# Machine Learning with WEKA

Eibe Frank

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
- Conclusions

# WEKA: the bird

*Copyright: Martin Kramer (mkramer@wxs.nl)*

# 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

# WEKA: versions

- There are several versions of WEKA:
  - WEKA 3.0: “book version” compatible with description in data mining book
  - WEKA 3.2: “GUI version” adds graphical user interfaces (book version is command-line only)
  - WEKA 3.3: “development version” with lots of improvements
- This talk is based on the latest snapshot of WEKA 3.3 (soon to be WEKA 3.4)

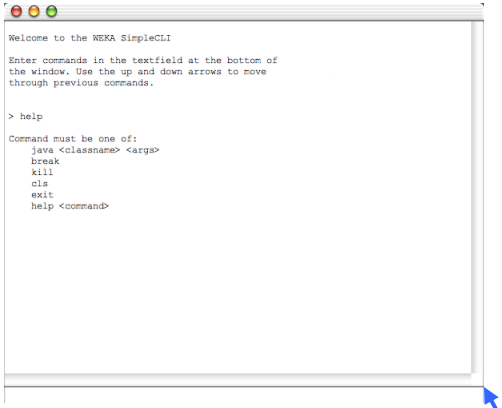
# WEKA only deals with “flat” files

```
@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
...
```

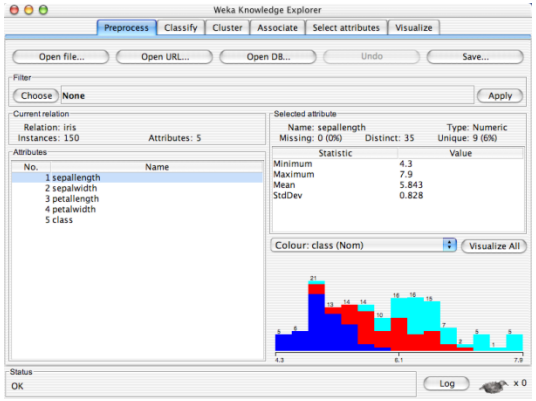
Flat file in ARFF format

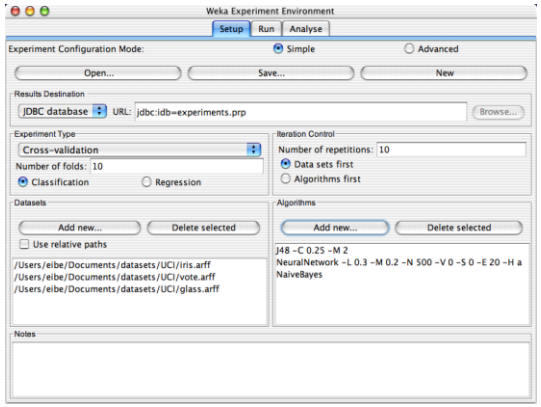
# WEKA only deals with “flat” files

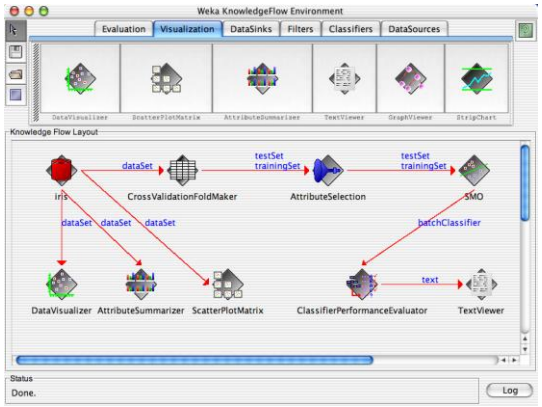

```
@relation heart-disease-simplified

@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
...
```


Welcome to the WEKA SimpleCLI
Enter commands in the textfield at the bottom of
the window. Use the up and down arrows to move
through previous commands.
> help
Command must be one of:
java <classname> <args>
break
kill
cls
exit
help <command>


Weka GUI Chooser
Waikato Environment for
Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow

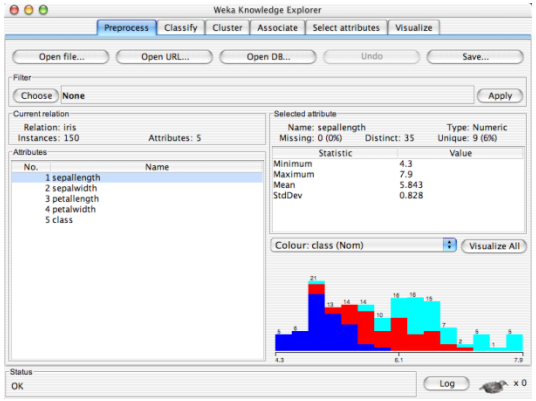

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Selected attribute
Name: sepallength
Missing: 0 (0%)
Distinct: 35
Type: Numeric
Unique: 9 (6%)
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Statistic
Value
Minimum 4.3
Maximum 7.9
Mean 5.843
StdDev 0.828
Colour: class (Nom)
Visualize All
Status
OK
Log
x 0

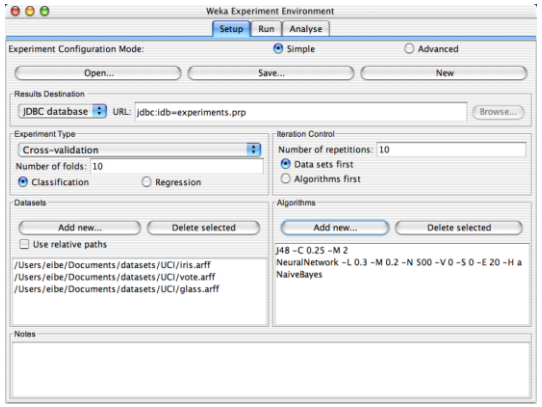

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
JDBC database
URL:
jdbc:idb=experiments.prp
Browse...
Experiment Type
Cross-validation
Number of folds: 10
Classification
Regression
Iteration Control
Number of repetitions: 10
Data sets first
Algorithms first
Datasets
Add new...
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
Algorithms
Add new...
Delete selected
J48 -C 0.25 -M 2
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
NaiveBayes
Notes

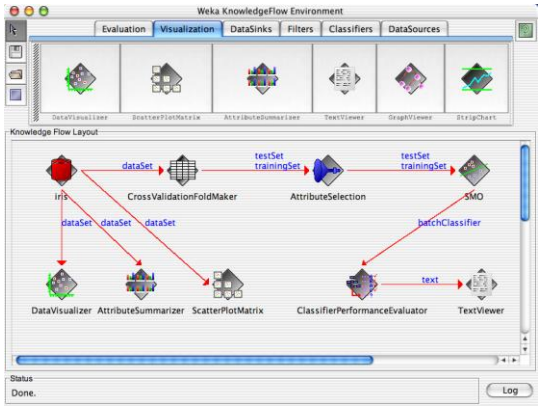

Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources
Knowledge Flow Layout
iris
dataSet
CrossValidationFoldMaker
testSet
trainingSet
AttributeSelection
SMO
batchClassifier
DataVisualizer
AttributeSummarizer
ScatterPlotMatrix
ClassifierPerformanceEvaluator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


Welcome to the WEKA SimpleCLI
> help
Command must be one of:
java <classname> <args>
break
kill
cls
exit
help <command>


Weka GUI Chooser
Waikato Environment for Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow

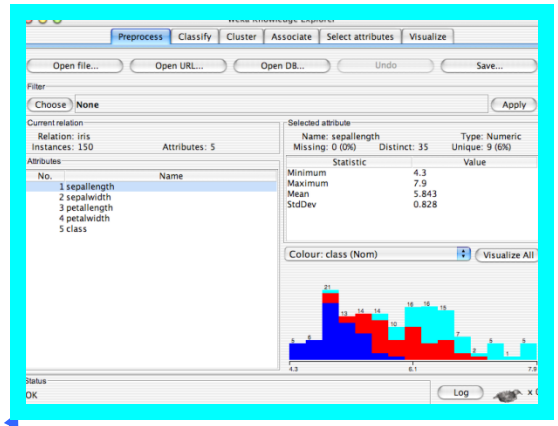

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Selected attribute
Name: sepallength
Missing: 0 (0%)
Distinct: 35
Type: Numeric
Unique: 9 (6%)
Statistic
Value
Minimum 4.3
Maximum 7.9
Mean 5.843
StdDev 0.828
Colour: class (Nom)
Visualize All
Status
OK
Log
x 0

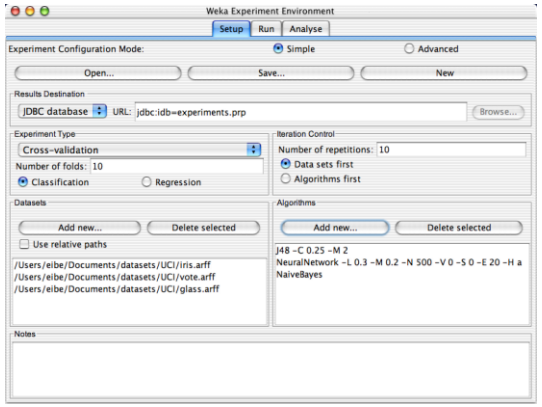

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
JDBC database
URL:
jdbc:idb=experiments.prp
Browse...
Experiment Type
Cross-validation
Number of folds: 10
Classification
Regression
Iteration Control
Number of repetitions: 10
Data sets first
Algorithms first
Datasets
Add new...
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
Algorithms
Add new...
Delete selected
J48 -C 0.25 -M 2
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
NaiveBayes
Notes

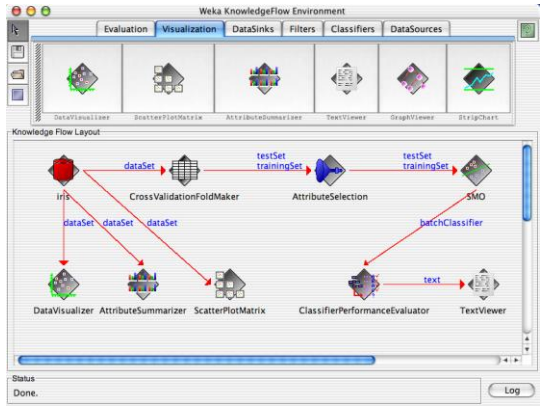

Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources
Knowledge Flow Layout
iris
dataSet
CrossValidationFoldMaker
testSet
trainingSet
AttributeSelection
SMO
batchClassifier
DataVisualizer
AttributeSummarizer
ScatterPlotMatrix
ClassifierPerformanceEvaluator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


Welcome to the WEKA SimpleCLI
> help
Command must be one of:
java
break
kill
cls
exit
help


Weka GUI Chooser
Waikato Environment for Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow

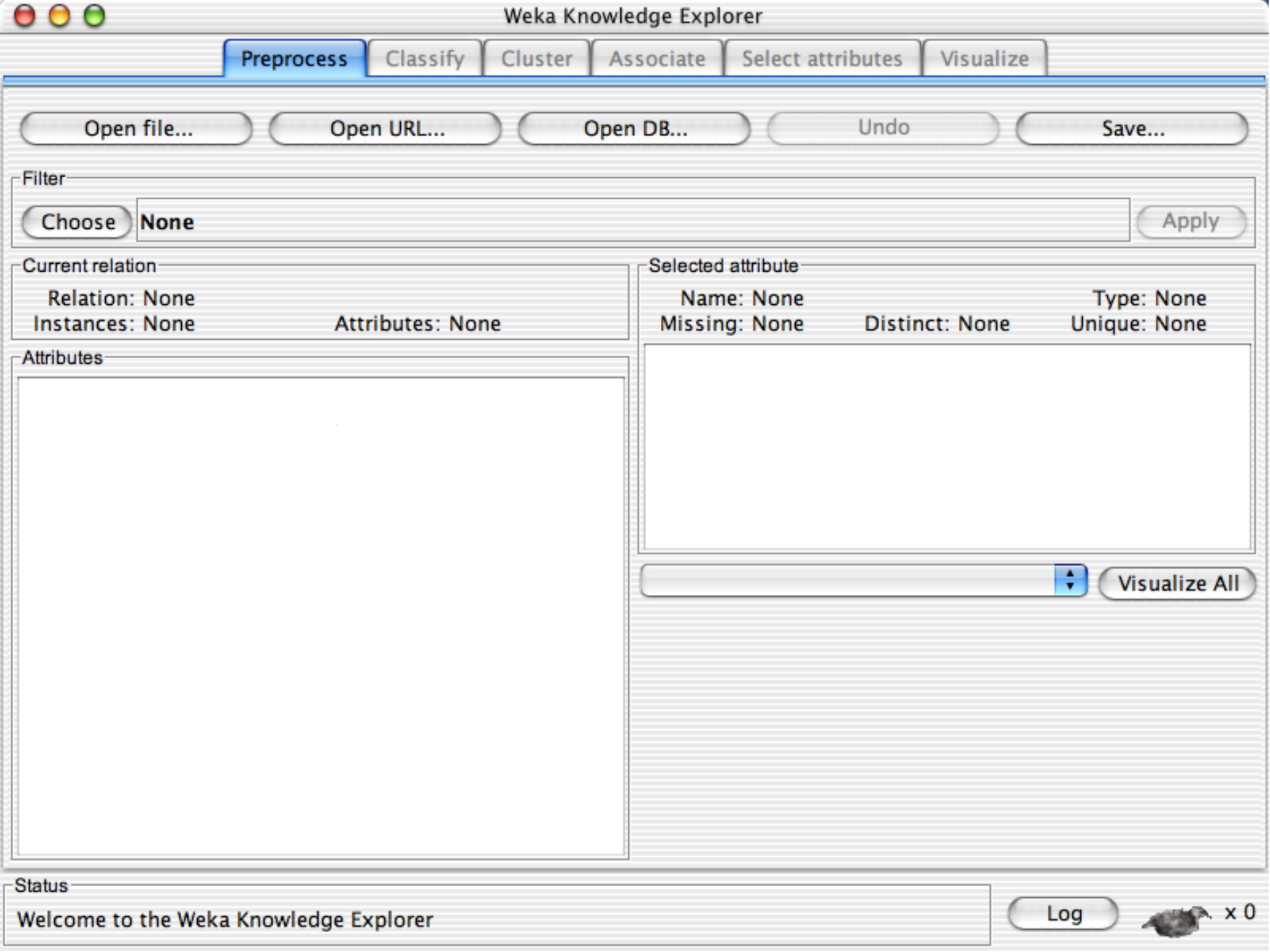

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Choose
None
Apply
Visualize All
Status
OK
Log

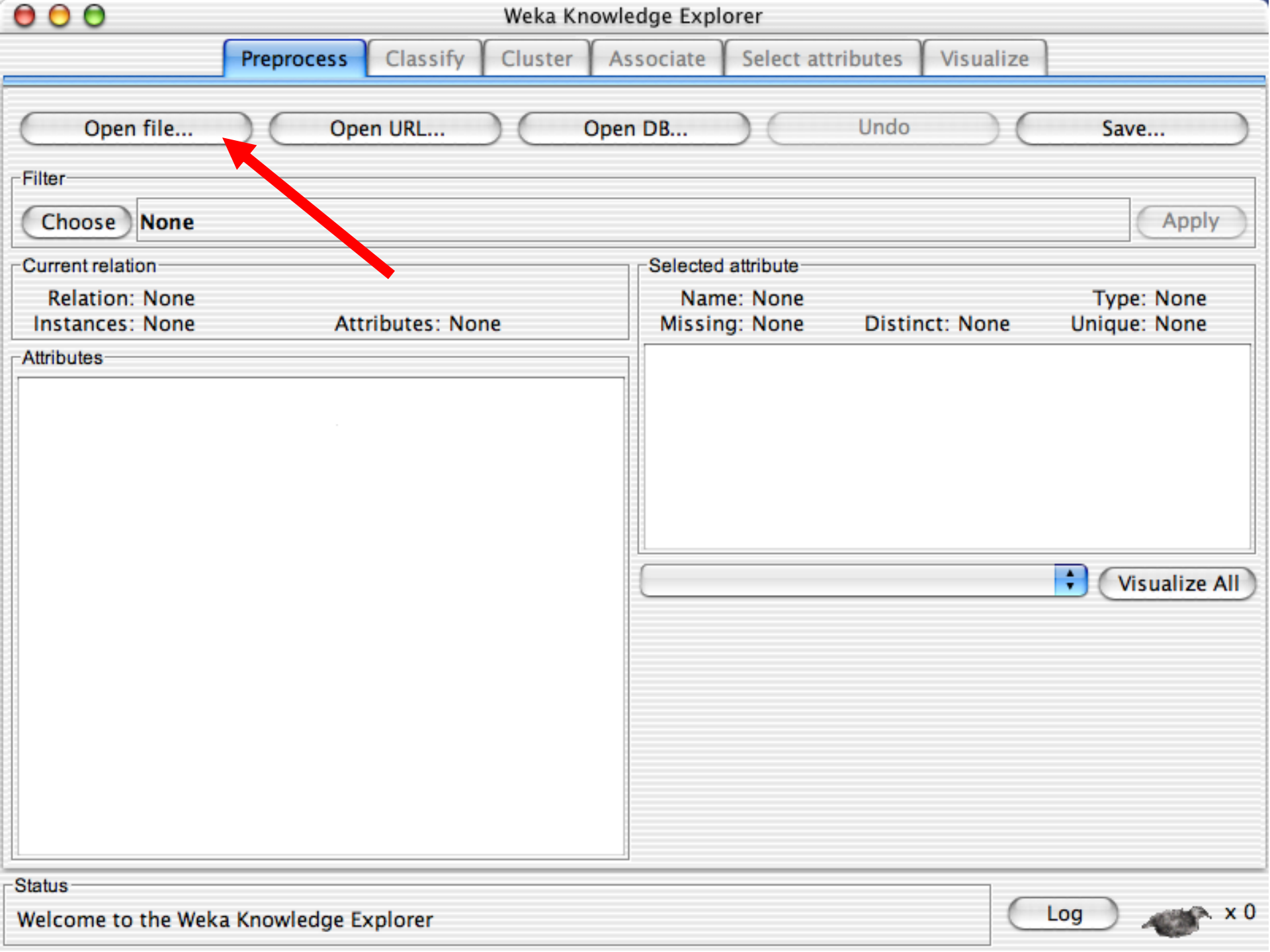

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
JDBC database
URL:
jdbc:idb=experiments.prp
Browse...
Experiment Type
Cross-validation
Number of folds: 10
Classification
Regression
Iteration Control
Number of repetitions: 10
Data sets first
Algorithms first
Datasets
Add new...
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
Algorithms
Add new...
Delete selected
J48 -C 0.25 -M 2
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
NaiveBayes
Notes


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

# 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 Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
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
Visualize All
Status
Welcome to the Weka Knowledge Explorer
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
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
Visualize All
Status
Welcome to the Weka Knowledge Explorer
Log
x 0

Weka Knowledge Explorer

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

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

Filter

Choose **None** | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

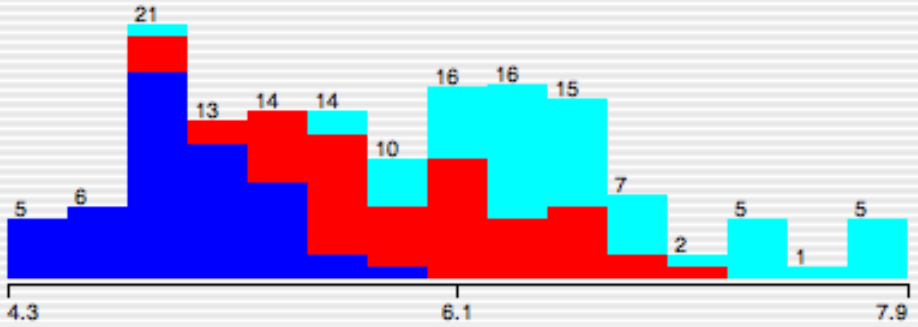
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

