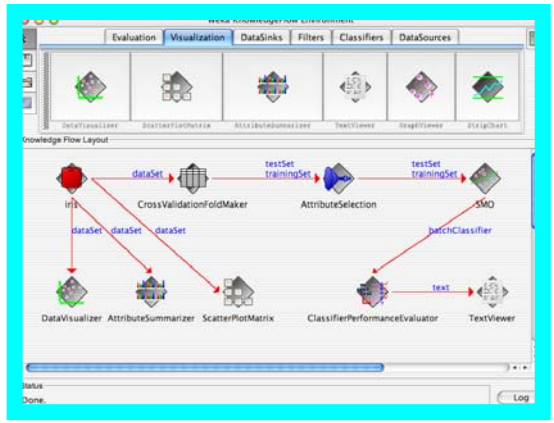

Setup
Run
Analyse
Simple
Advanced
Open...
Save...
New
JDBC database
jdbc:idb=experiments.prp
Cross-validation
Number of folds: 10
Classification
Regression
Data sets first
Algorithms first
Add new...
Delete selected
Use relative paths


Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources
Knowledge Flow Layout
CrossValidationFoldMaker
AttributeSelection
DataVisualizer
AttributeSummarizer
ScatterPlotMatrix
ClassifierPerformanceEvaluator
TextViewer
Status
Done.
Log


Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log



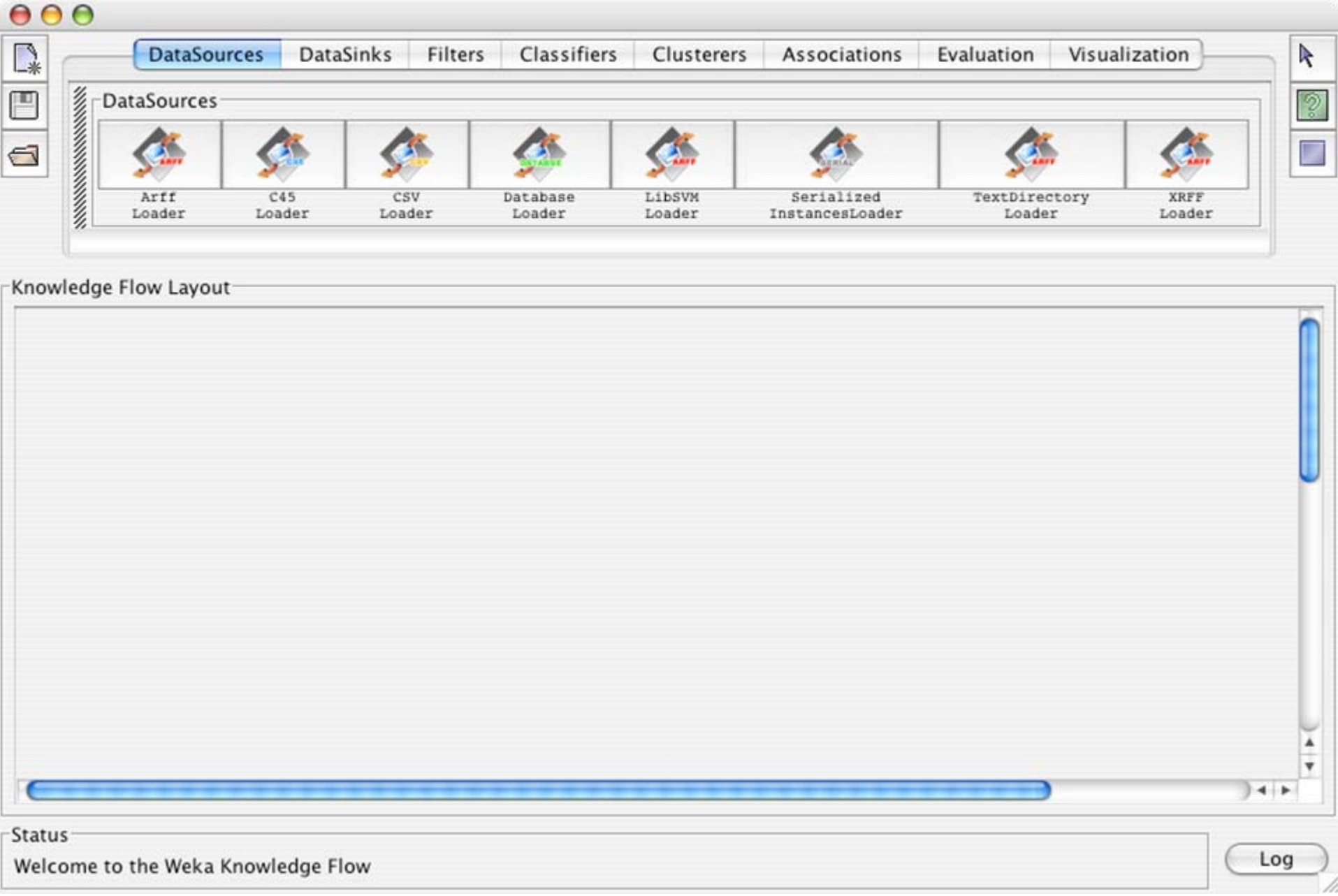


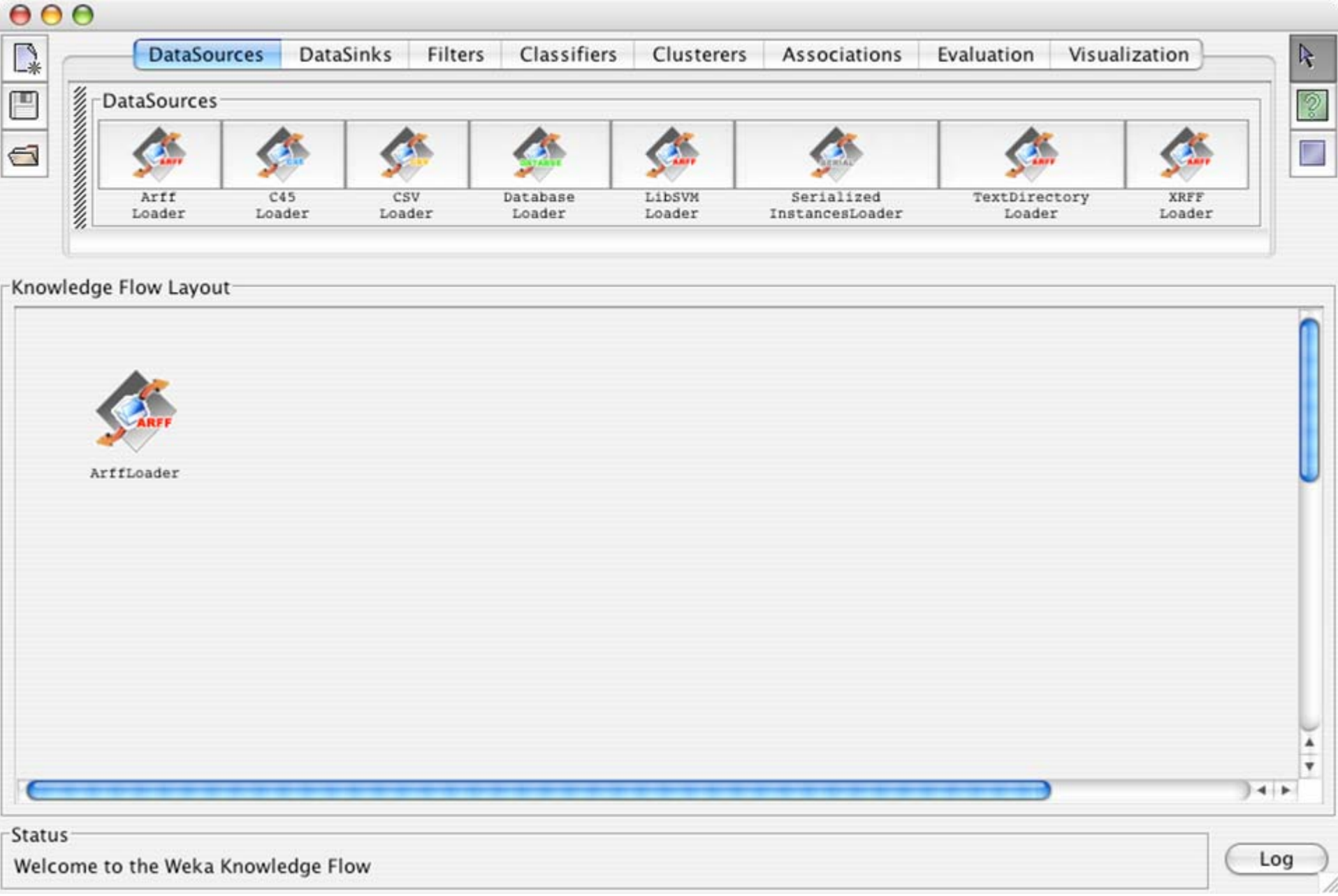


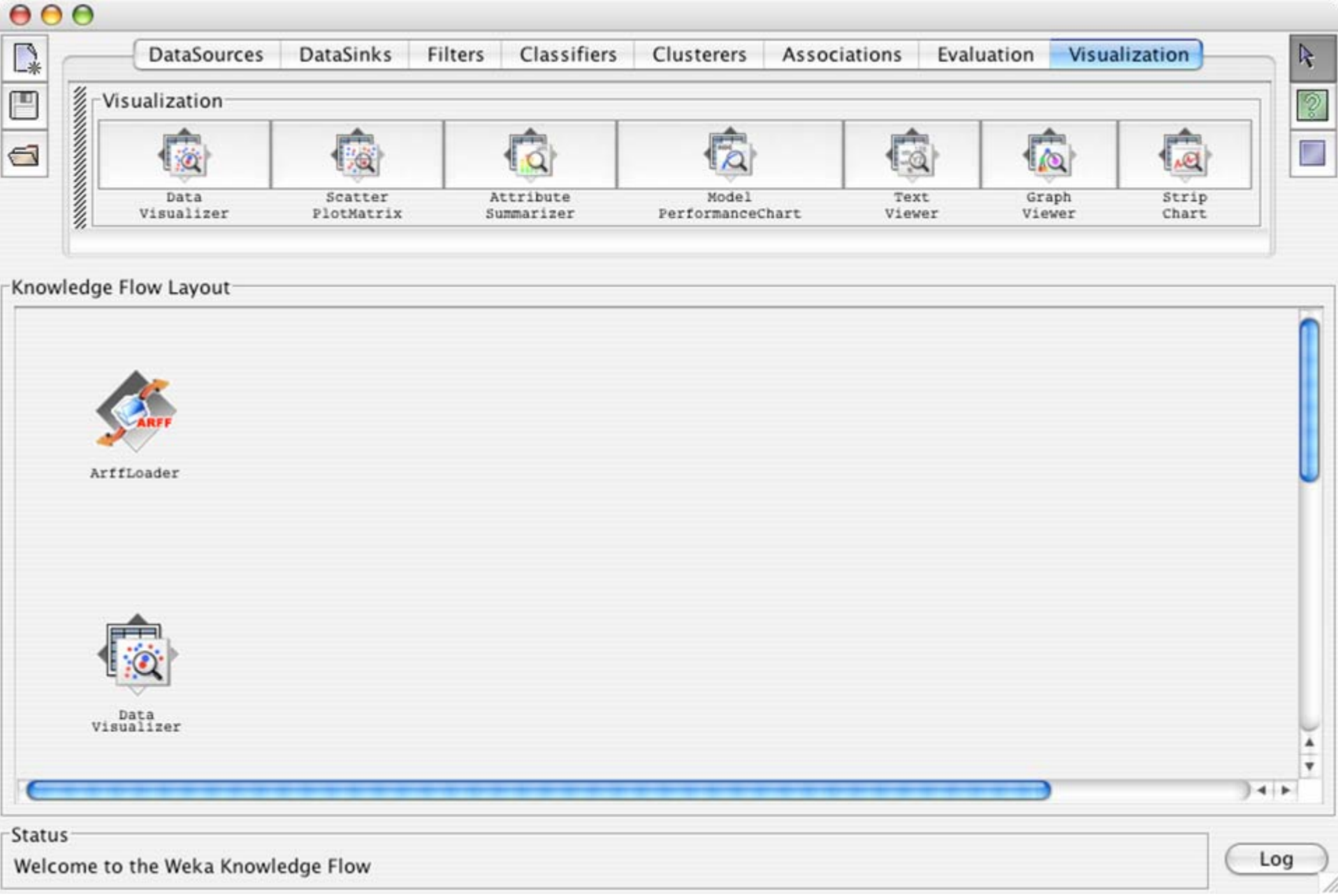
The Knowledge Flow GUI

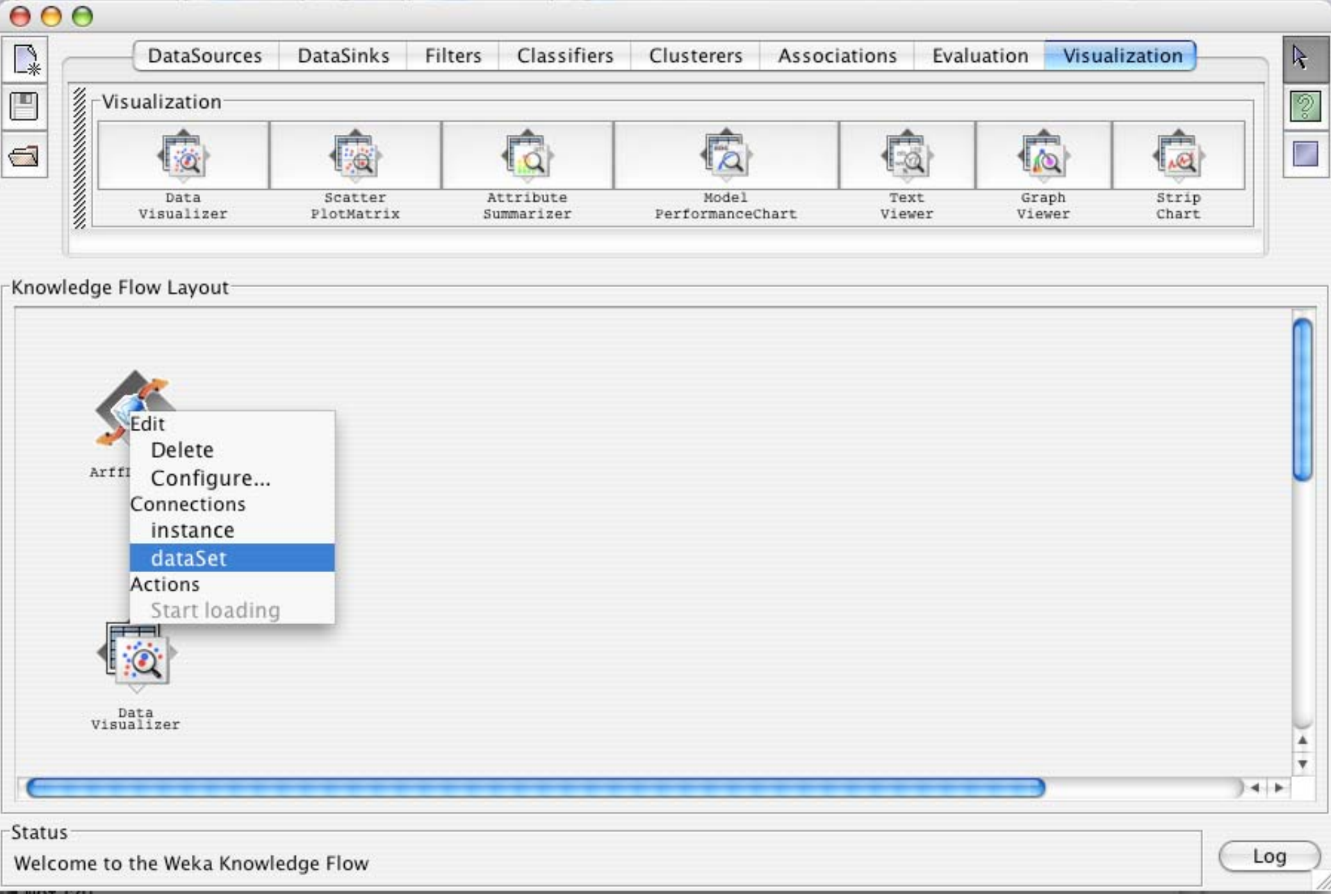
- Java-Beans-based interface for setting up and running machine learning experiments
- Data sources, classifiers, etc. are beans and can be connected graphically
- Data “flows” through components: e.g.,