Colour: class (Nom) | Visualize All




Status

OK

Log x 0

Weka Knowledge Explorer

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

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

Filter

Choose **None** | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

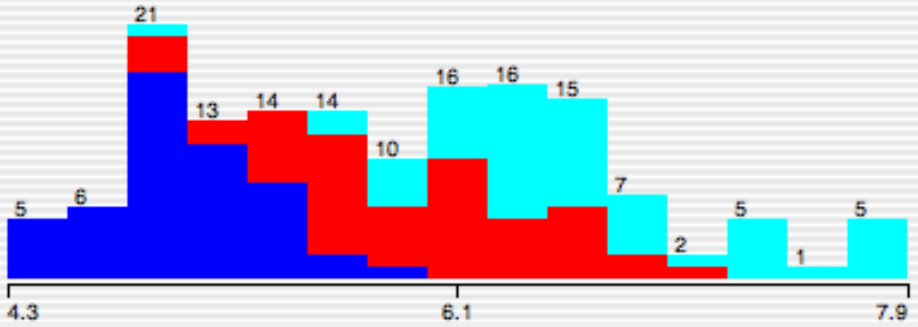
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

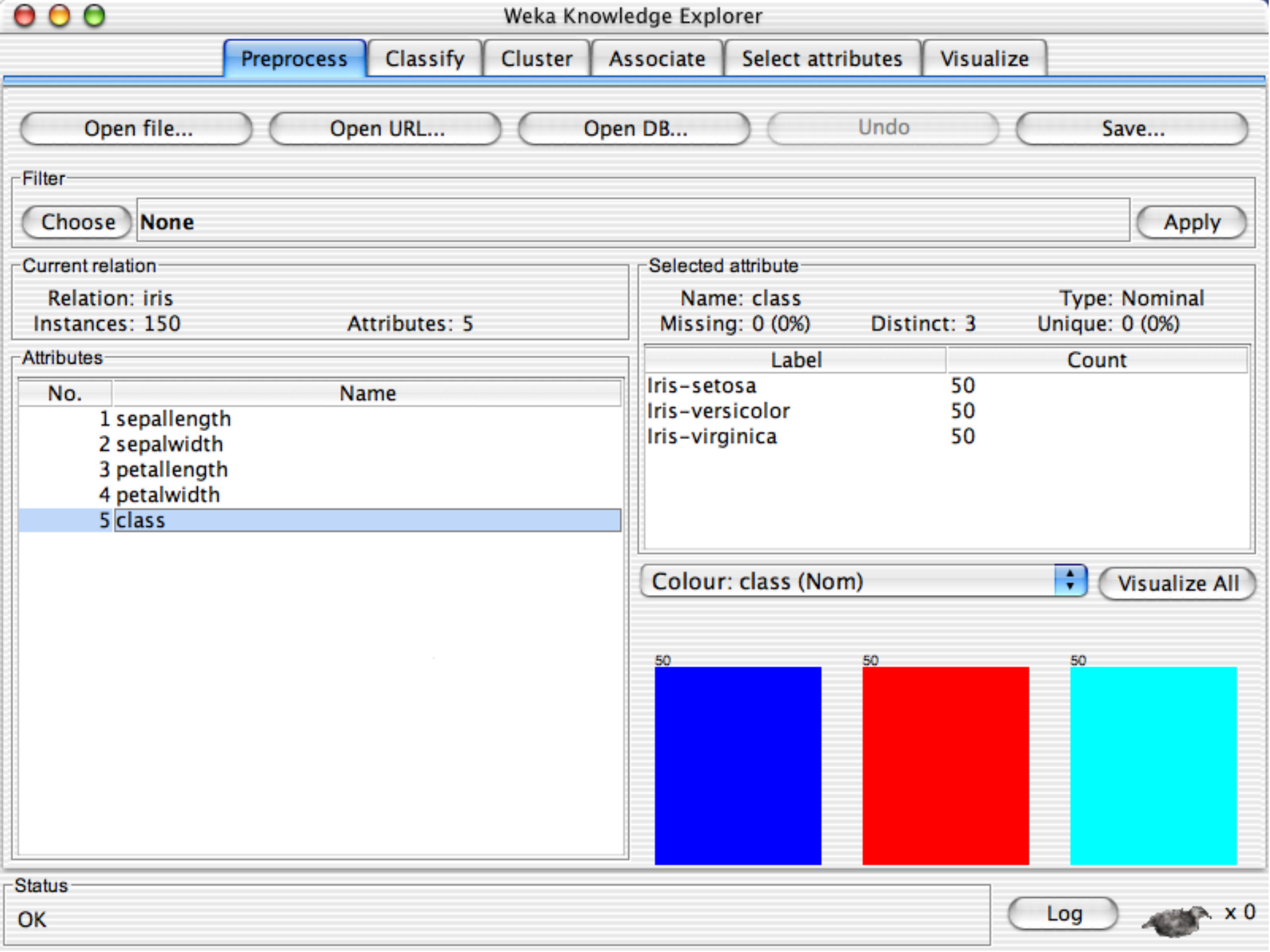
Colour: class (Nom) | Visualize All

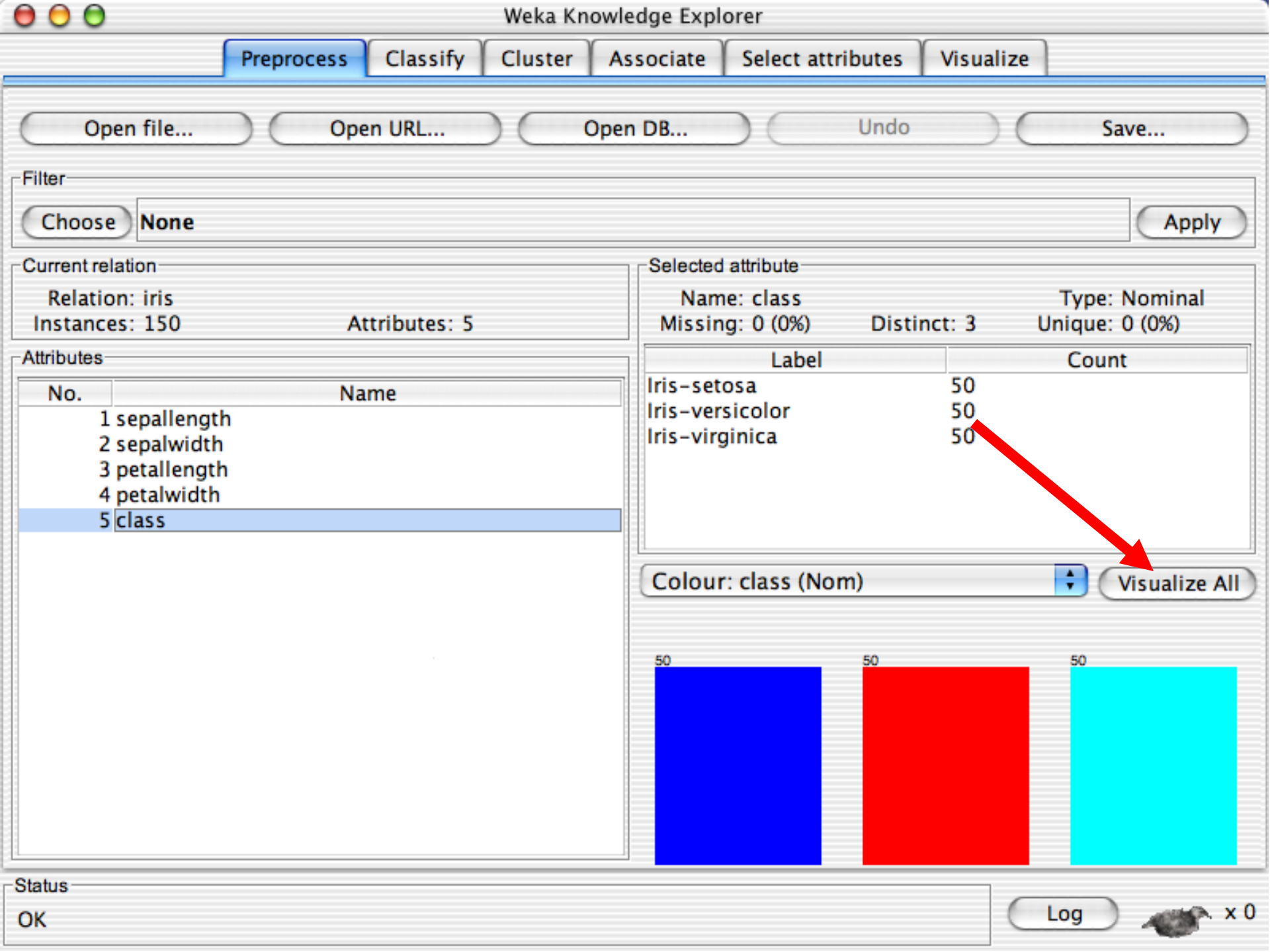


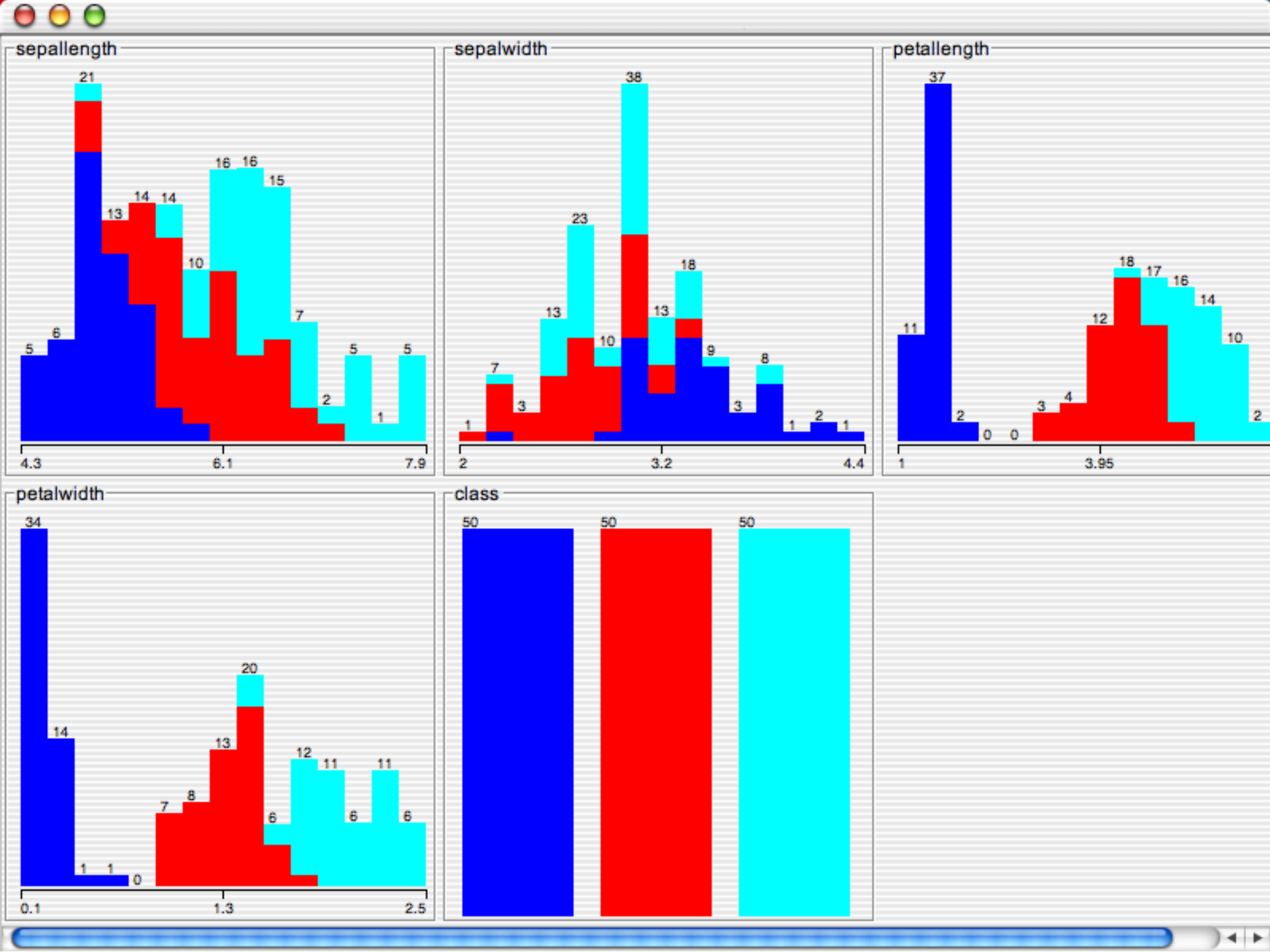

Status

OK

Log | x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
Colour: class (Nom)
Visualize All
50
50
50
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
Colour: class (Nom)
Visualize All
50
50
50
Status
OK
Log
x 0


sepallength
21
16 16
14 14
13
15
10
7
6
5
5
5
2
1
4.3
6.1
7.9
sepalwidth
38
23
18
13
13
10
9
8
7
3
3
2
1
1
2
3.2
4.4
petallength
37
18
17
16
14
12
11
10
4
3
2
2
0
0
1
3.95
petalwidth
34
20
14
13
12
11
11
8
7
6
6
6
1
1
0
0.1
1.3
2.5
class
50
50
50

Weka Knowledge Explorer

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

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

Filter

Choose **None** | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

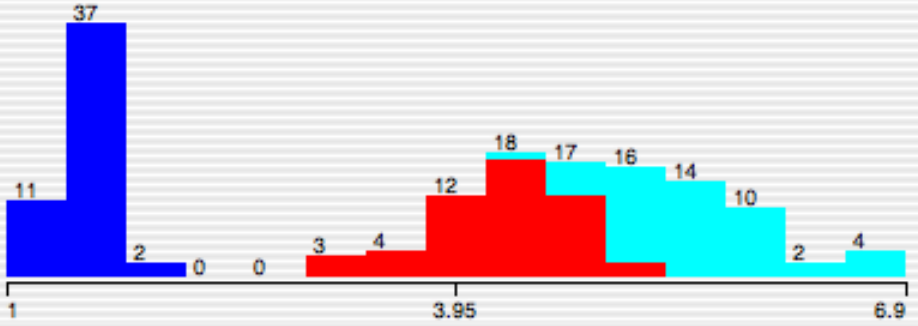
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

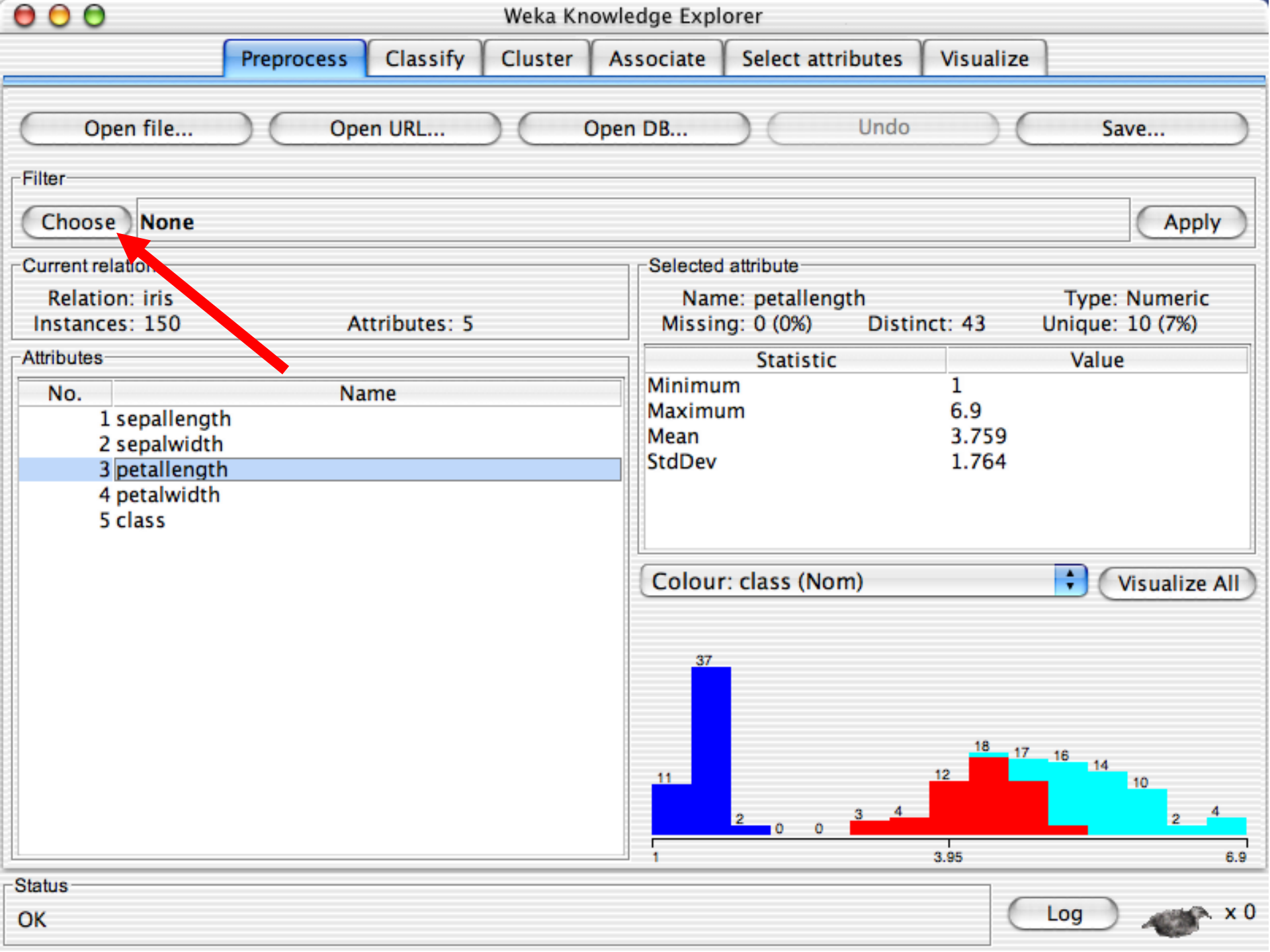
Colour: class (Nom) | Visualize All

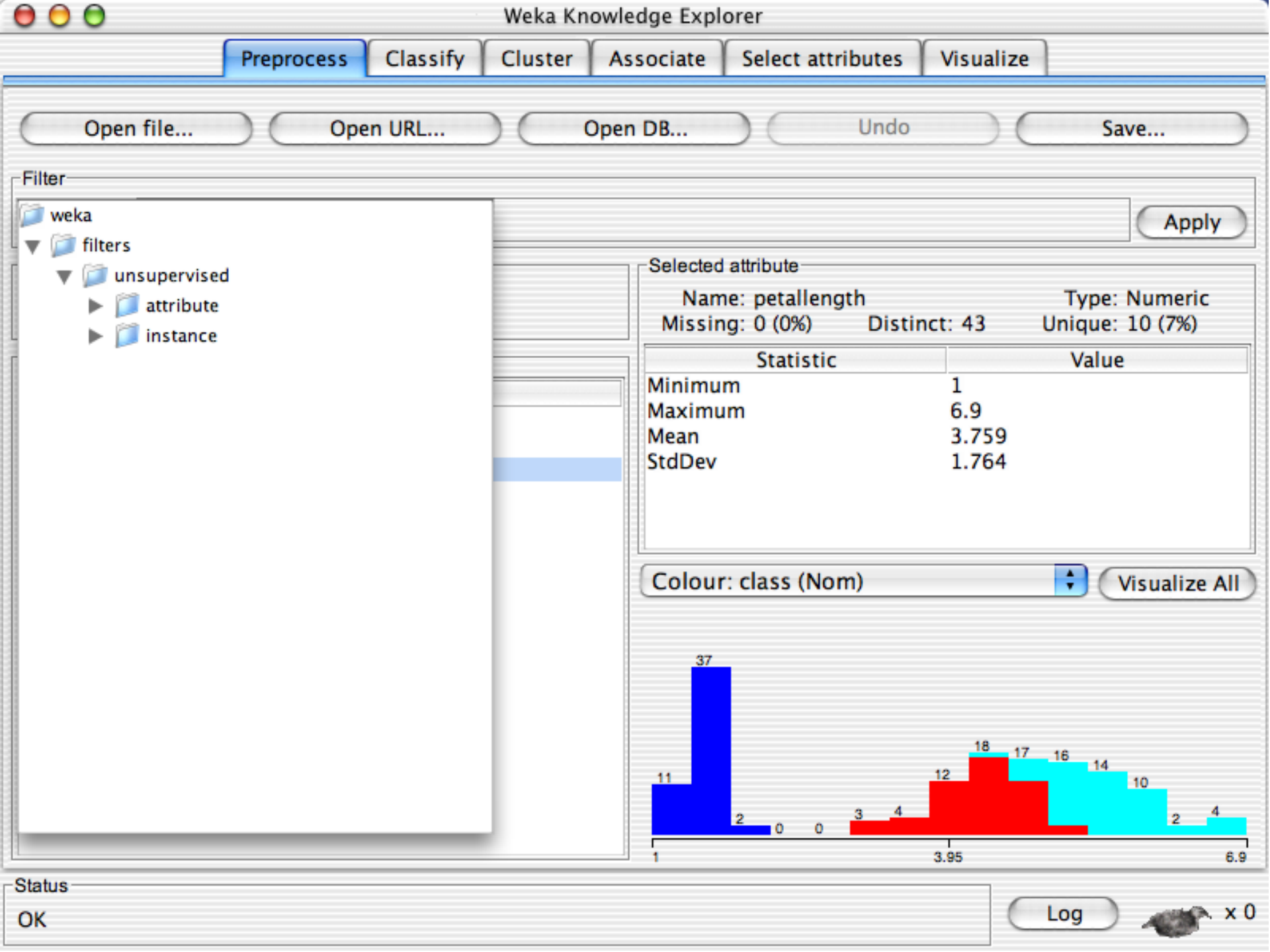


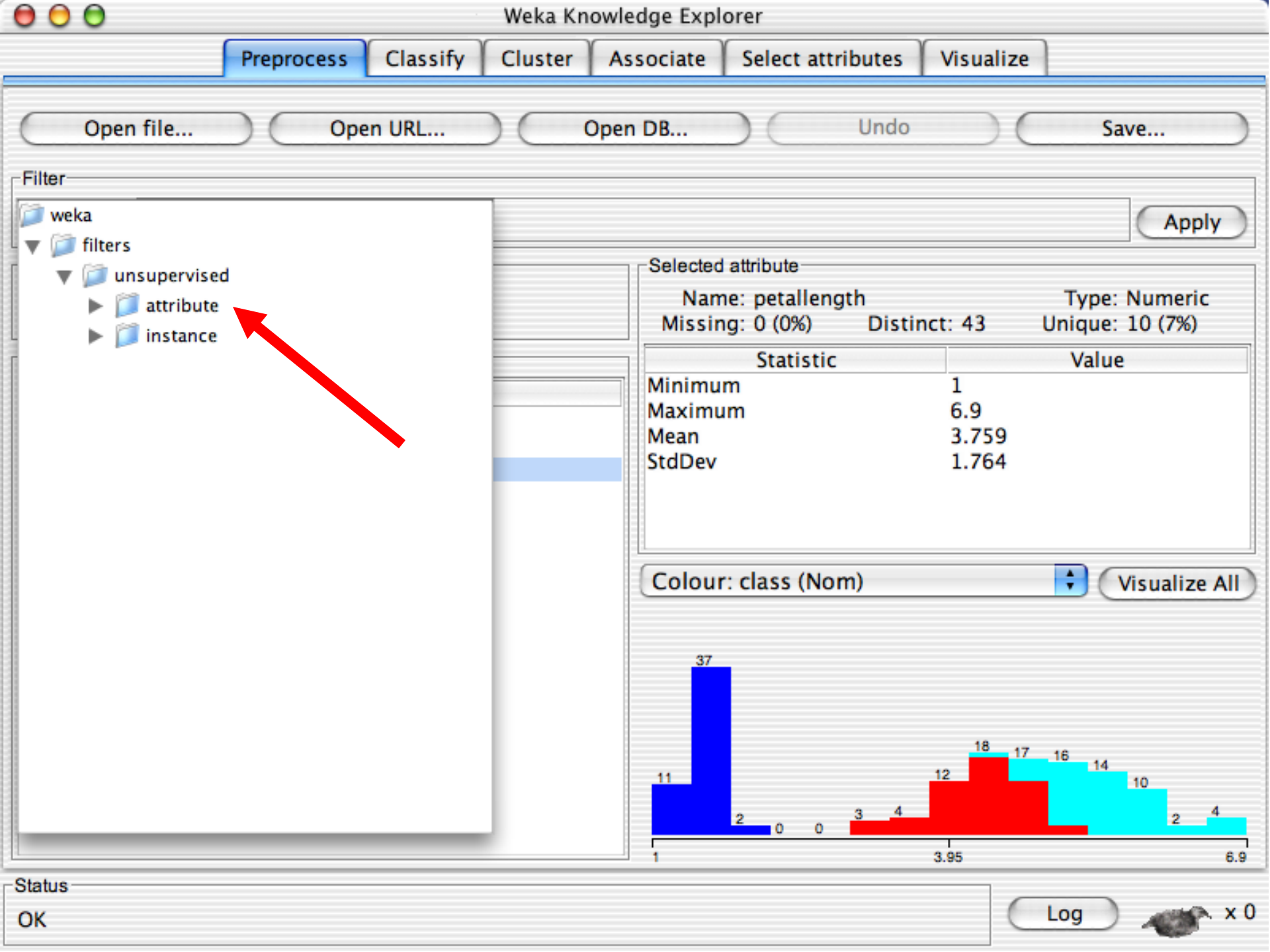

Status

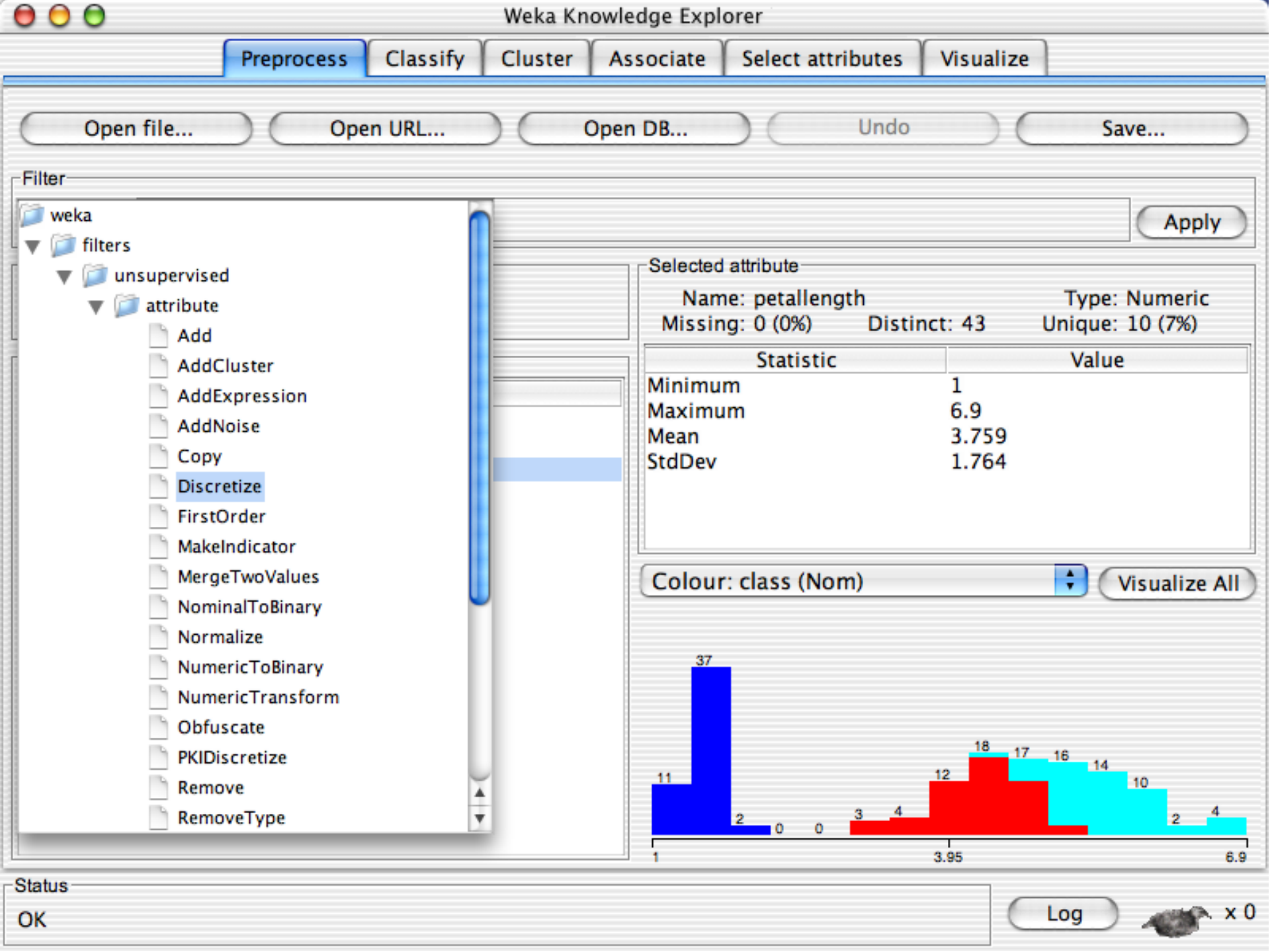
OK

Log | x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
Colour: class (Nom)
Visualize All
37
11
2
0
0
3
4
12
18
17
16
14
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
weka
filters
unsupervised
attribute
instance
Apply
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
Colour: class (Nom)
Visualize All
37
11
2
0
0
3
4
12
18
17
16
14
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Apply
weka
filters
unsupervised
attribute
instance
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
Colour: class (Nom)
Visualize All
37
11
2
0
0
3
4
12
18
17
16
14
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
weka
filters
unsupervised
attribute
Add
AddCluster
AddExpression
AddNoise
Copy
Discretize
FirstOrder
MakeIndicator
MergeTwoValues
NominalToBinary
Normalize
NumericToBinary
NumericTransform
Obfuscate
PKIDiscretize
Remove
RemoveType
Apply
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
Colour: class (Nom)
Visualize All
37
11
2
0
0
3
4
12
18
17
16
14
10
2
4
1
3.95
6.9
Status
OK
Log
x 0

Weka Knowledge Explorer

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

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

Filter

Choose **Discretize** -B 10 -R first-last | Apply

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

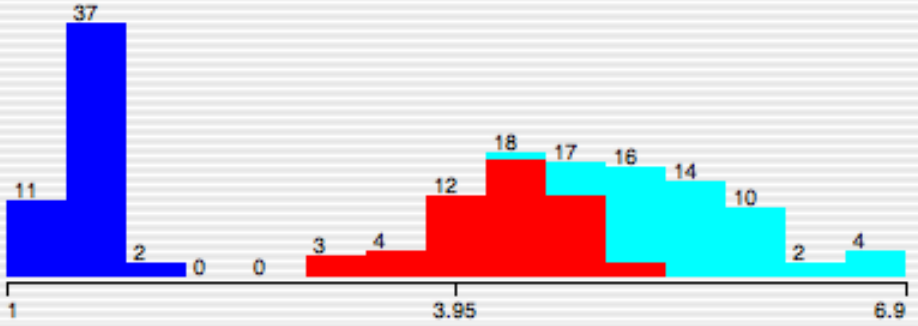
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