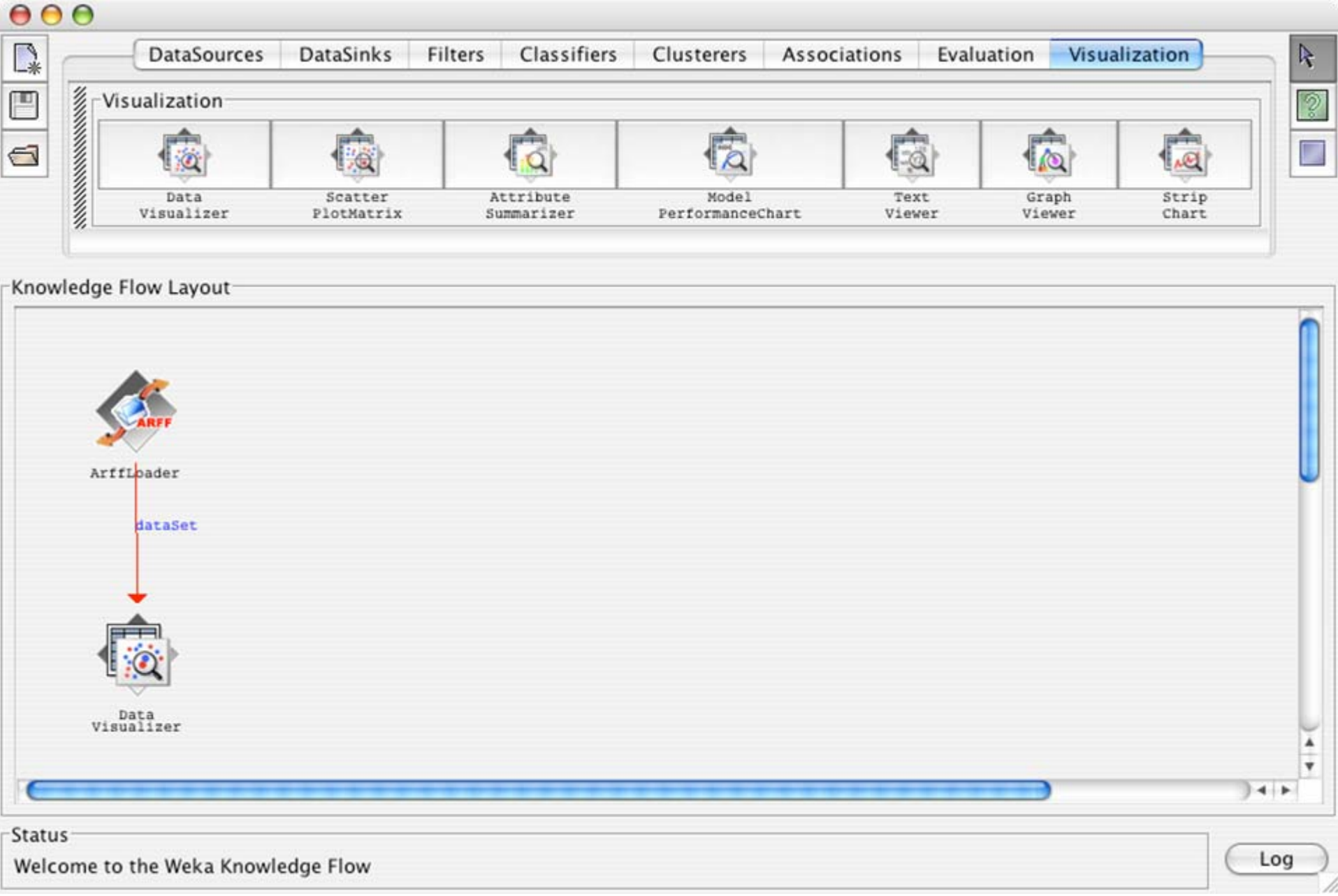
  “data source” -> “filter” -> “classifier” -> “evaluator”
- Layouts can be saved and loaded again later
- cf. Clementine ™

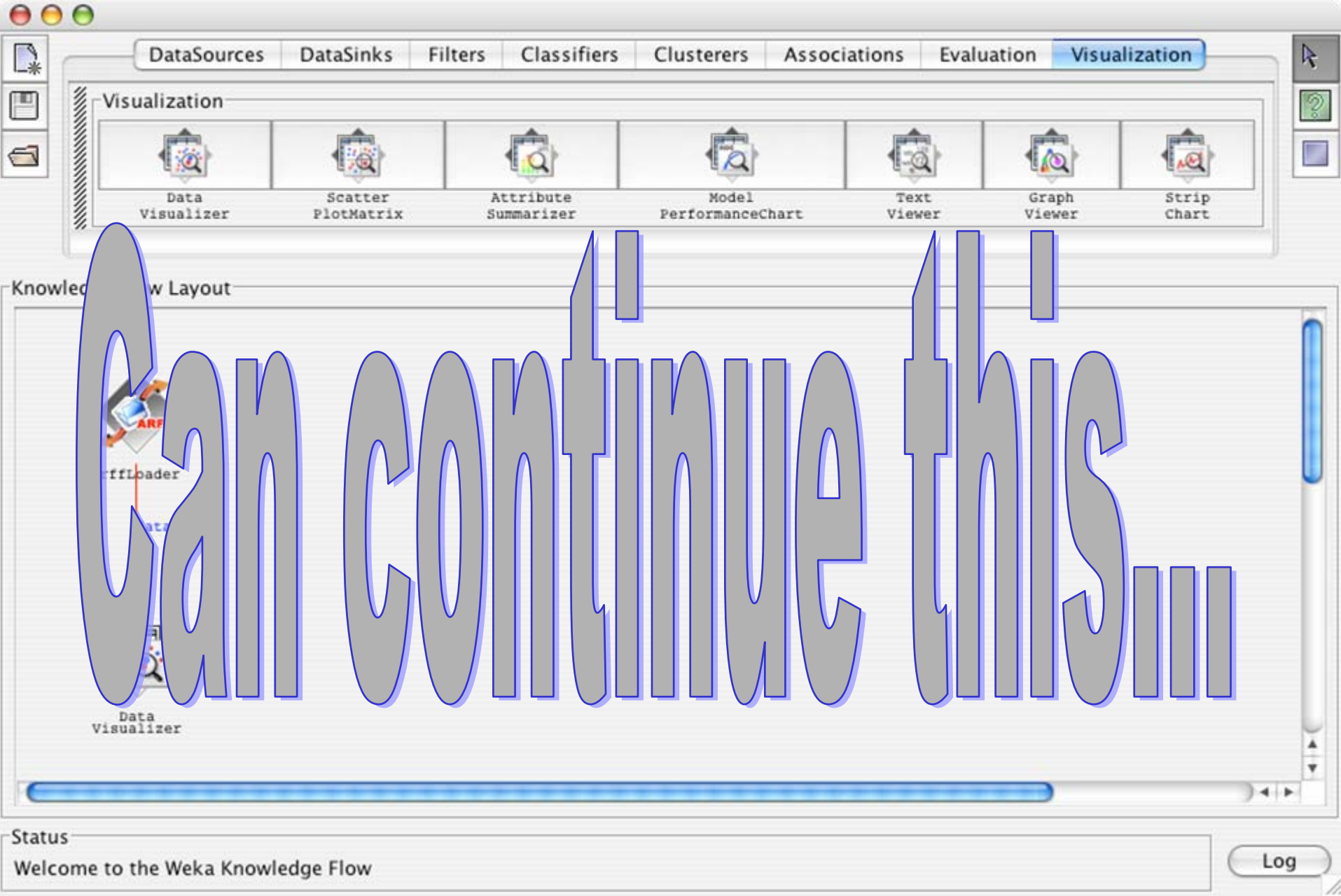

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
DataSources
Arff Loader
C45 Loader
CSV Loader
Database Loader
LibSVM Loader
Serialized InstancesLoader
TextDirectory Loader
XRFF Loader
Knowledge Flow Layout
Status
Welcome to the Weka Knowledge Flow
Log