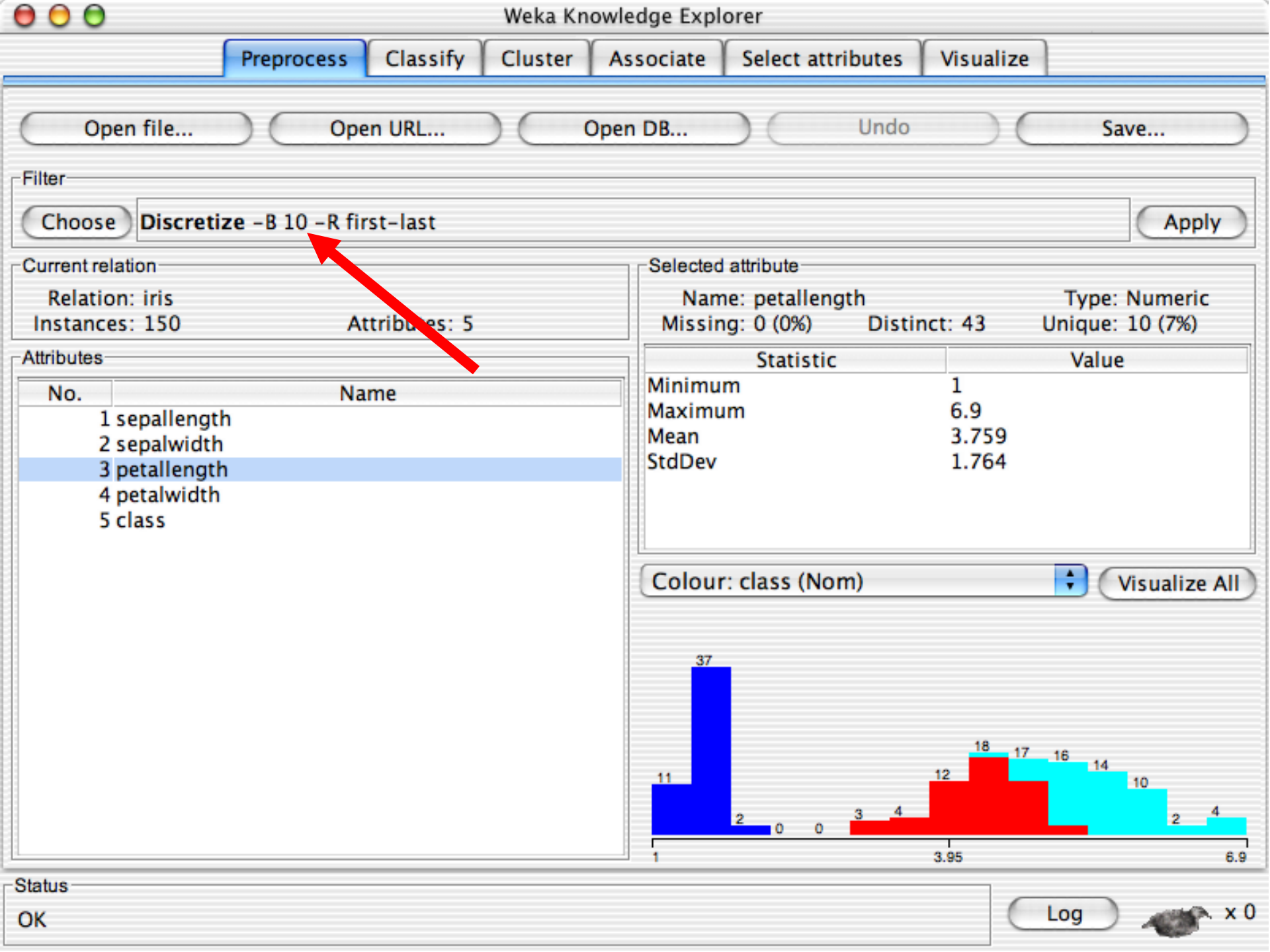
Colour: class (Nom) | Visualize All

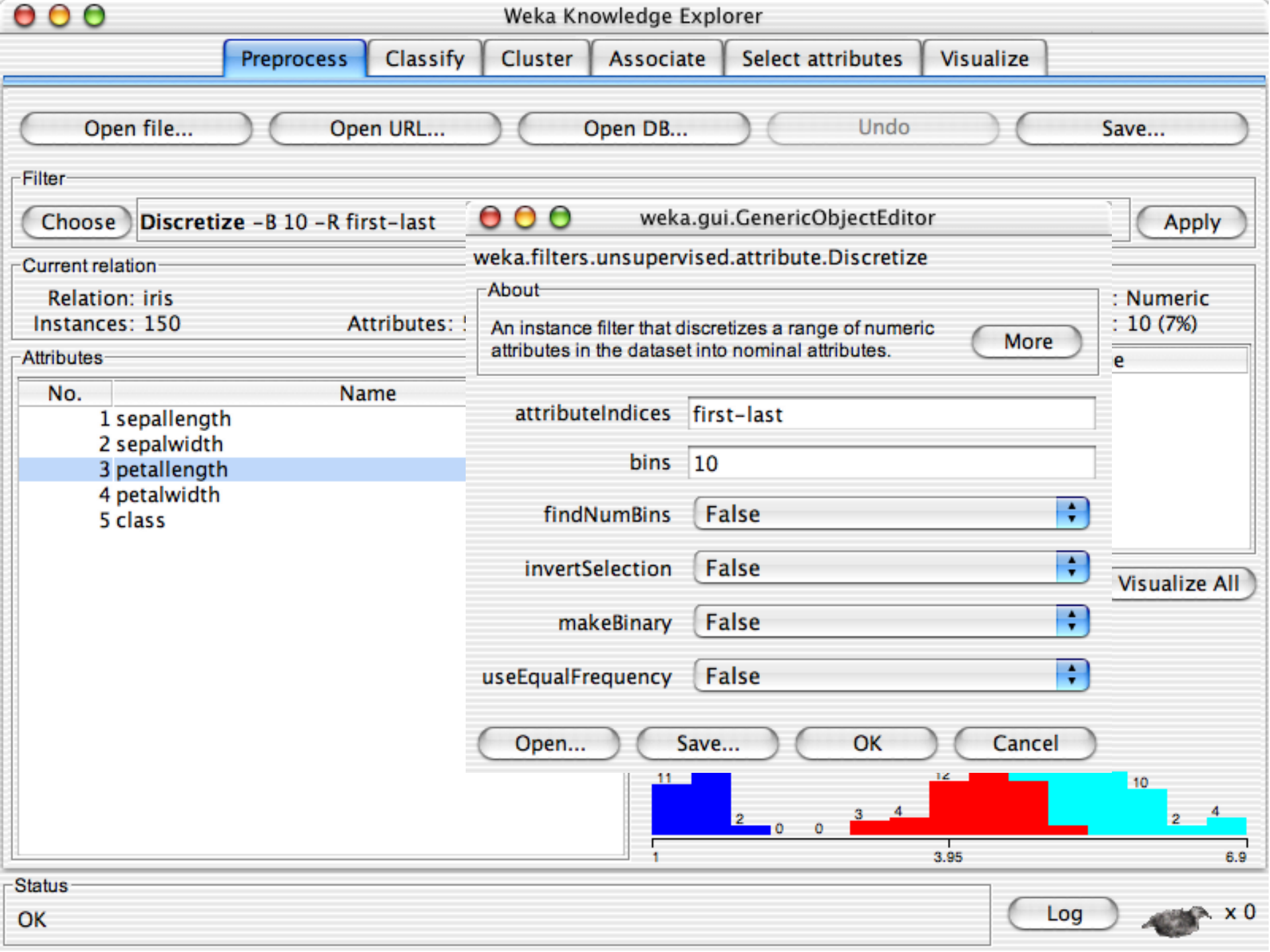


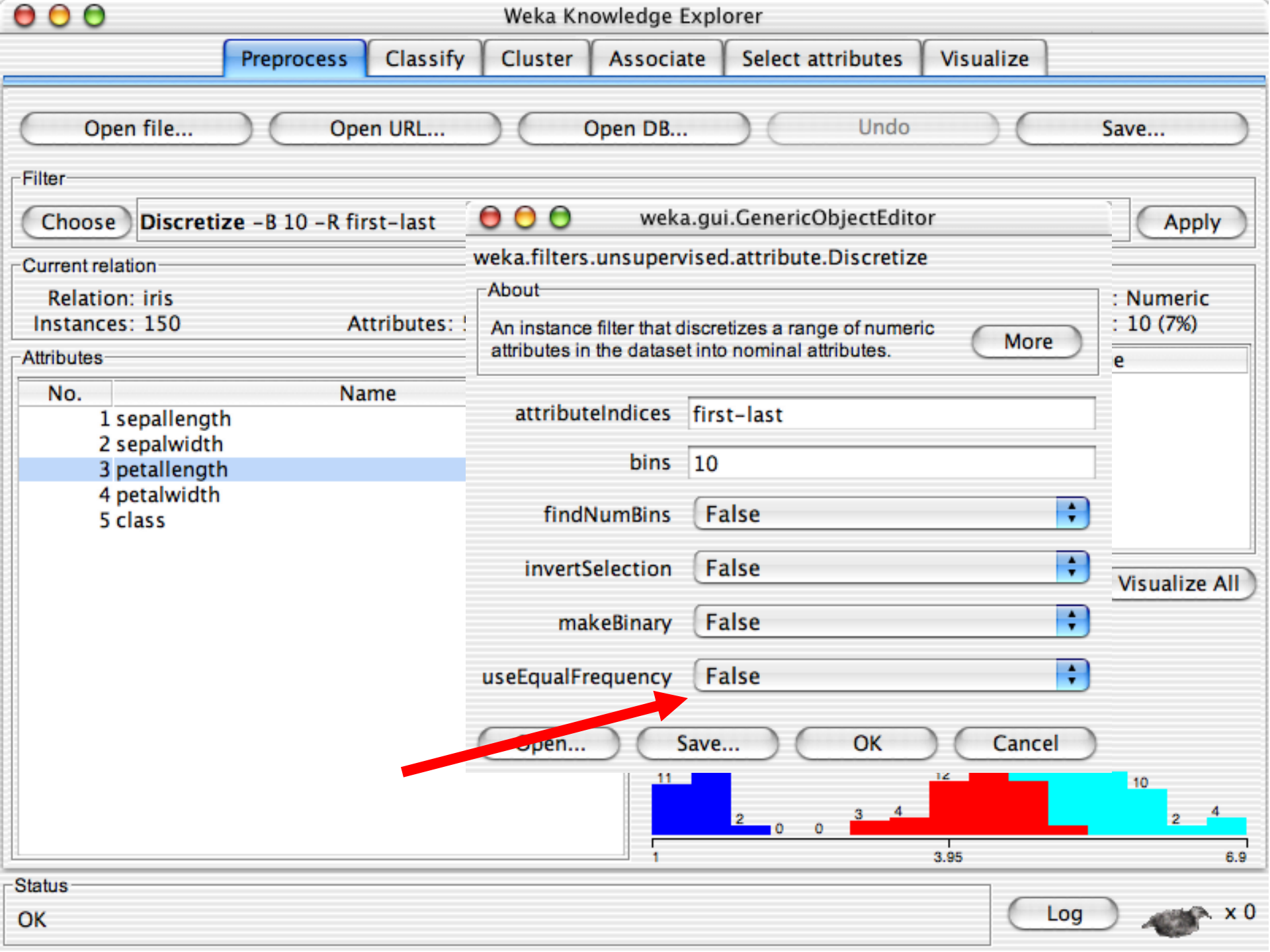

Status

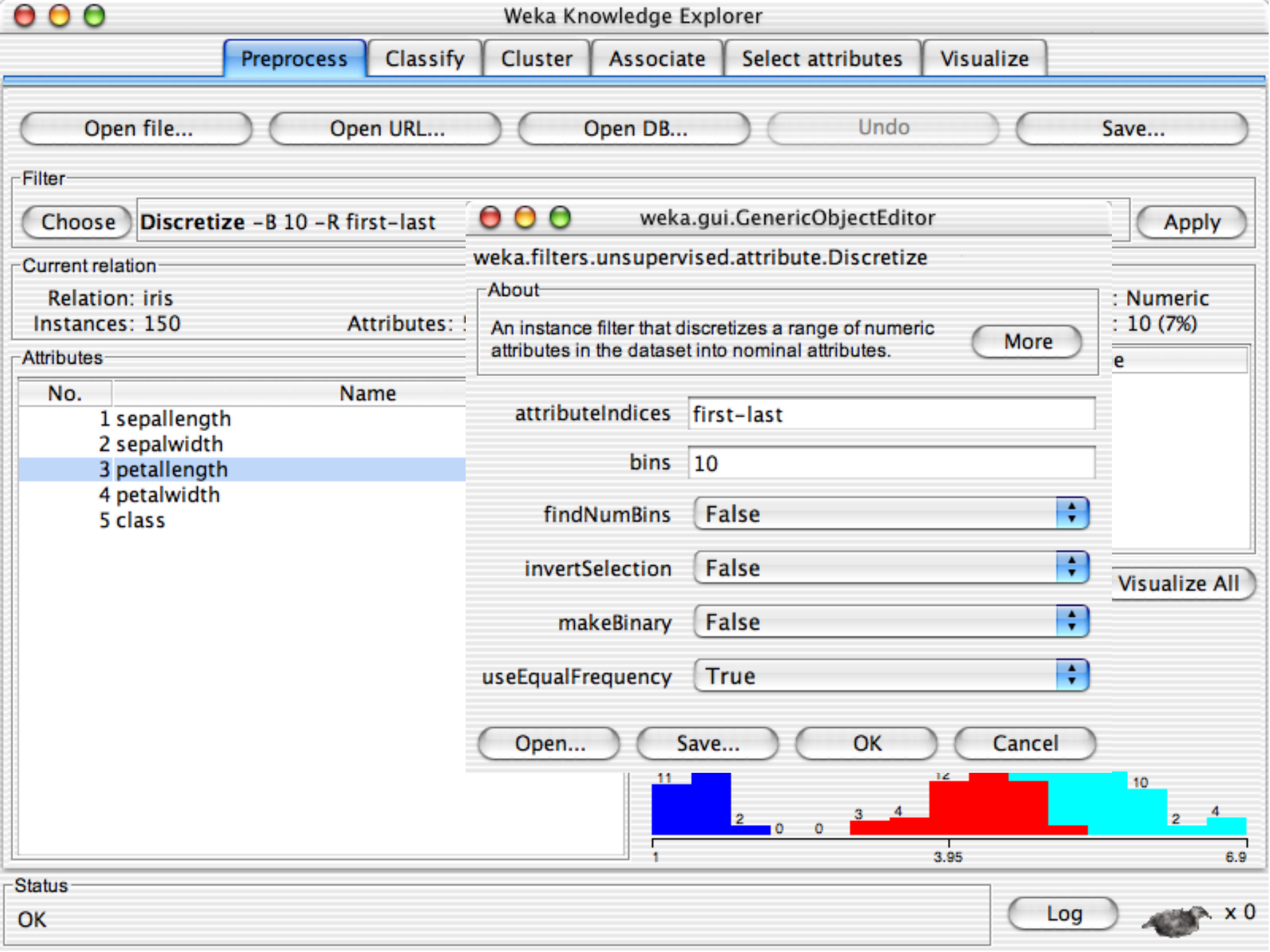
OK

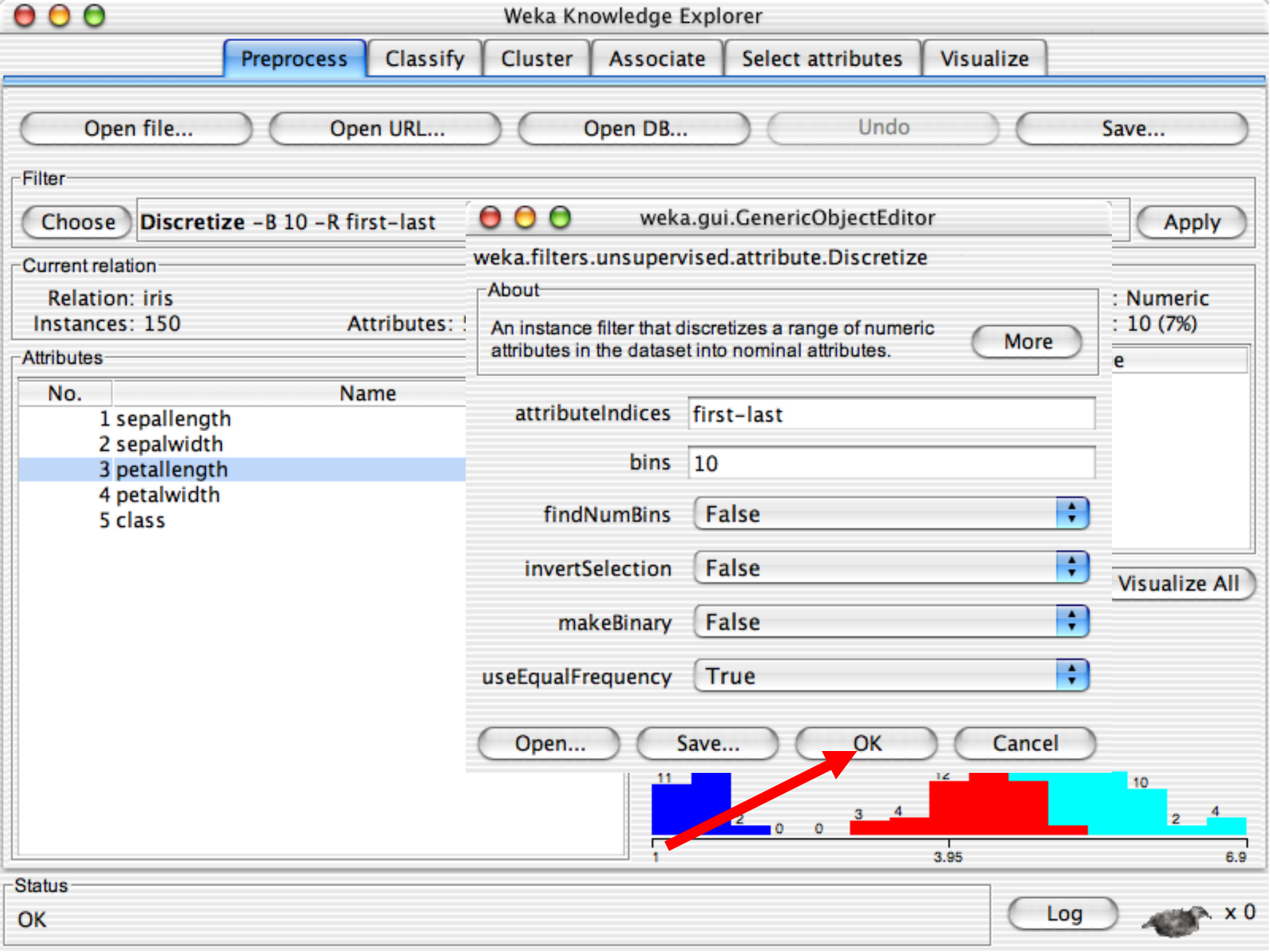
Log | x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -B 10 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
Colour: class (Nom)
Visualize All
37
11
2
0
0
3
4
12
18
17
16
14
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -B 10 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes:
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
attributeIndices
first-last
bins
10
findNumBins
False
invertSelection
False
makeBinary
False
useEqualFrequency
False
Open...
Save...
OK
Cancel
: Numeric
: 10 (7%)
e
Visualize All
11
2
0
0
3
4
12
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -B 10 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes:
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
attributeIndices
first-last
bins
10
findNumBins
False
invertSelection
False
makeBinary
False
useEqualFrequency
False
Open...
Save...
OK
Cancel
: Numeric
: 10 (7%)
Visualize All
11
2
0
0
3
4
12
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -B 10 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes:
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
attributeIndices
first-last
bins
10
findNumBins
False
invertSelection
False
makeBinary
False
useEqualFrequency
True
Open...
Save...
OK
Cancel
: Numeric
: 10 (7%)
e
Visualize All
11
2
0
0
3
4
12
10
2
4
1
3.95
6.9
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -B 10 -R first-last
Apply
Current relation
Relation: iris
Instances: 150
Attributes:
Attributes
No.
Name
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
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
attributeIndices
first-last
bins
10
findNumBins
False
invertSelection
False
makeBinary
False
useEqualFrequency
True
Open...
Save...
OK
Cancel
: Numeric
: 10 (7%)
e
Visualize All
11
2
0
0
3
4
12
10
2
4
1
3.95
6.9
Status
OK
Log
x 0

Weka Knowledge Explorer

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

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

Filter

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

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

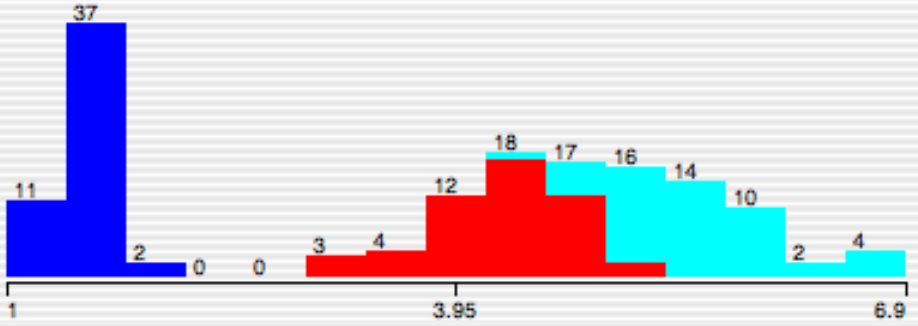
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

Colour: class (Nom) | Visualize All




Status

OK

Log x 0

Weka Knowledge Explorer

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

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

Filter

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

Current relation

Relation: iris
Instances: 150 Attributes: 5

Attributes

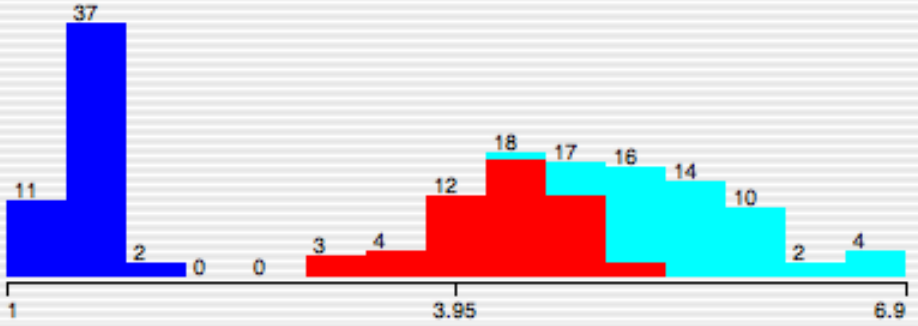
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

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 |

Colour: class (Nom) | Visualize All




Status

OK

Log x 0

Weka Knowledge Explorer

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

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

Filter

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

Current relation

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

Attributes

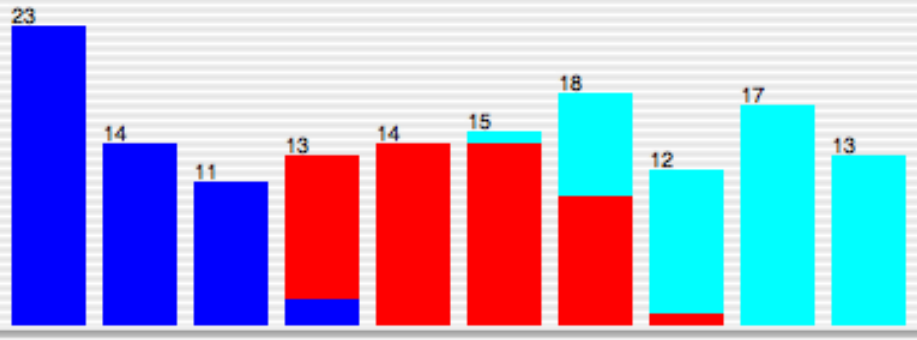
| No. | Name |
|---|---|
| 1 | sepallength |
| 2 | sepalwidth |
| 3 | petallength |
| 4 | petalwidth |
| 5 | class |

Selected attribute

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

| Label | Count |
|---|---|
| '(-inf-1.45]' | 23 |
| '(1.45-1.55]' | 14 |
| '(1.55-1.8]' | 11 |
| '(1.8-3.95]' | 13 |
| '(3.95-4.35]' | 14 |
| '(4.35-4.65]' | 15 |
| '(4.65-5.05]' | 18 |

Colour: class (Nom) | Visualize All



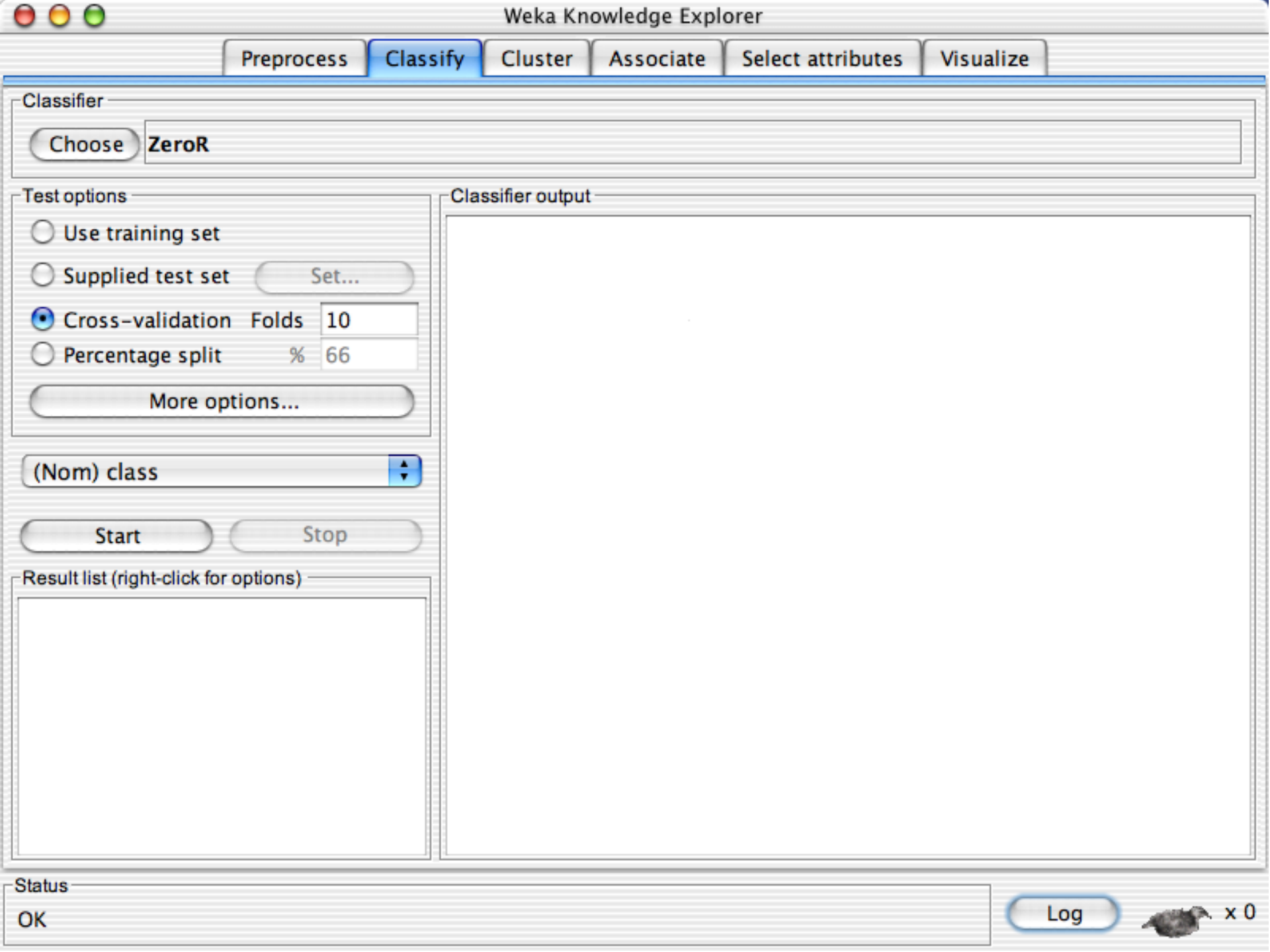

Status

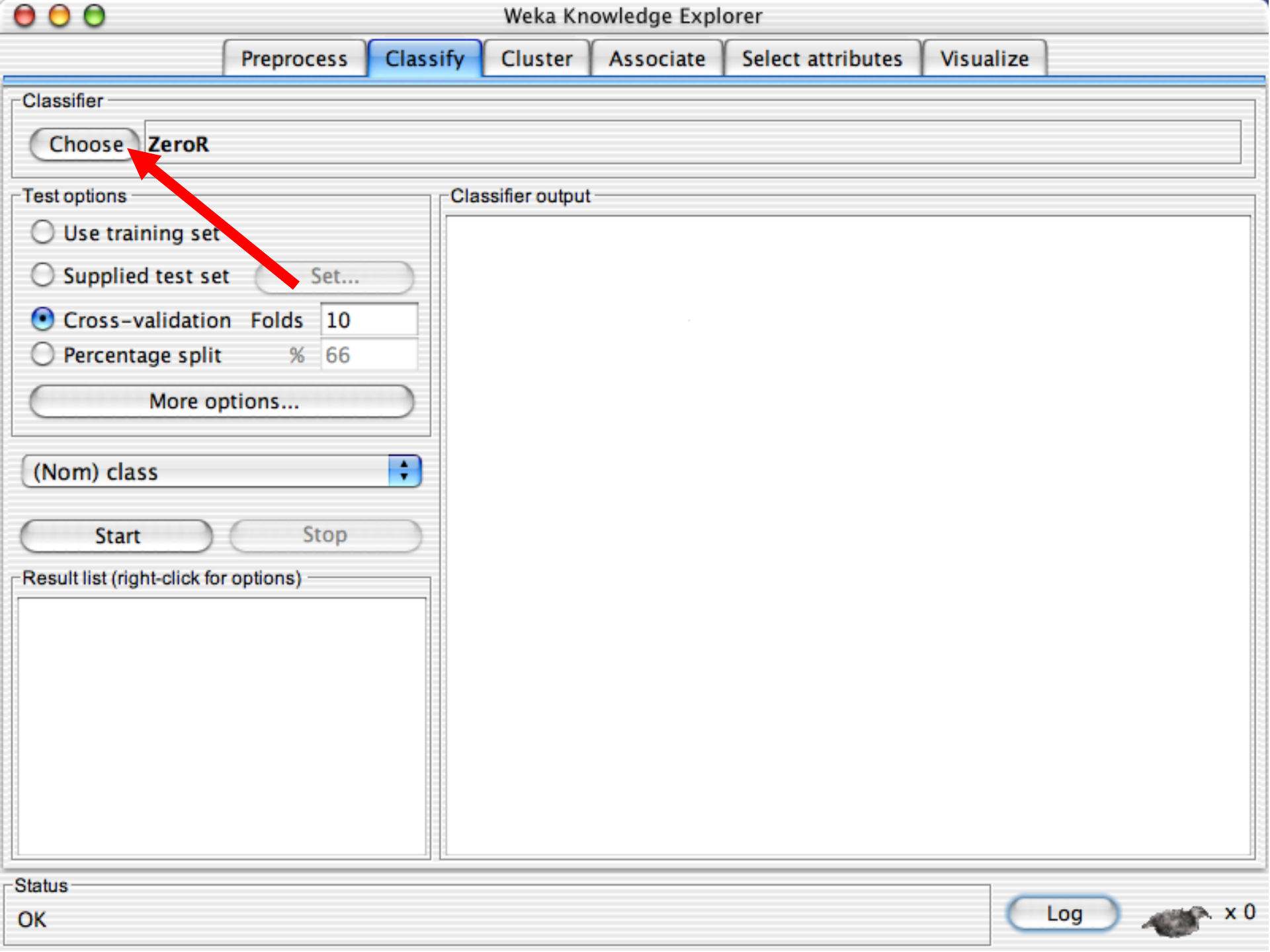
OK

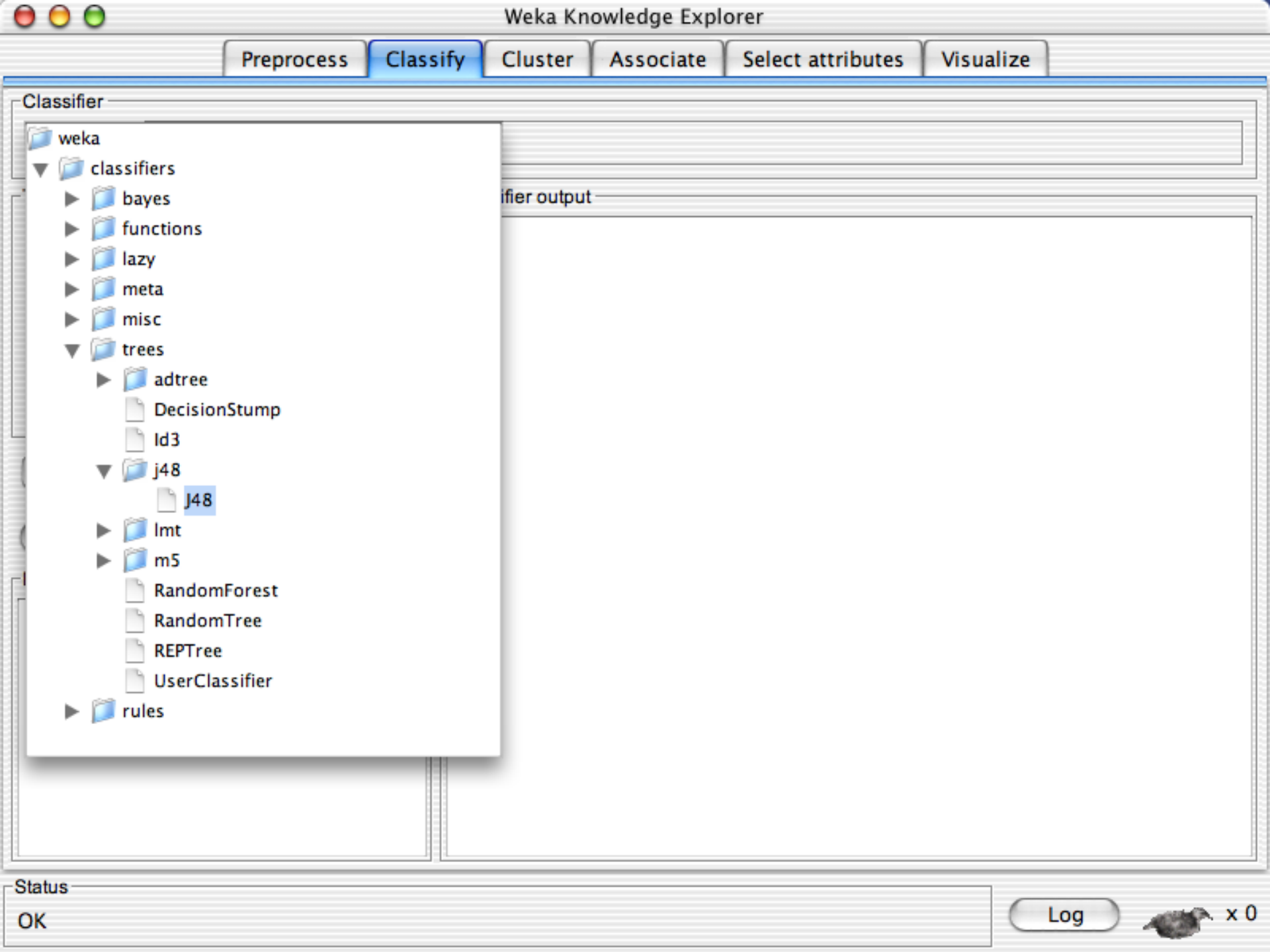
Log | x 0

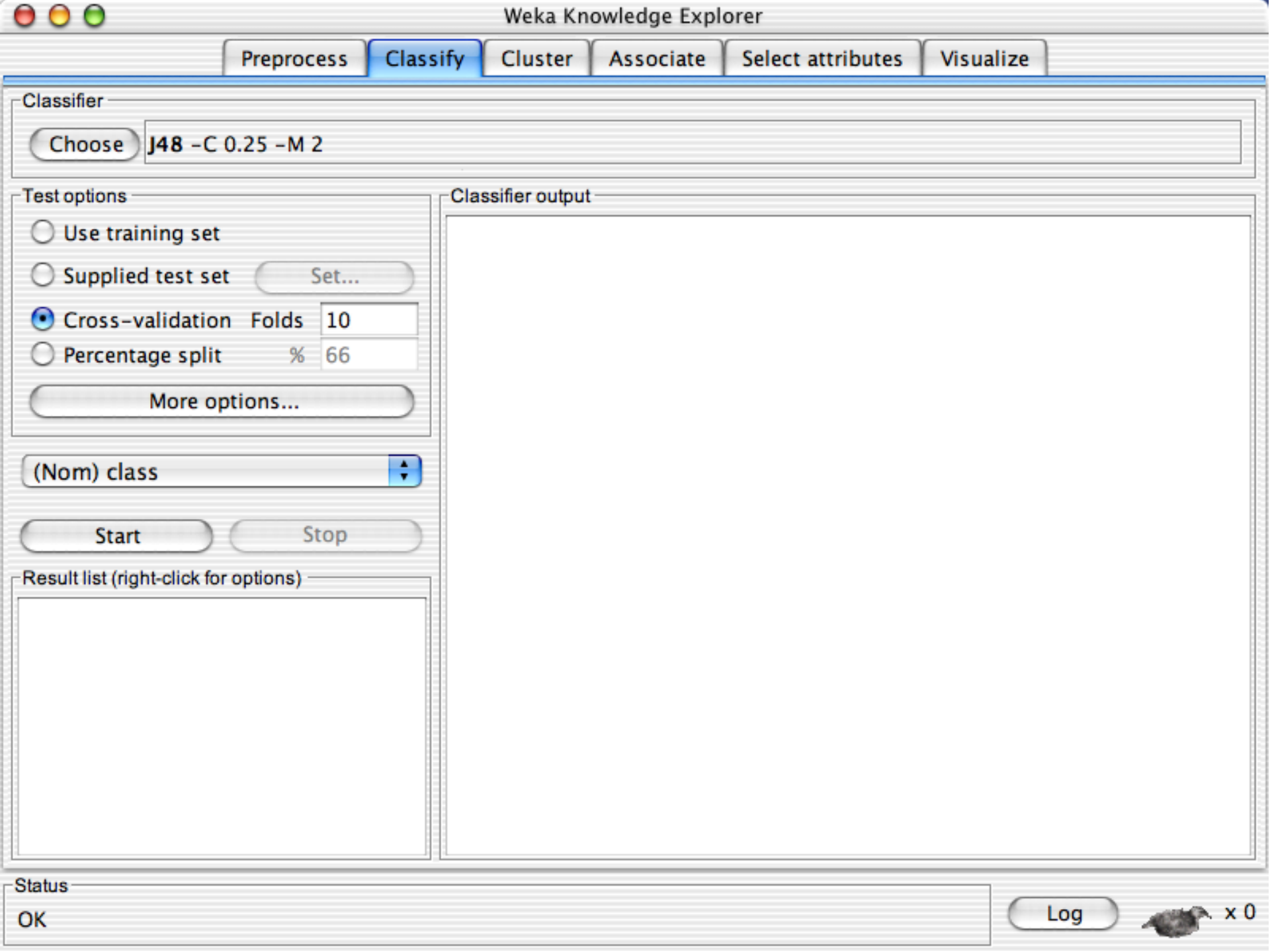
# Explorer: building “classifiers”

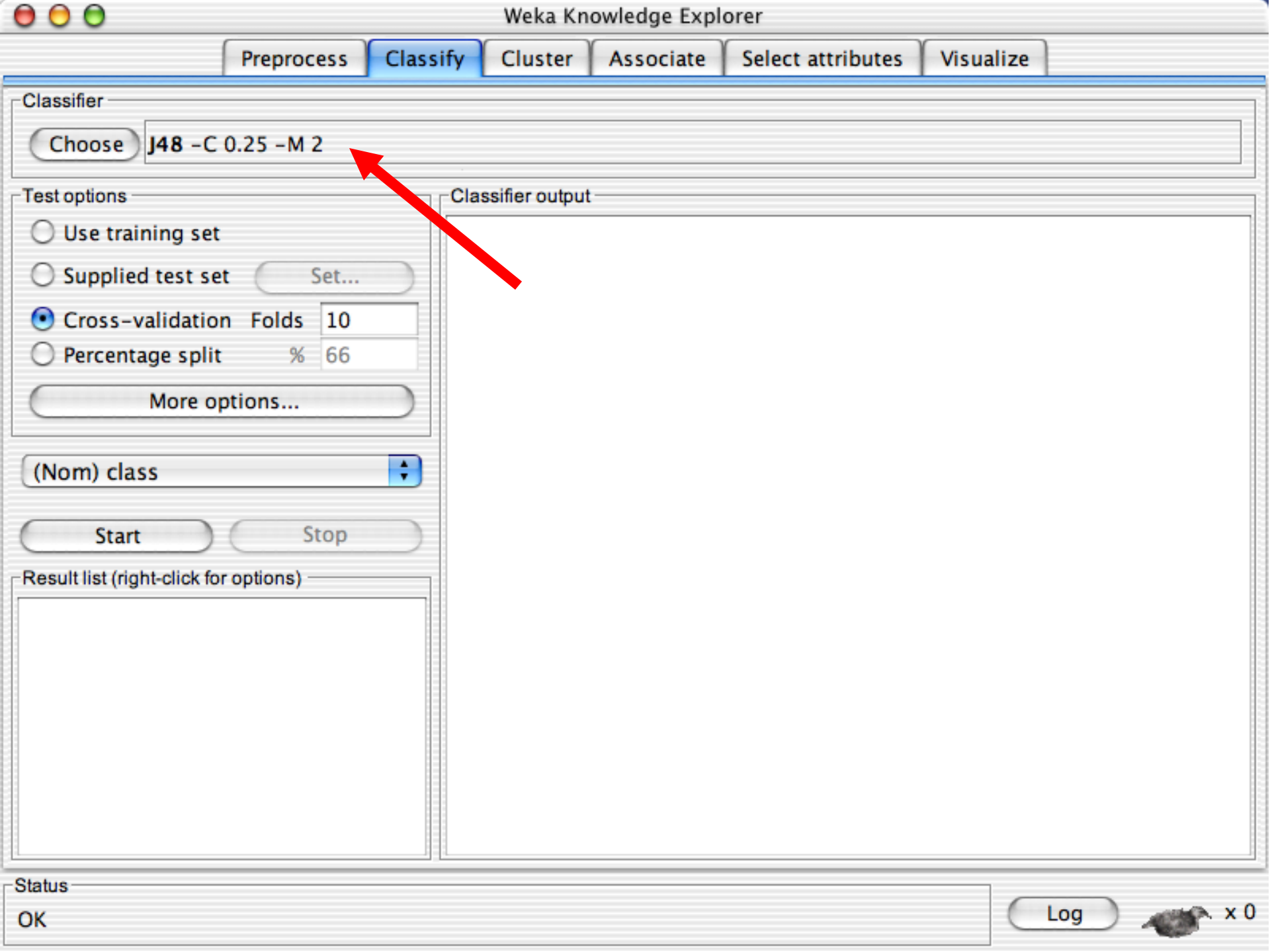
- Classifiers in WEKA are models for predicting nominal or numeric quantities
- Implemented learning schemes include:
  - Decision trees and lists, instance-based classifiers, support vector machines, multi-layer perceptrons, logistic regression, Bayes’ nets, …
- “Meta”-classifiers include:
  - Bagging,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

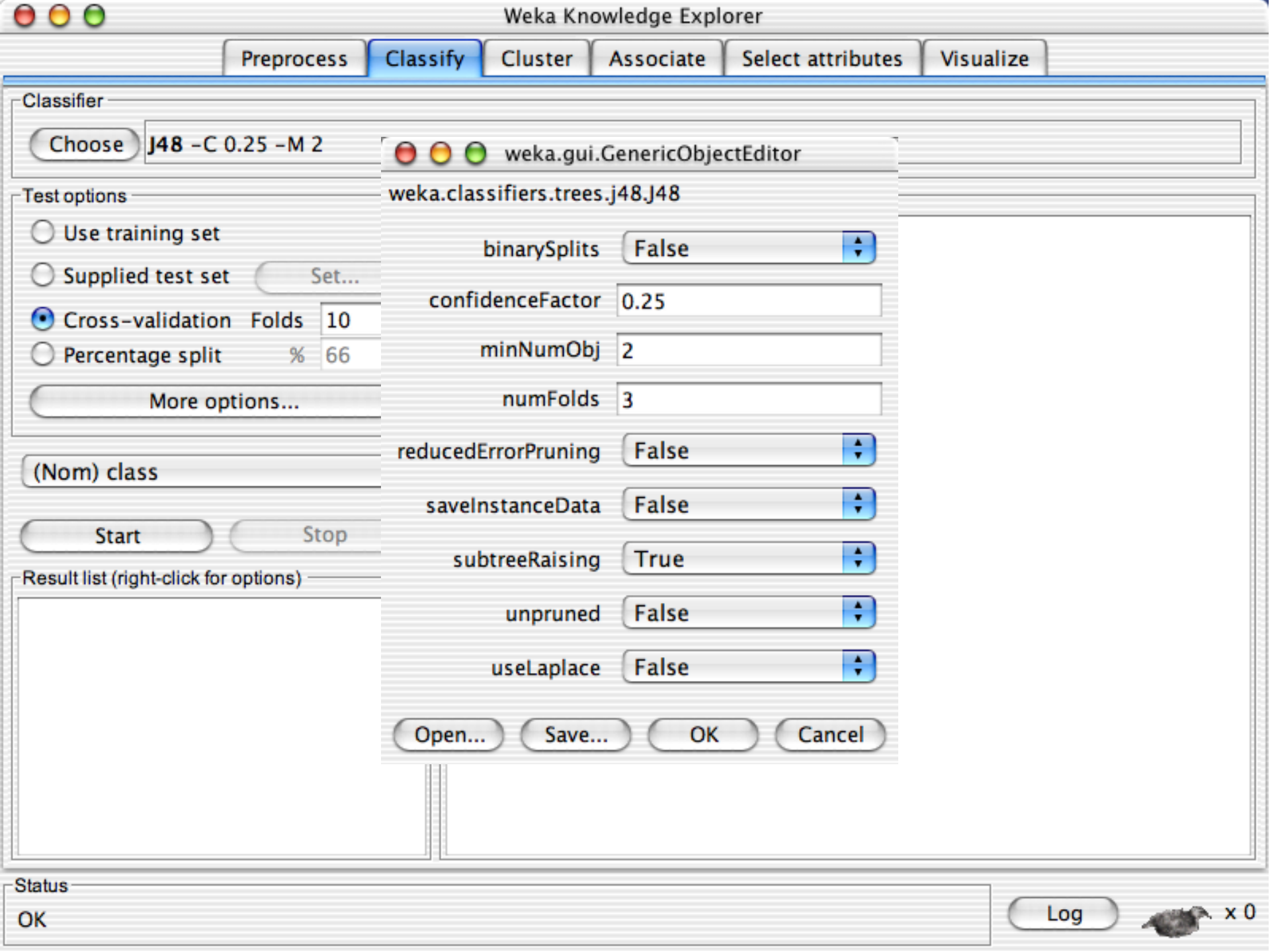

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for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