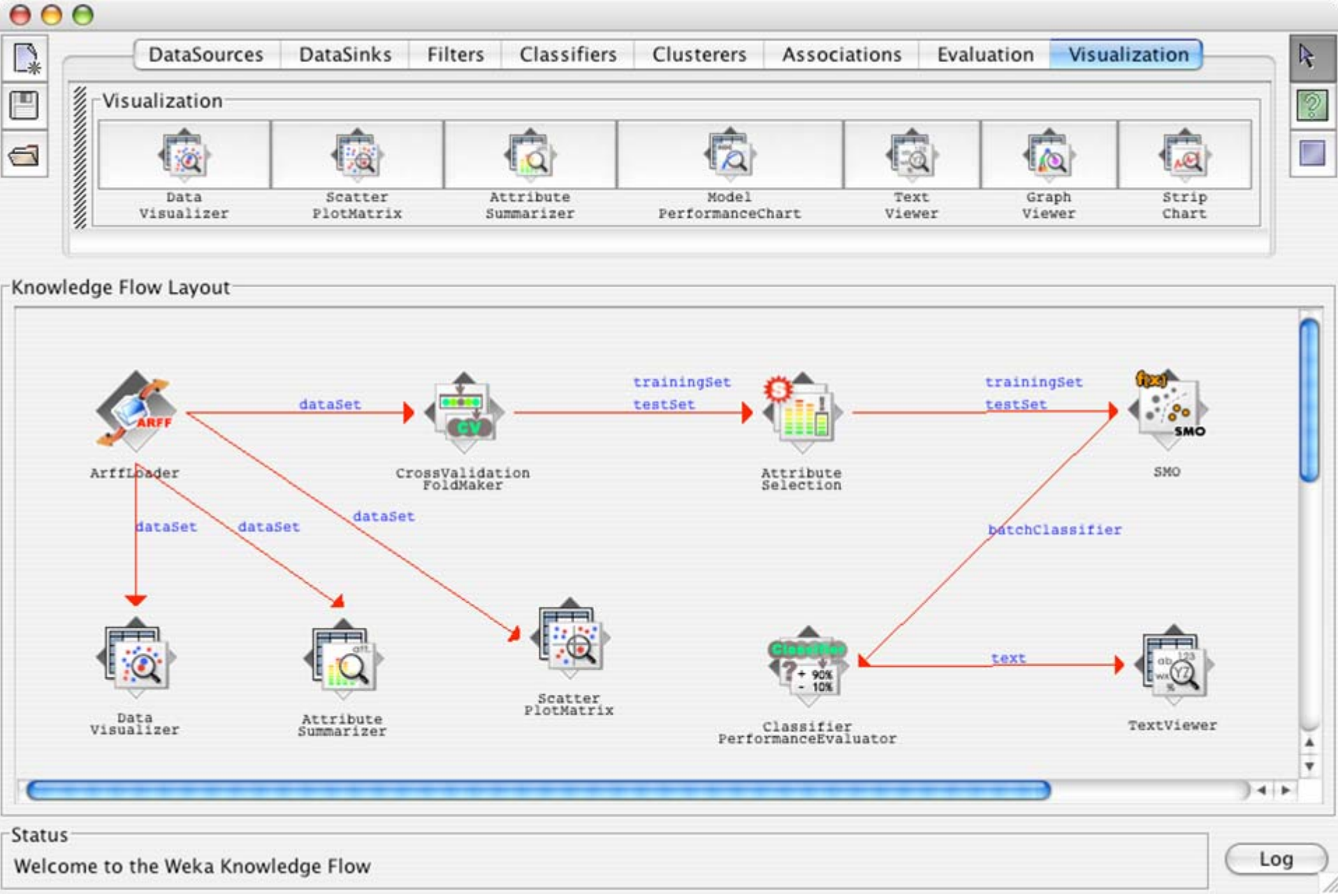

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
DataSources
Arff Loader
C45 Loader
CSV Loader
Database Loader
LibSVM Loader
Serialized InstancesLoader
TextDirectory Loader
XRFF Loader
Knowledge Flow Layout
ARFF
ArffLoader
Status
Welcome to the Weka Knowledge Flow
Log

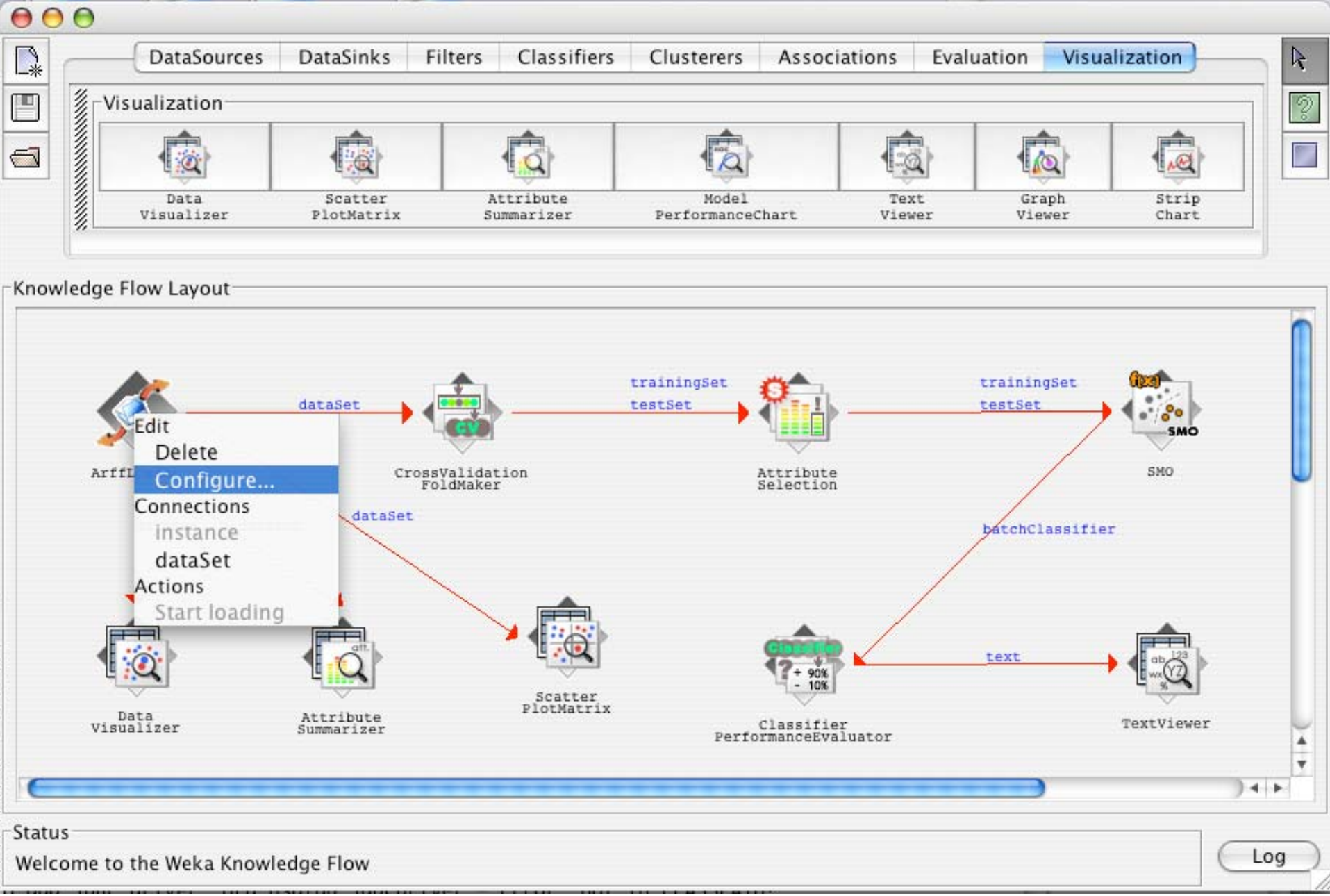

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
ARFF
ArffLoader
Data Visualizer
Status
Welcome to the Weka Knowledge Flow
Log