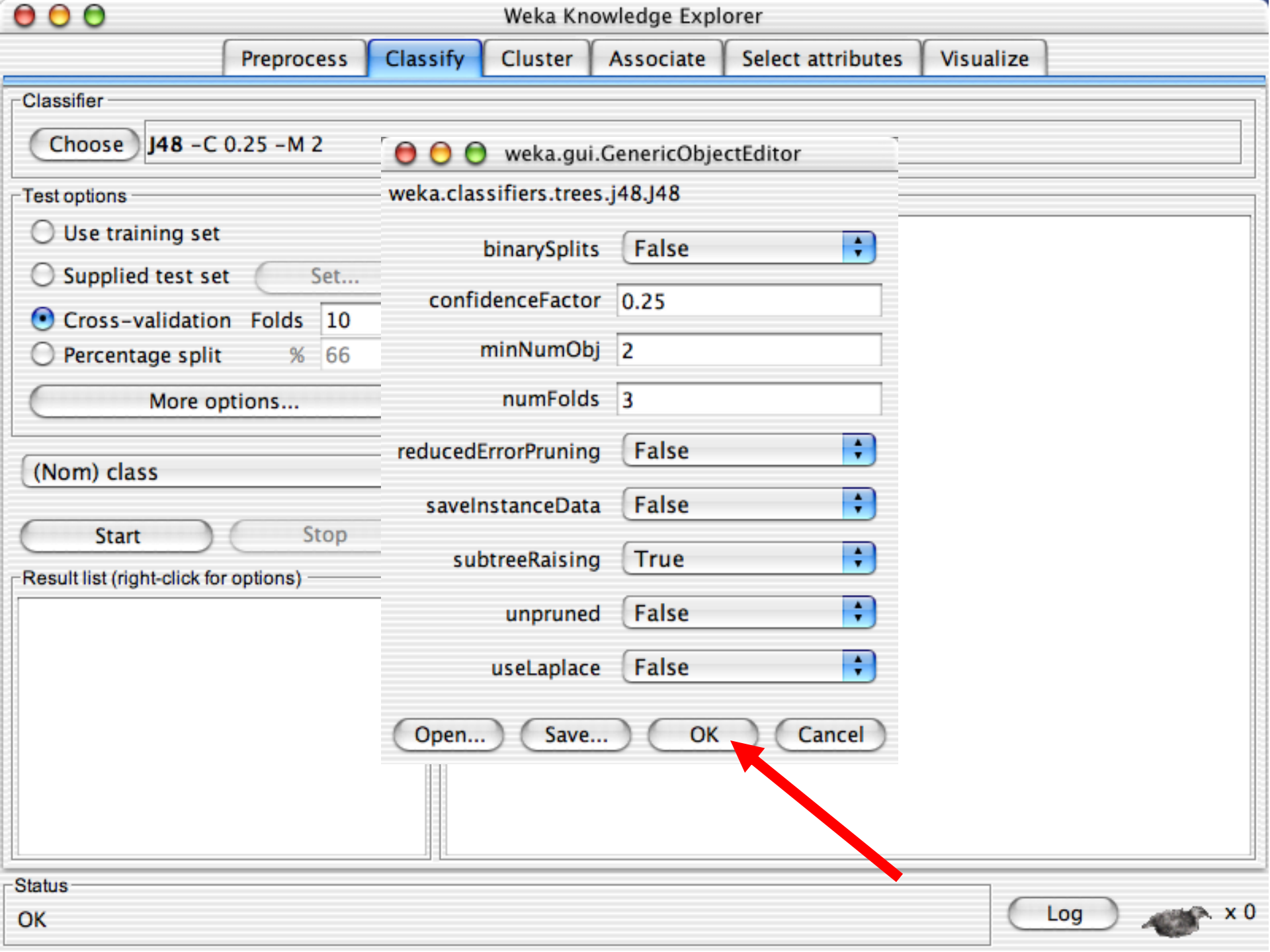

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
ZeroR
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click for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

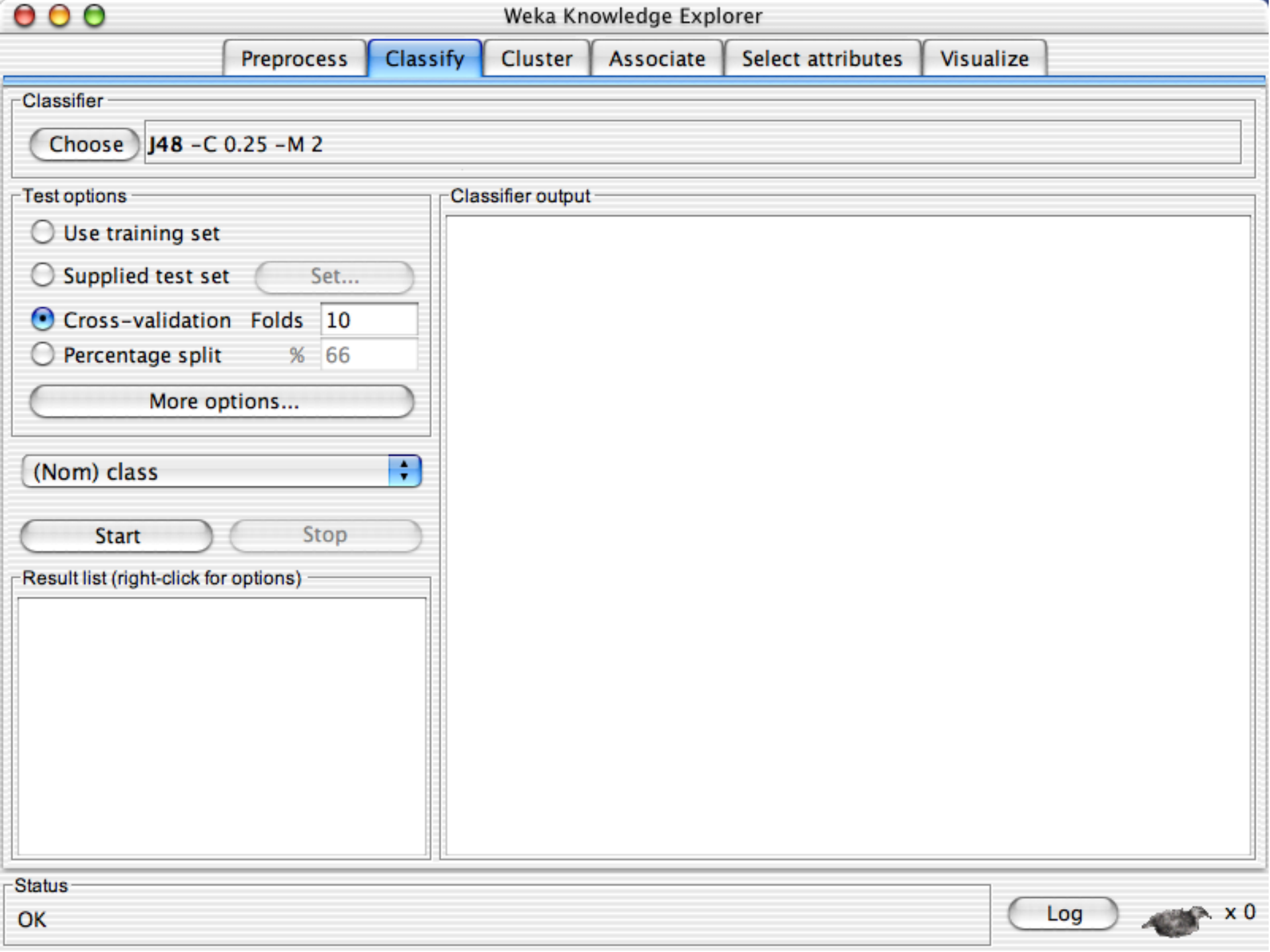

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
misc
trees
adtree
DecisionStump
Id3
j48
J48
lmt
m5
RandomForest
RandomTree
REPTree
UserClassifier
rules
ifier output
Status
OK
Log
x 0

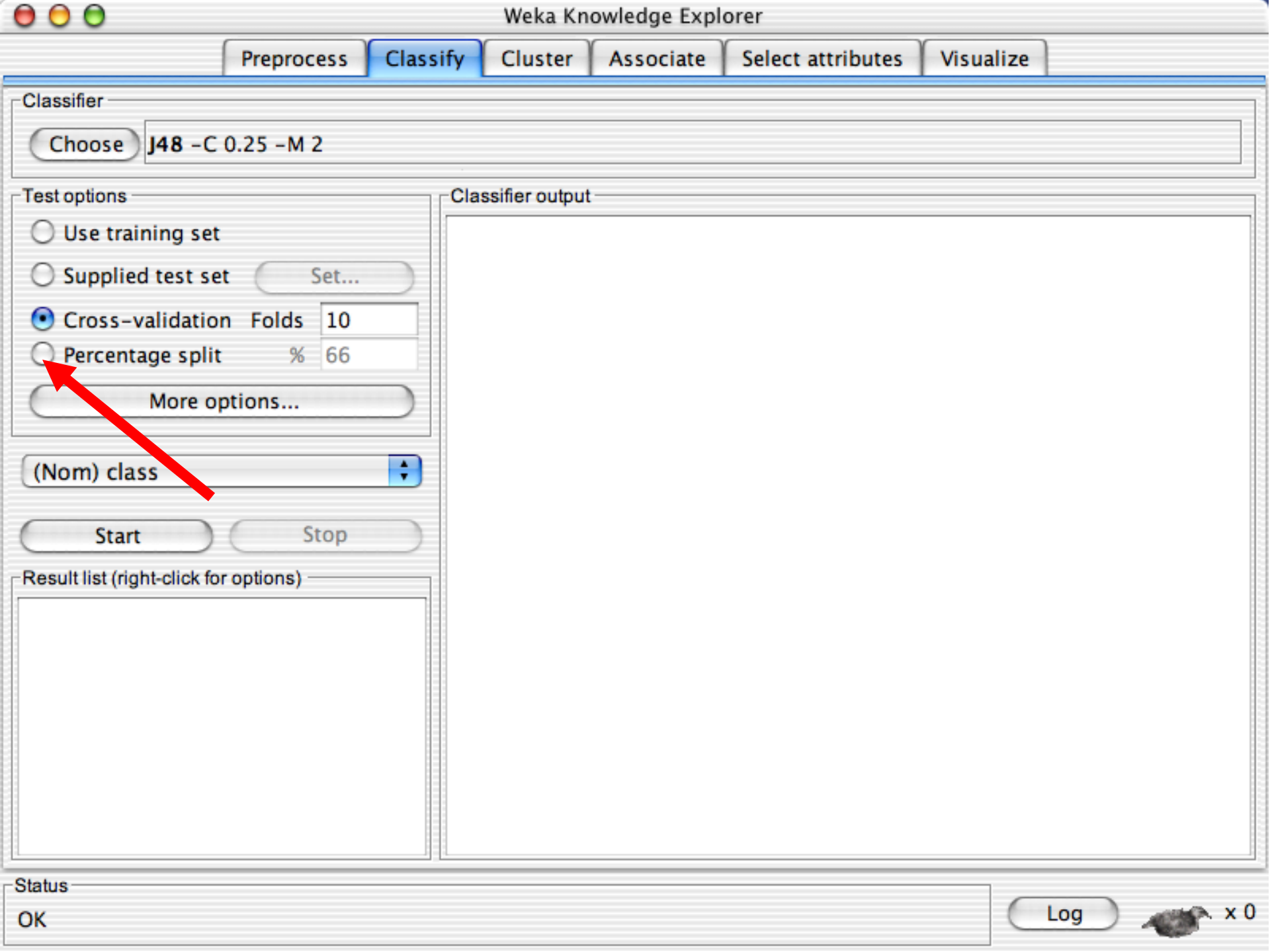

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click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

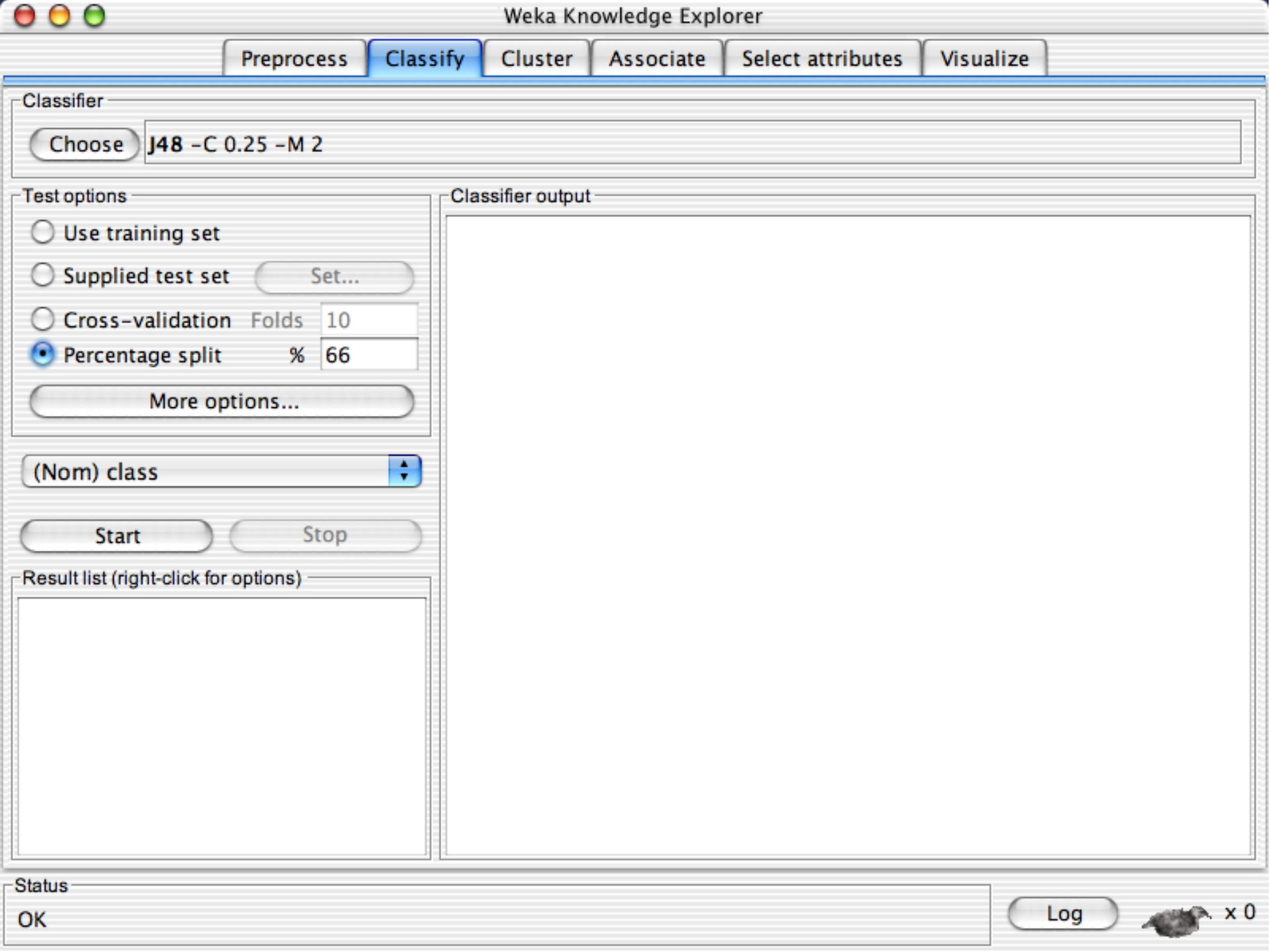

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click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

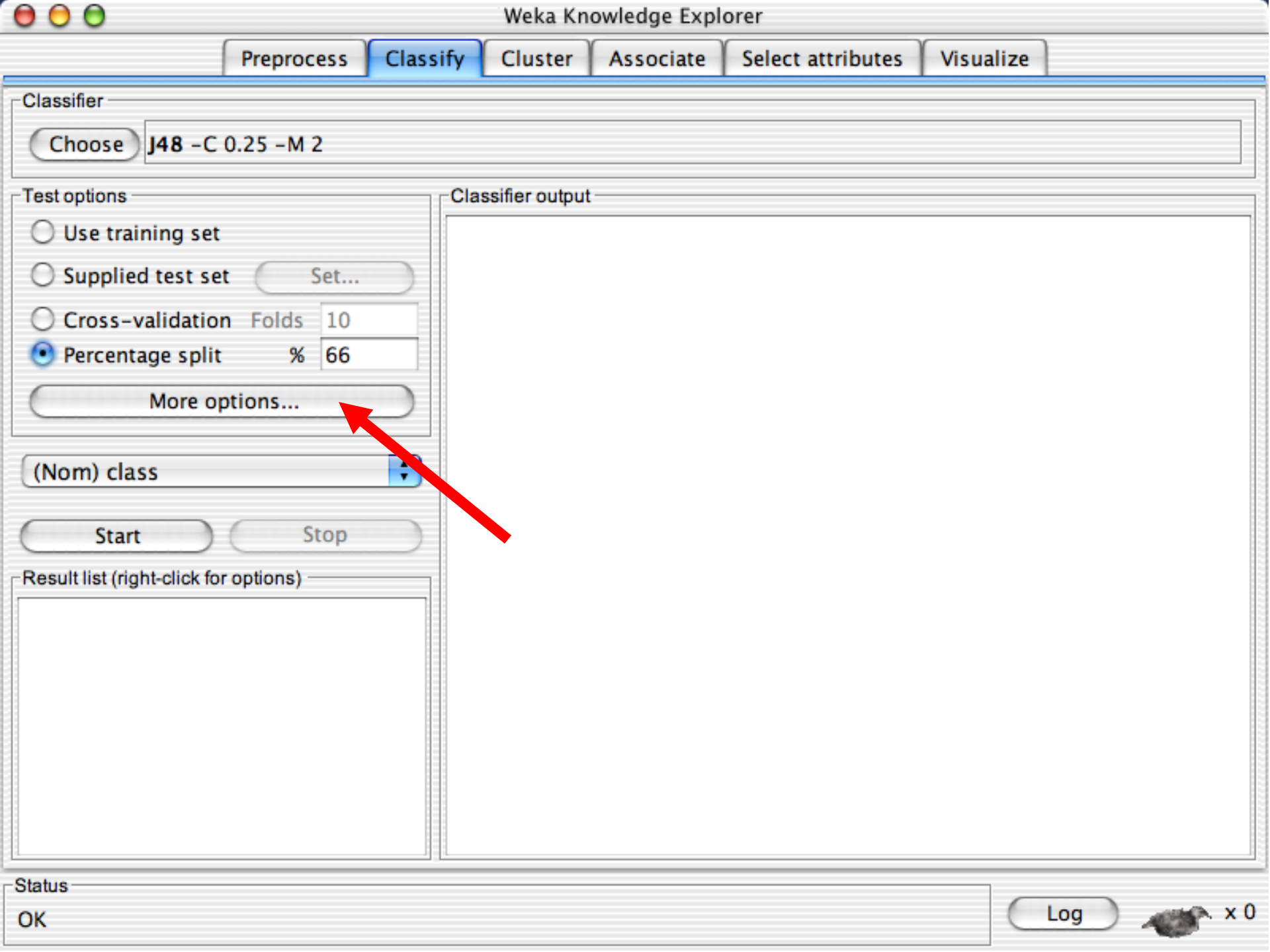

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
weka.gui.GenericObjectEditor
weka.classifiers.trees.j48.J48
binarySplits
False
confidenceFactor
0.25
minNumObj
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
subtreeRaising
True
unpruned
False
useLaplace
False
Open...
Save...
OK
Cancel
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)
Status
OK
Log
x 0