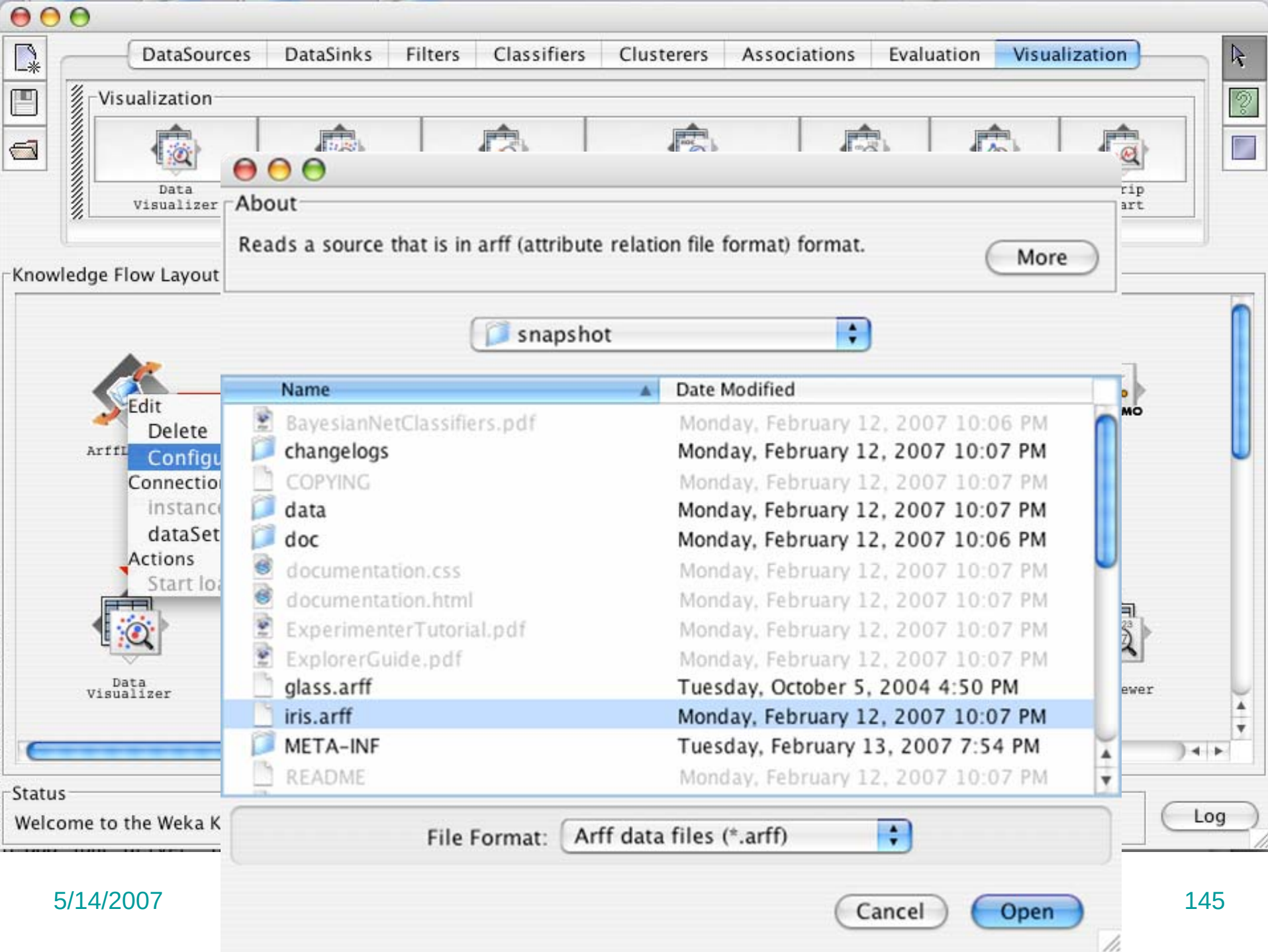

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
Edit
Delete
Configure...
Connections
instance
dataSet
Actions
Start loading
Data Visualizer
Status
Welcome to the Weka Knowledge Flow
Log

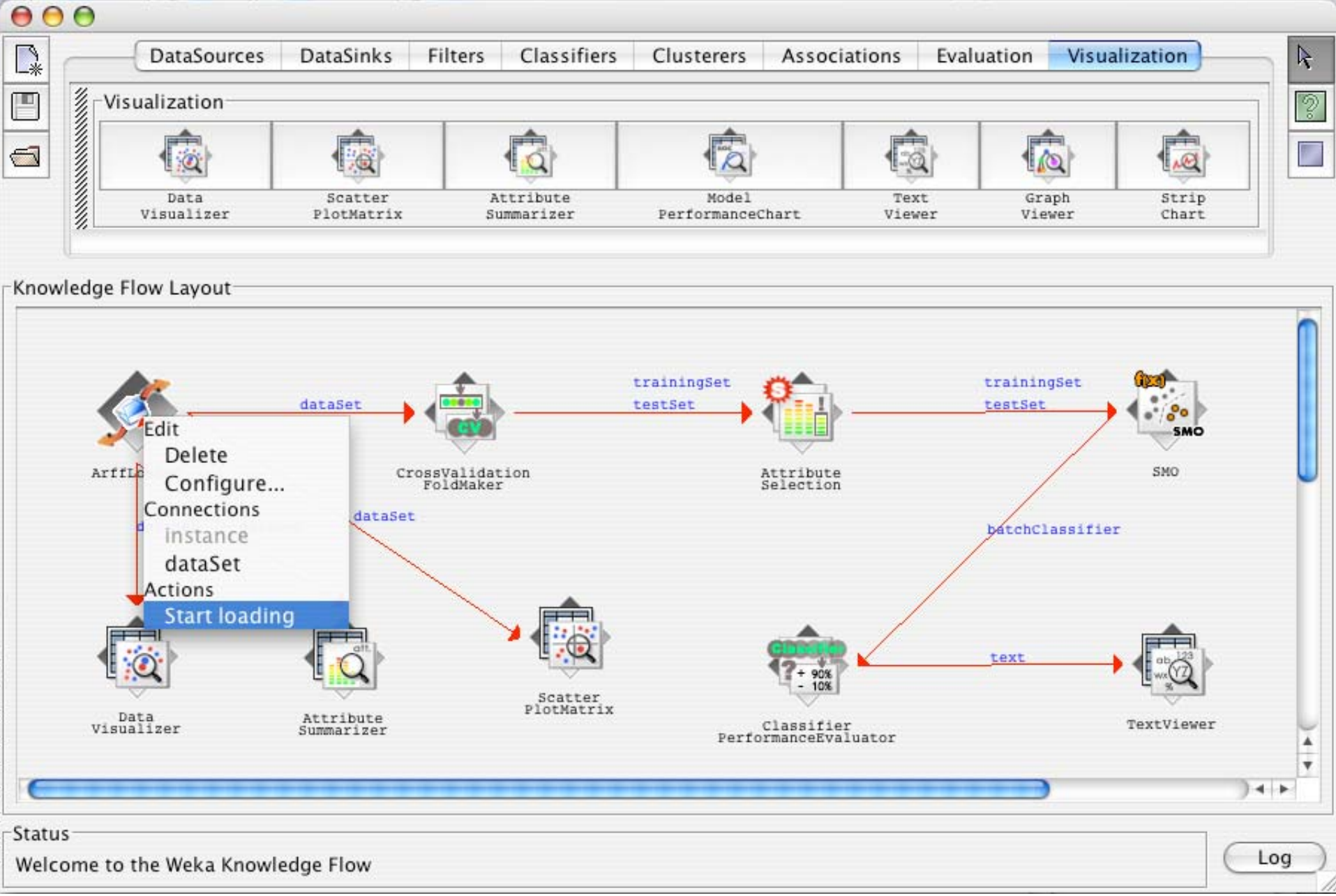

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
ARFF
ArffLoader
dataSet
Data Visualizer
Status
Welcome to the Weka Knowledge Flow
Log