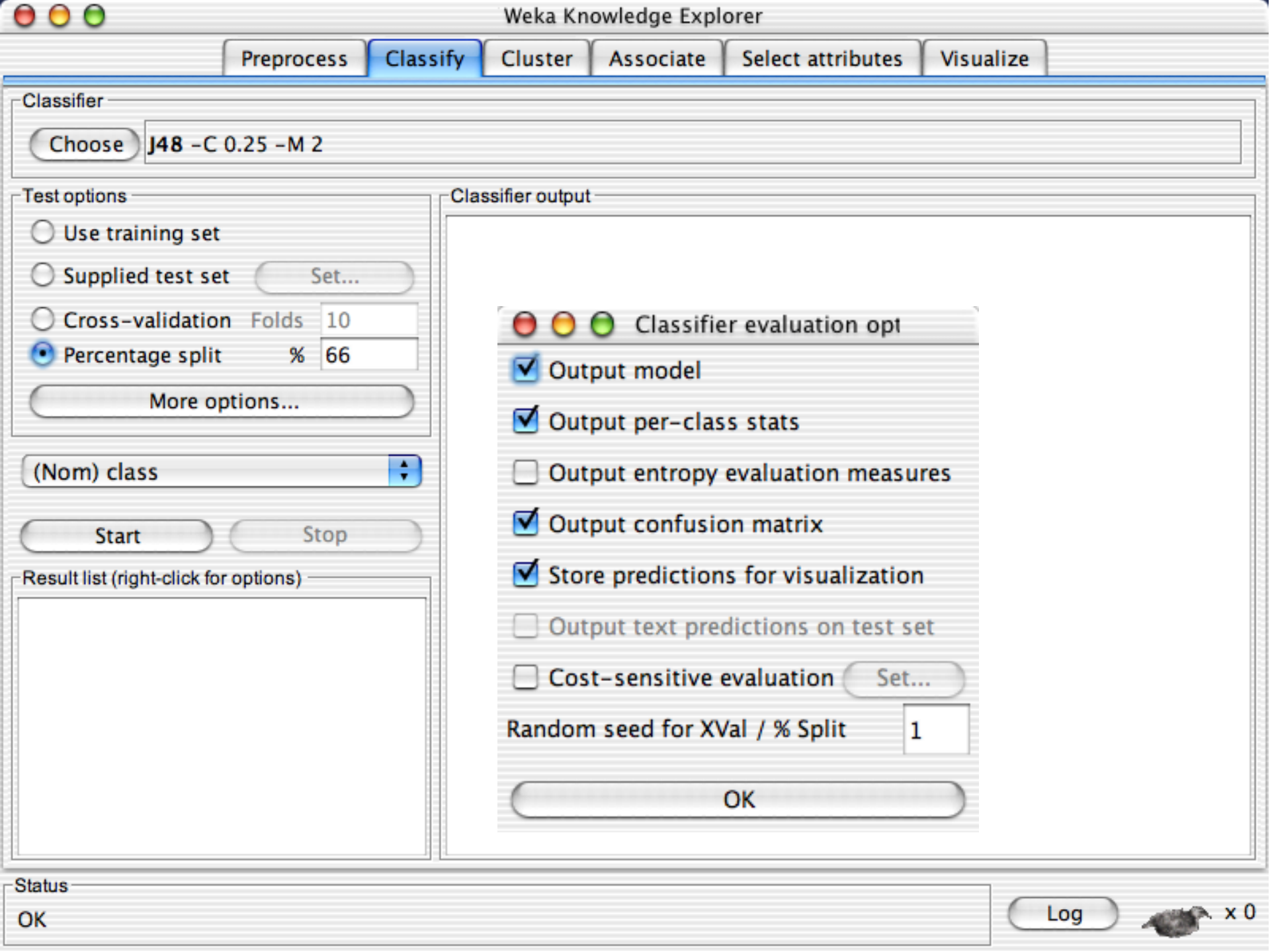

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
weka.gui.GenericObjectEditor
weka.classifiers.trees.j48.J48
binarySplits
False
confidenceFactor
0.25
minNumObj
2
numFolds
3
reducedErrorPruning
False
saveInstanceData
False
subtreeRaising
True
unpruned
False
useLaplace
False
Open...
Save...
OK
Cancel
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)
Status
OK
Log
x 0

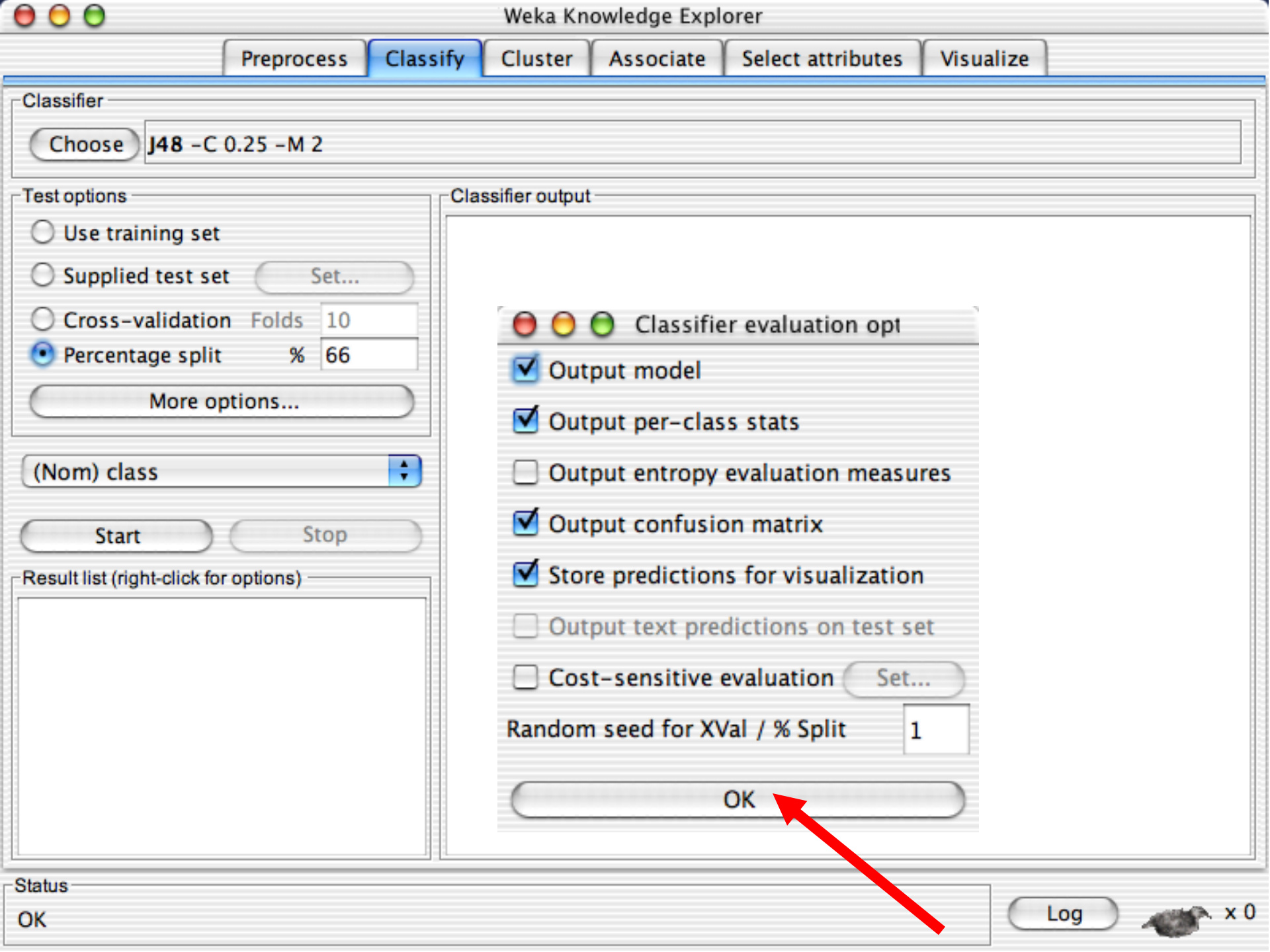

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click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

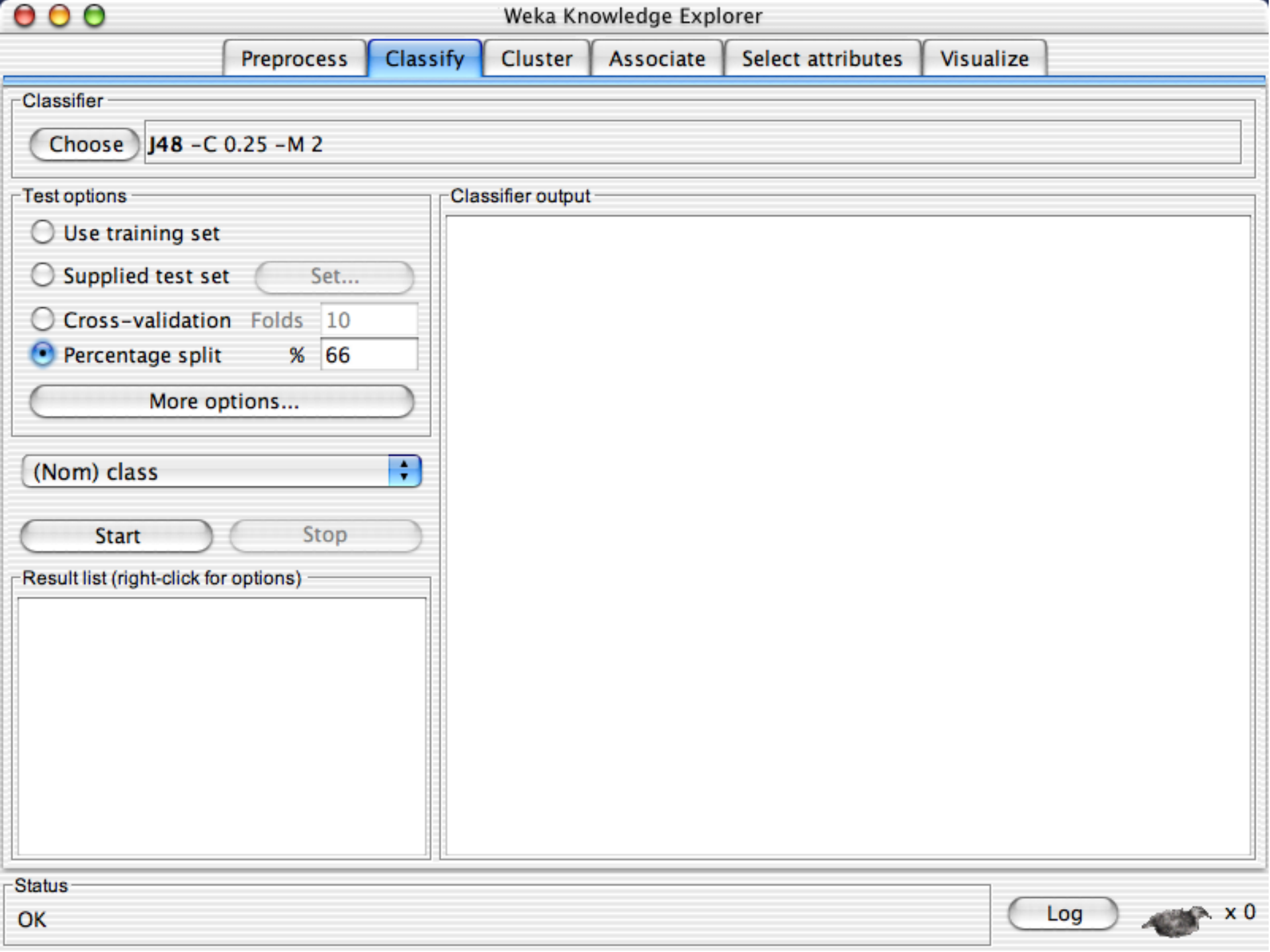

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click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

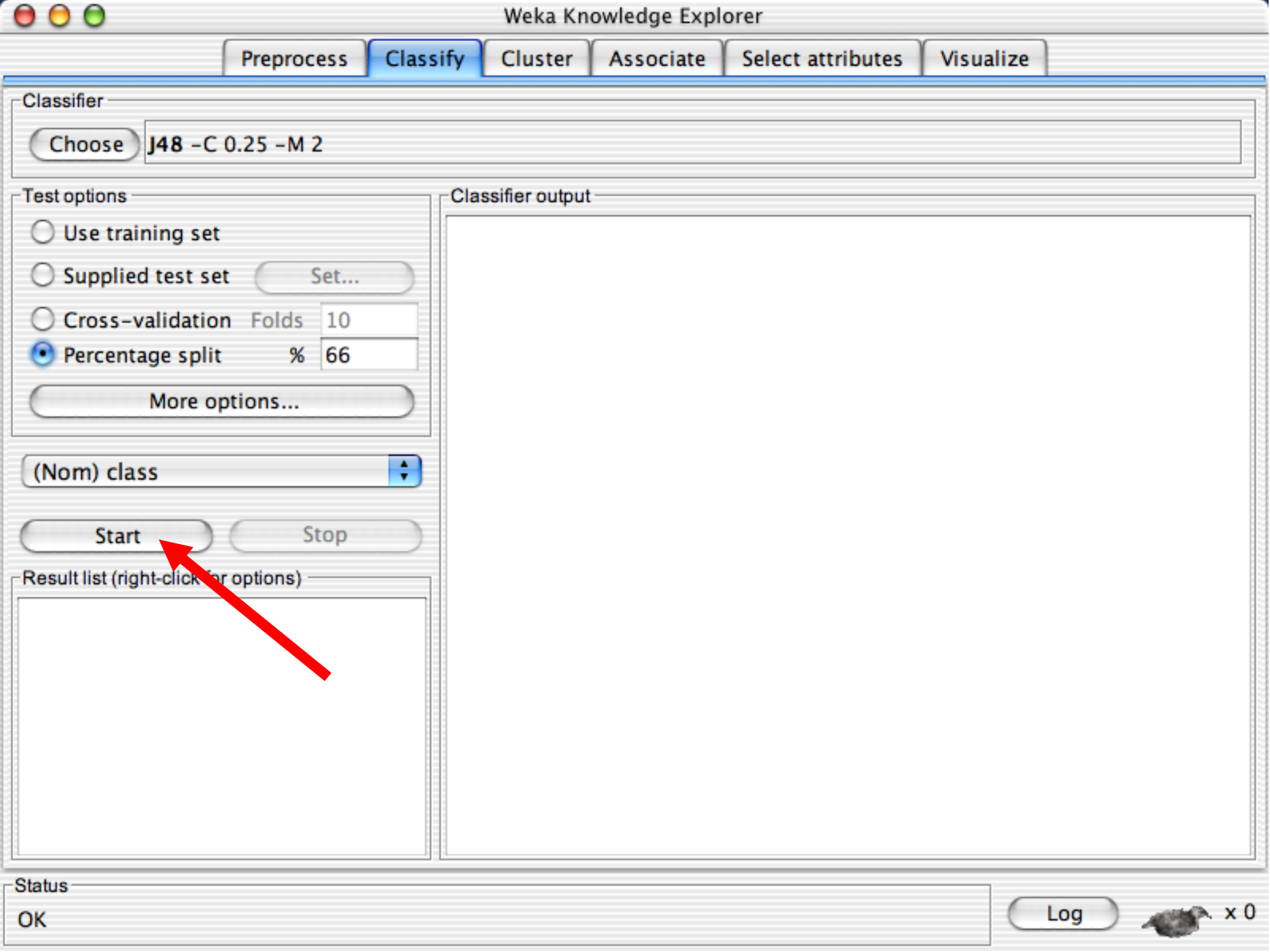

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
(Nom) class
Start
Stop
Result list (right-click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


Weka Knowledge 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 Knowledge 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)
Classifier evaluation opt
Output model
Output per-class stats
Output entropy evaluation measures
Output confusion matrix
Store predictions for visualization
Output text predictions on test set
Cost-sensitive evaluation
Set...
Random seed for XVal / % Split
1
OK
Status
OK
Log
x 0


Weka Knowledge 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)
Classifier evaluation opt
Output model
Output per-class stats
Output entropy evaluation measures
Output confusion matrix
Store predictions for visualization
Output text predictions on test set
Cost-sensitive evaluation
Set...
Random seed for XVal / % Split
1
OK
Status
OK
Log
x 0


Weka Knowledge 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 Knowledge 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 Knowledge 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)

11:49:05 - trees.j48.J48

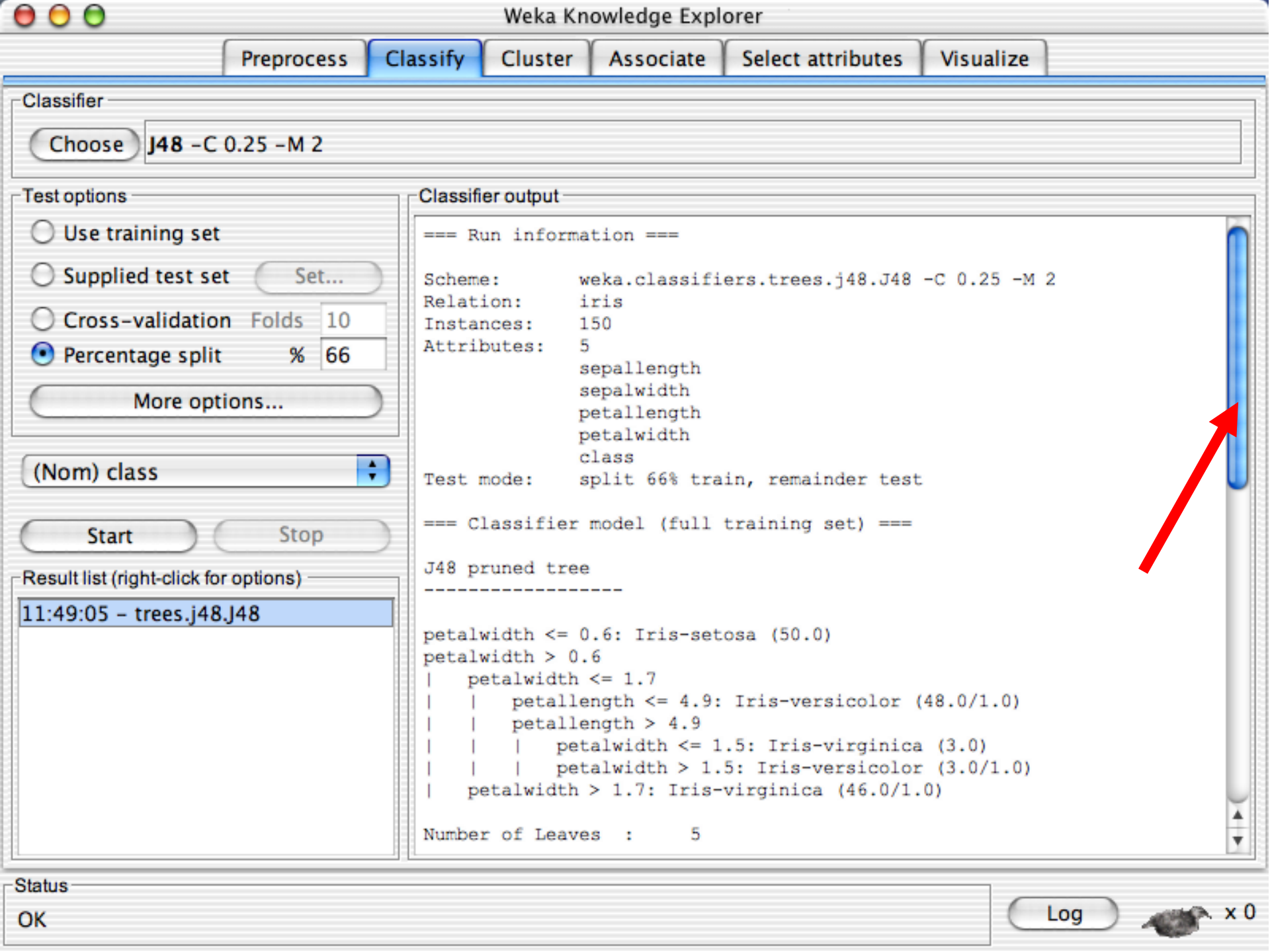
Classifier output

```
=== Run information ===

Scheme:       weka.classifiers.trees.j48.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
```

Status

OK

Log x 0


Weka Knowledge 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)
11:49:05 - trees.j48.J48
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.j48.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
Status
OK
Log
x 0

Weka Knowledge 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)