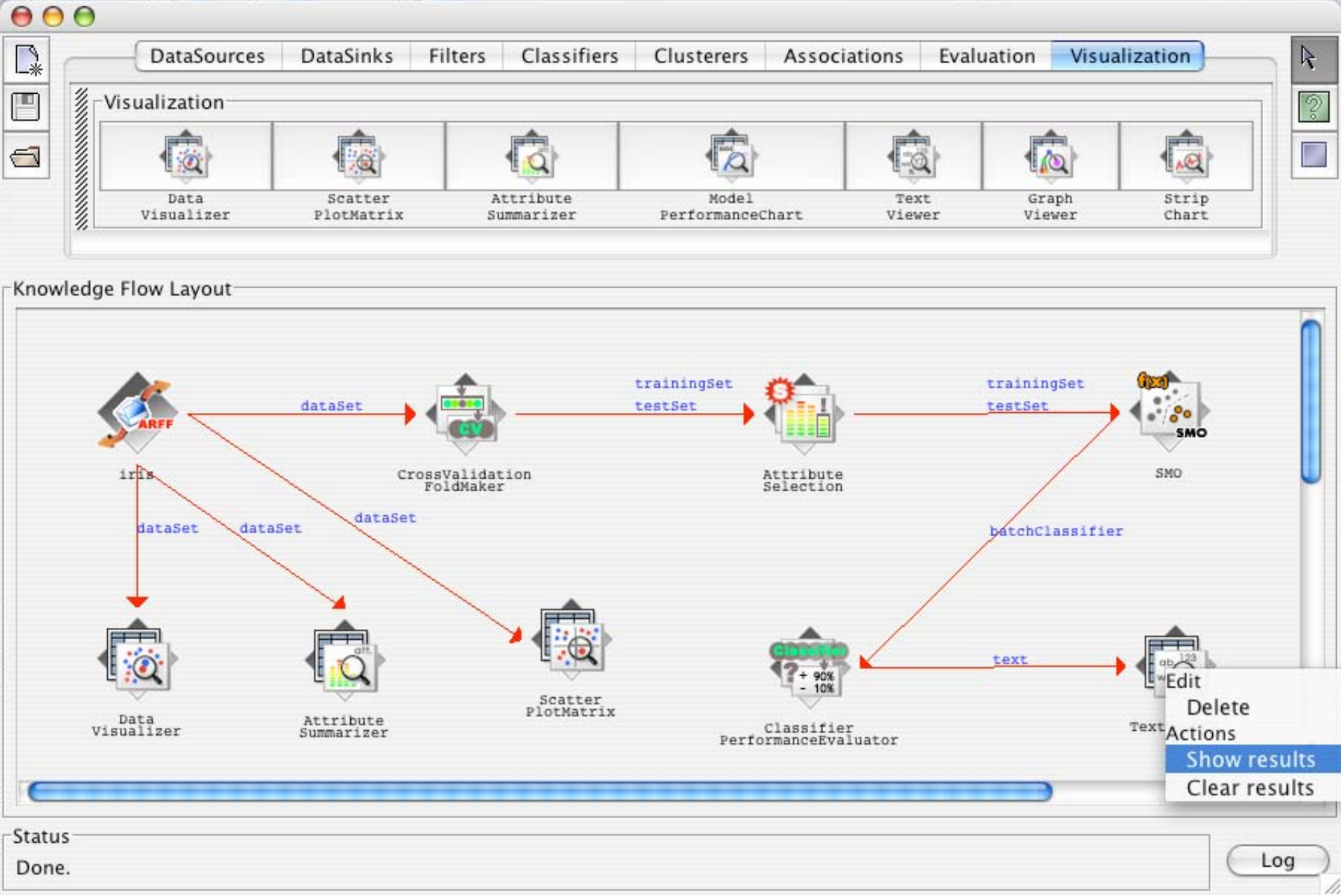

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Can continue this...
Data Visualizer
Status
Welcome to the Weka Knowledge Flow
Log

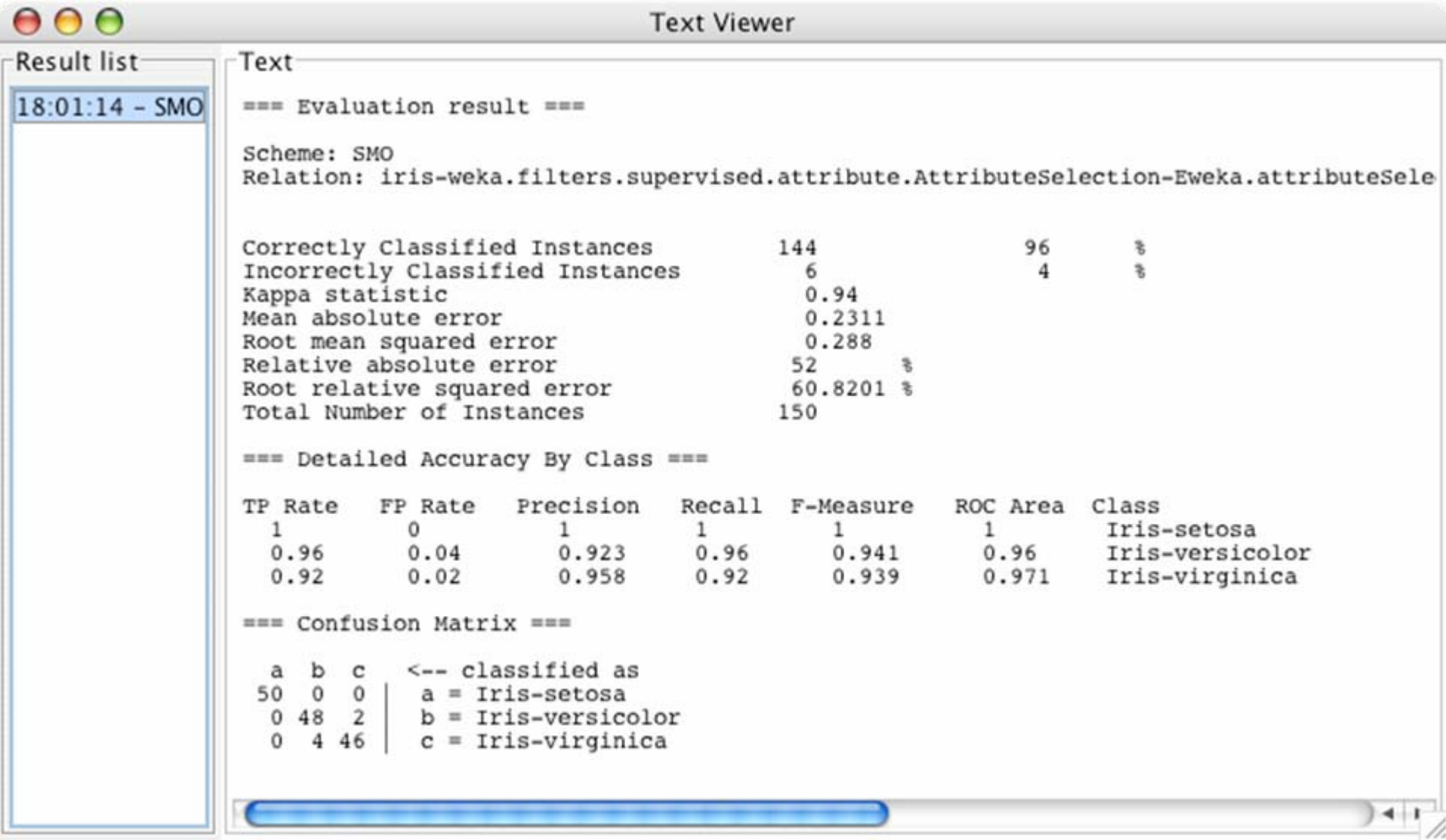

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
ARFF
ArffLoader
dataSet
CV
CrossValidation FoldMaker
trainingSet
testSet
Attribute Selection
trainingSet
testSet
SMO
SMO
dataSet
dataSet
dataSet
batchClassifier
Data Visualizer
Attribute Summarizer
Scatter PlotMatrix
Classifier
+ 90%
- 10%
Classifier PerformanceEvaluator
text
TextViewer
Status
Welcome to the Weka Knowledge Flow
Log


DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
dataSet
trainingSet
testSet
trainingSet
testSet
Edit
Delete
Configure...
Connections
instance
dataSet
Actions
Start loading
CrossValidation FoldMaker
Attribute Selection
SMO
dataSet
batchClassifier
Data Visualizer
Attribute Summarizer
Scatter PlotMatrix
Classifier PerformanceEvaluator
text
TextViewer
Status
Welcome to the Weka Knowledge Flow
Log


DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
About
Reads a source that is in arff (attribute relation file format) format.
More
Knowledge Flow Layout
snapshot
Name
Date Modified
BayesianNetClassifiers.pdf
Monday, February 12, 2007 10:06 PM
changelogs
Monday, February 12, 2007 10:07 PM
COPYING
Monday, February 12, 2007 10:07 PM
data
Monday, February 12, 2007 10:07 PM
doc
Monday, February 12, 2007 10:06 PM
documentation.css
Monday, February 12, 2007 10:07 PM
documentation.html
Monday, February 12, 2007 10:07 PM
ExperimenterTutorial.pdf
Monday, February 12, 2007 10:07 PM
ExplorerGuide.pdf
Monday, February 12, 2007 10:07 PM
glass.arff
Tuesday, October 5, 2004 4:50 PM
iris.arff
Monday, February 12, 2007 10:07 PM
META-INF
Tuesday, February 13, 2007 7:54 PM
README
Monday, February 12, 2007 10:07 PM
Edit
Delete
Actions
Data Visualizer
Status
Welcome to the Weka K
Log
File Format:
Arff data files (*.arff)
Cancel
Open

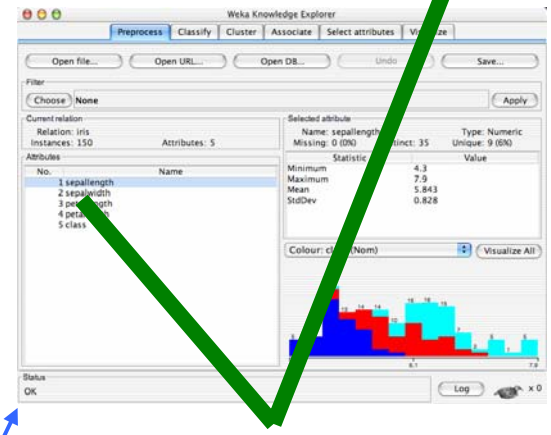

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
dataSet
trainingSet
testSet
trainingSet
testSet
batchClassifier
text
Edit
Delete
Configure...
Connections
instance
dataSet
Actions
Start loading
CrossValidation FoldMaker
Attribute Selection
SMO
Data Visualizer
Attribute Summarizer
Scatter PlotMatrix
Classifier PerformanceEvaluator
TextViewer
Status
Welcome to the Weka Knowledge Flow
Log

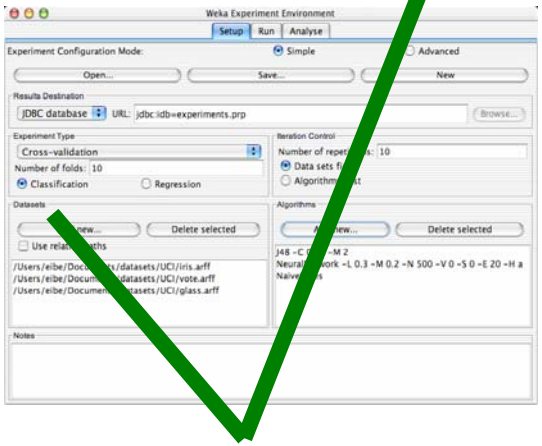

DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
ARFF
iris
dataSet
CV
CrossValidation FoldMaker
trainingSet
testSet
Attribute Selection
trainingSet
testSet
SMO
SMO
dataSet
dataSet
dataSet
batchClassifier
Data Visualizer
Attribute Summarizer
Scatter PlotMatrix
+ 90%
- 10%
Classifier PerformanceEvaluator
text
Text
Edit
Delete
Actions
Show results
Clear results
Status
Done.
Log

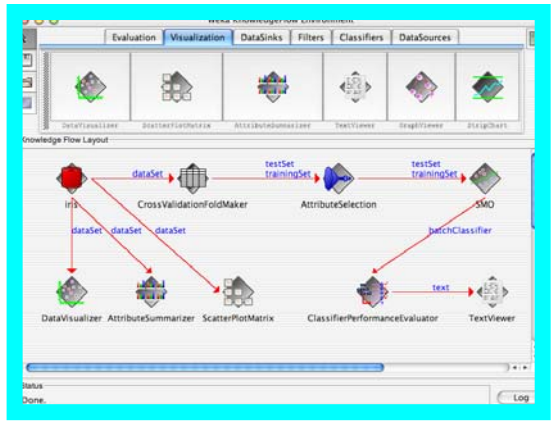

Text Viewer
Result list
18:01:14 - SMO
Text
=== Evaluation result ===
Scheme: SMO
Relation: iris-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSele
Correctly Classified Instances 144 96 %
Incorrectly Classified Instances 6 4 %
Kappa statistic 0.94
Mean absolute error 0.2311
Root mean squared error 0.288
Relative absolute error 52 %
Root relative squared error 60.8201 %
Total Number of Instances 150
=== Detailed Accuracy By Class ===
TP Rate FP Rate Precision Recall F-Measure ROC Area Class
1 0 1 1 1 1 Iris-setosa
0.96 0.04 0.923 0.96 0.941 0.96 Iris-versicolor
0.92 0.02 0.958 0.92 0.939 0.971 Iris-virginica
=== Confusion Matrix ===
a b c <-- classified as
50 0 0 | a = Iris-setosa
0 48 2 | b = Iris-versicolor
0 4 46 | c = Iris-virginica


DataSources
DataSinks
Filters
Classifiers
Clusterers
Associations
Evaluation
Visualization
Visualization
Data Visualizer
Scatter PlotMatrix
Attribute Summarizer
Model PerformanceChart
Text Viewer
Graph Viewer
Strip Chart
Knowledge Flow Layout
ARFF
ArffLoader
instance
lazy
K=
IBK
IBk
incrementalClassifier
+ 90%
- 10%
Incremental ClassifierEvaluator
chart
StripChart
Strip Chart
Legend
Accuracy
RMSE (prob)
0.87
0.91
0.36
0.19
0.16
1.0
0.5
0.0
0
500
1000
Status
Done.
Log

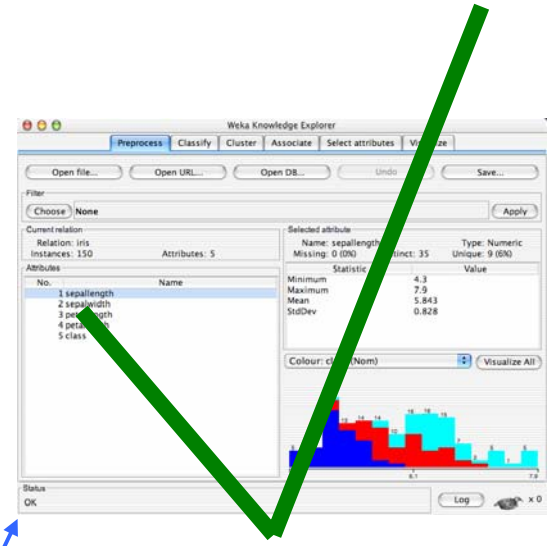

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log

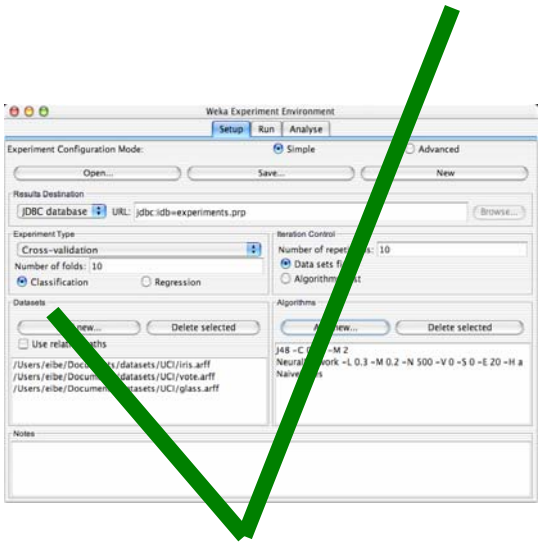


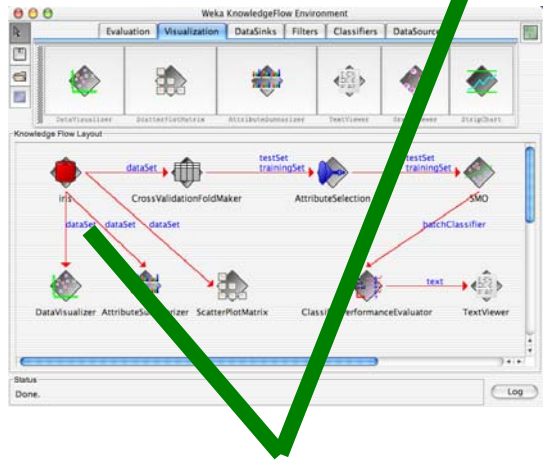




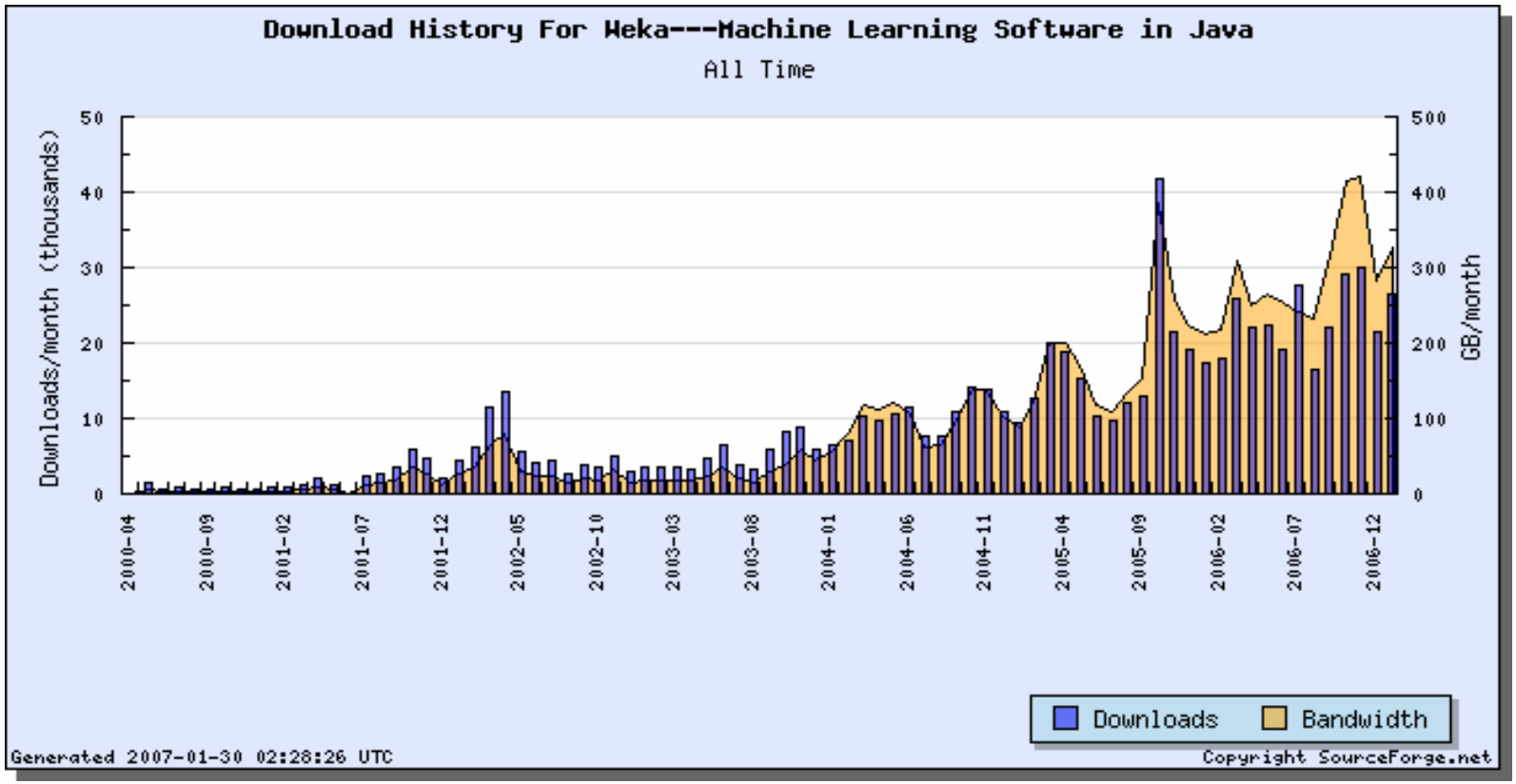

Weka GUI Chooser
Waikato Environment for Knowledge Analysis
Version 3.5.5
(c) 1999 - 2007
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow
ArffViewer
Log



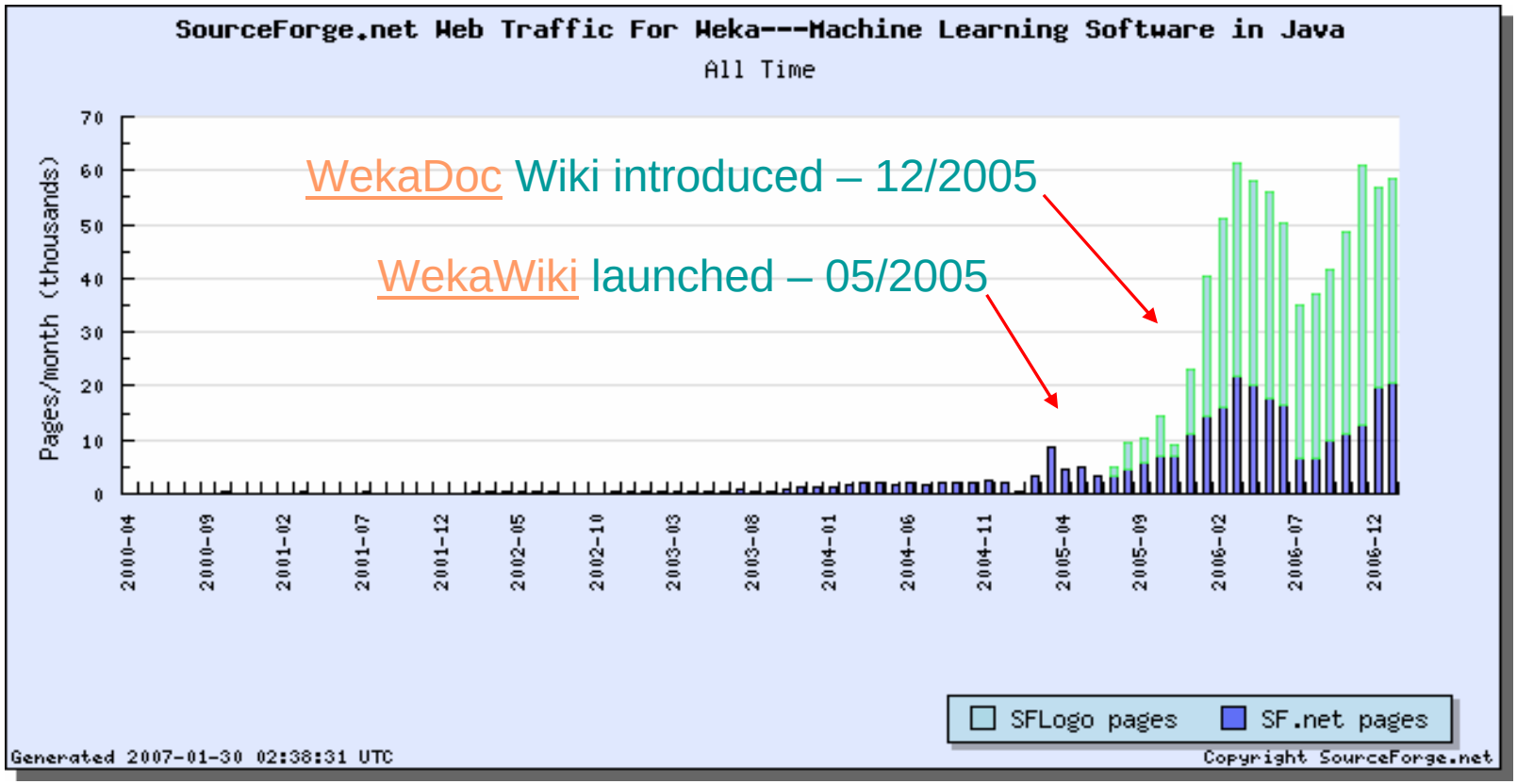




Sourceforge.net – Downloads



Sourceforge.net – Web Traffic



Projects based on WEKA

- 45 projects currently (30/01/07) listed on the WekaWiki
- Incorporate/wrap WEKA
  - GRB Tool Shed - a tool to aid gamma ray burst research
  - YALE - facility for large scale ML experiments
  - GATE - NLP workbench with a WEKA interface
  - Judge - document clustering and classification
  - RWeka - an R interface to Weka
- Extend/modify WEKA
  - BioWeka - extension library for knowledge discovery in biology
  - WekaMetal - meta learning extension to WEKA
  - Weka-Parallel - parallel processing for WEKA
  - Grid Weka - grid computing using WEKA
  - Weka-CG - computational genetics tool library

WEKA and PENTAHO

- Pentaho – The leader in Open Source Business Intelligence (BI)
- September 2006 – Pentaho acquires the Weka project (exclusive license and SF.net page)
- Weka will be used/integrated as data mining component in their BI suite
- Weka will be still available as GPL open source software
- Most likely to evolve 2 editions:
  - Community edition
  - BI oriented edition

Limitations of WEKA

- Traditional algorithms need to have all data in main memory
- ==> big datasets are an issue
- Solution:
  - Incremental schemes
  - Stream algorithms

    MOA "**M**assive **O**nline **A**nalysis"

    (not only a *flightless* bird, but also *extinct*!)

Conclusion: try it yourself!

- WEKA is available at

  http://www.cs.waikato.ac.nz/ml/weka

- Also has a list of projects based on WEKA
- (probably incomplete list of) WEKA contributors:

  Abdelaziz Mahoui, Alexander K. Seewald, Ashraf M. Kibriya, Bernhard Pfahringer, Brent Martin, Peter Flach, Eibe Frank, Gabi Schmidberger, Ian H. Witten, J. Lindgren, Janice Boughton, Jason Wells, Len Trigg, Lucio de Souza Coelho, Malcolm Ware, Mark Hall, Remco Bouckaert, Richard Kirkby, Shane Butler, Shane Legg, Stuart Inglis, Sylvain Roy, Tony Voyle, Xin Xu, Yong Wang, Zhihai Wang