11:49:05 - trees.j48.J48

Classifier output

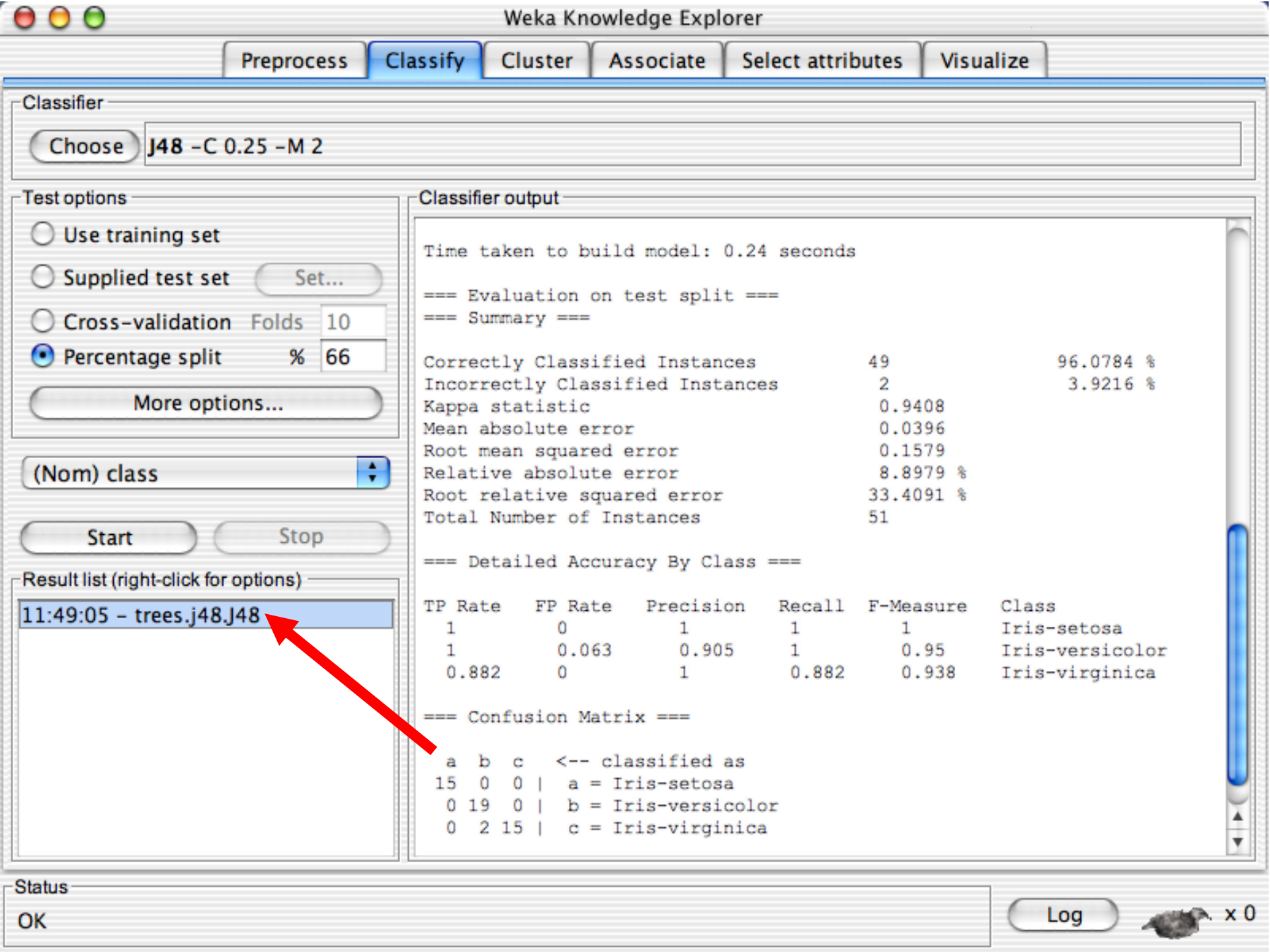
```
Time taken to build model: 0.24 seconds

=== 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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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
```

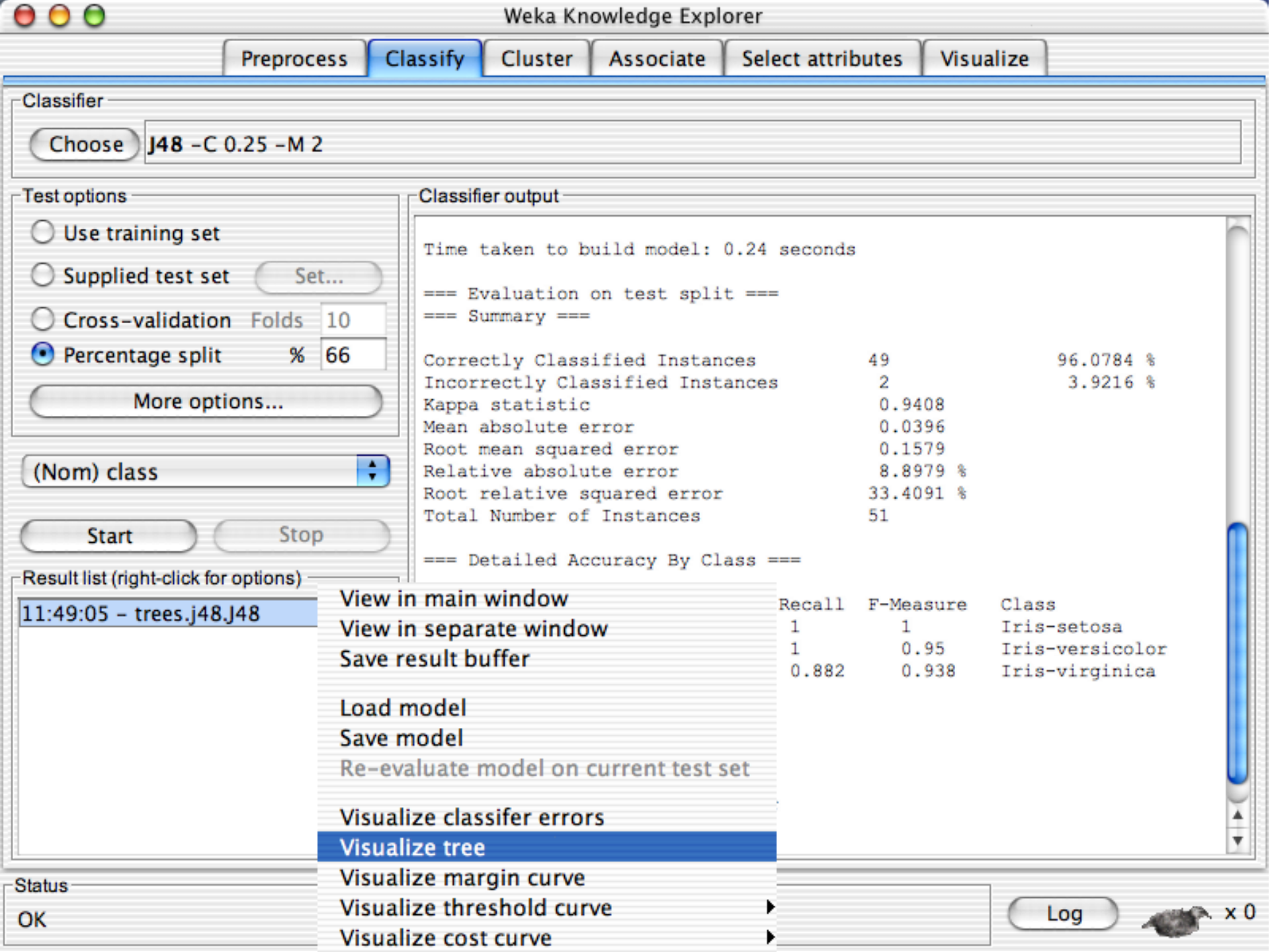
Status

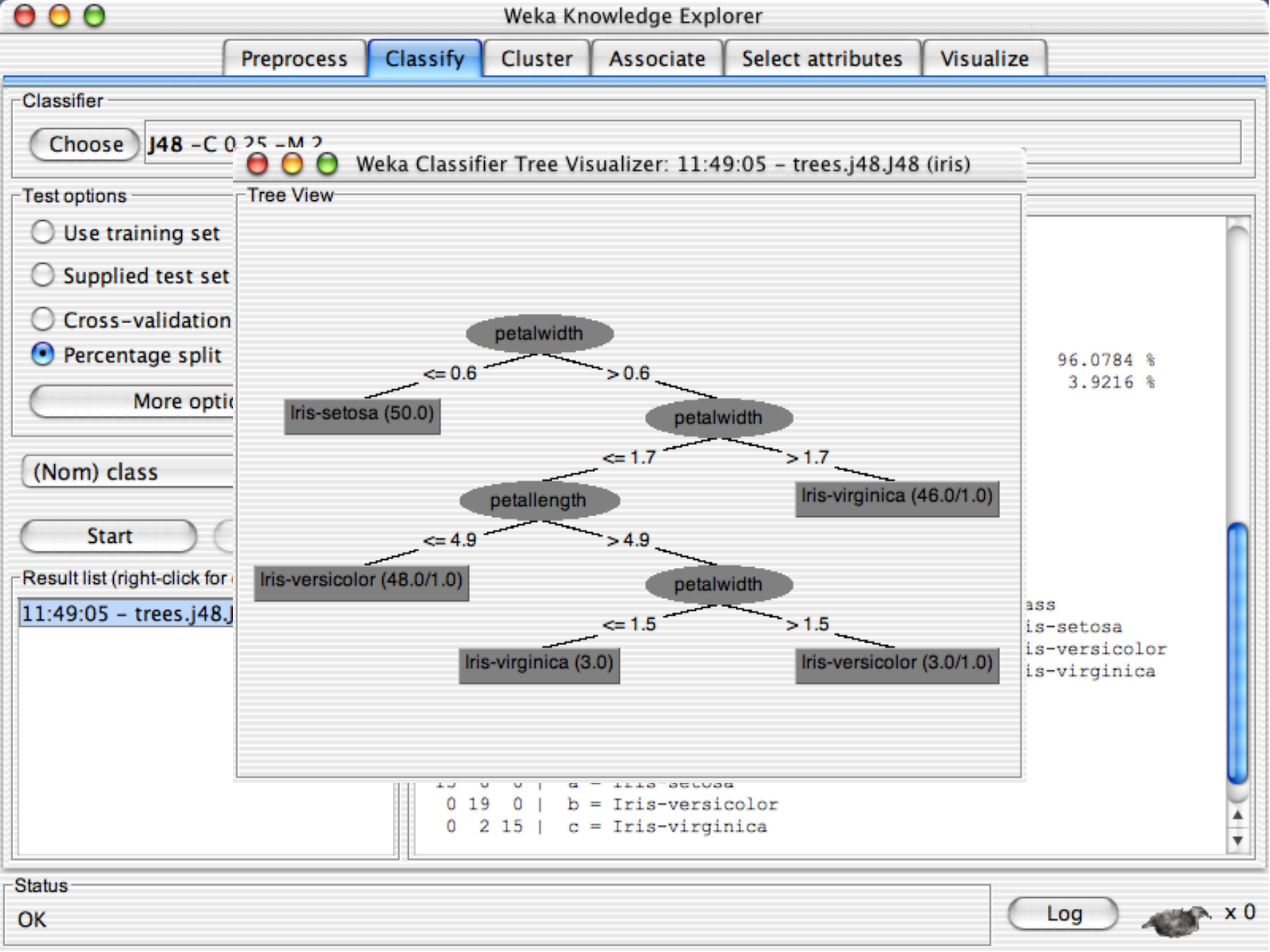
OK

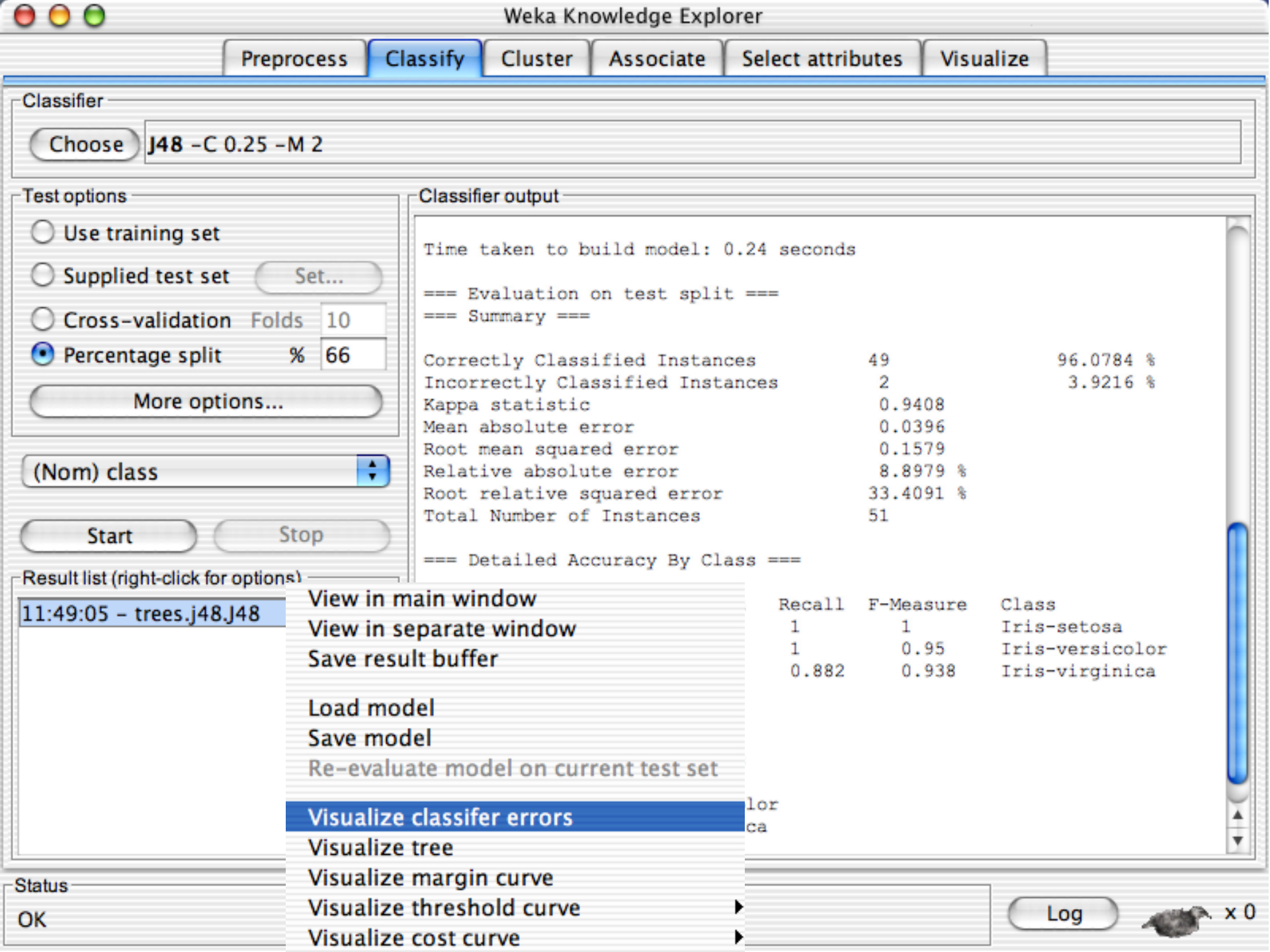
Log

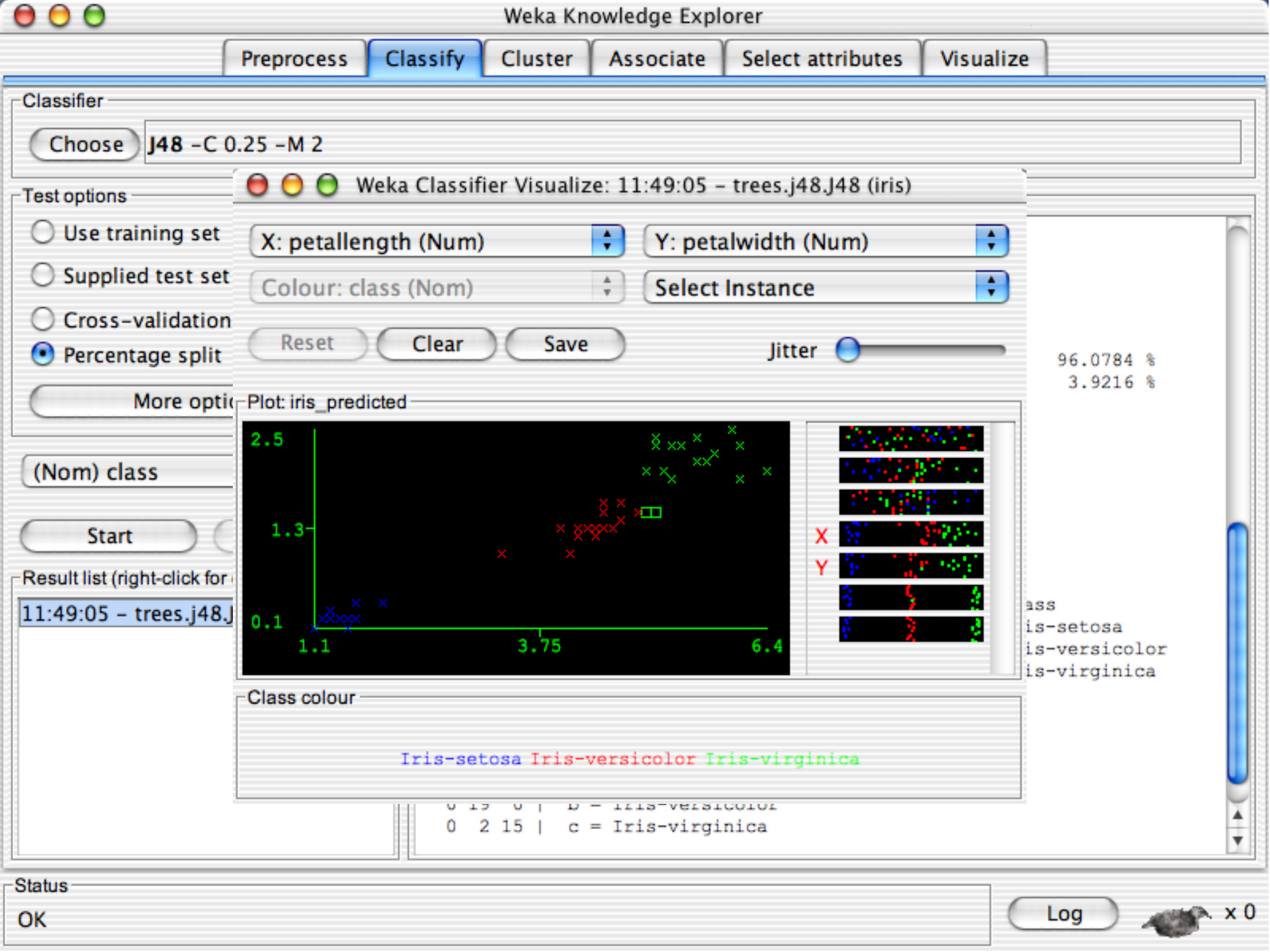
x 0


Weka Knowledge 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)
11:49:05 - trees.j48.J48
Classifier output
Time taken to build model: 0.24 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 Class
1 0 1 1 1 Iris-setosa
1 0.063 0.905 1 0.95 Iris-versicolor
0.882 0 1 0.882 0.938 Iris-virginica
=== 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 Knowledge 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)
11:49:05 - trees.j48.J48
Classifier output
Time taken to build model: 0.24 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 Class
1 1 Iris-setosa
1 0.95 Iris-versicolor
0.882 0.938 Iris-virginica
View in main window
View in separate window
Save result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifer errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Weka Classifier Tree Visualizer: 11:49:05 - trees.j48.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)
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optic
(Nom) class
Start
Result list (right-click for
11:49:05 - trees.j48.J
96.0784 %
3.9216 %
ass
is-setosa
is-versicolor
is-virginica
0 19 0 | b = Iris-versicolor
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0


Weka Knowledge 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)
11:49:05 - trees.j48.J48
Classifier output
Time taken to build model: 0.24 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 Class
1 1 Iris-setosa
1 0.95 Iris-versicolor
0.882 0.938 Iris-virginica
View in main window
View in separate window
Save result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifer errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
J48 -C 0.25 -M 2
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optic
(Nom) class
Start
Result list (right-click for
11:49:05 - trees.j48.J
Weka Classifier Visualize: 11:49:05 - trees.j48.J48 (iris)
X: petallength (Num)
Y: petalwidth (Num)
Colour: class (Nom)
Select Instance
Reset
Clear
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 %
ass
is-setosa
is-versicolor
is-virginica
0 2 15 | c = Iris-virginica
Status
OK
Log
x 0

Weka Knowledge 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)

11:49:05 - trees.j48.J48

Classifier output

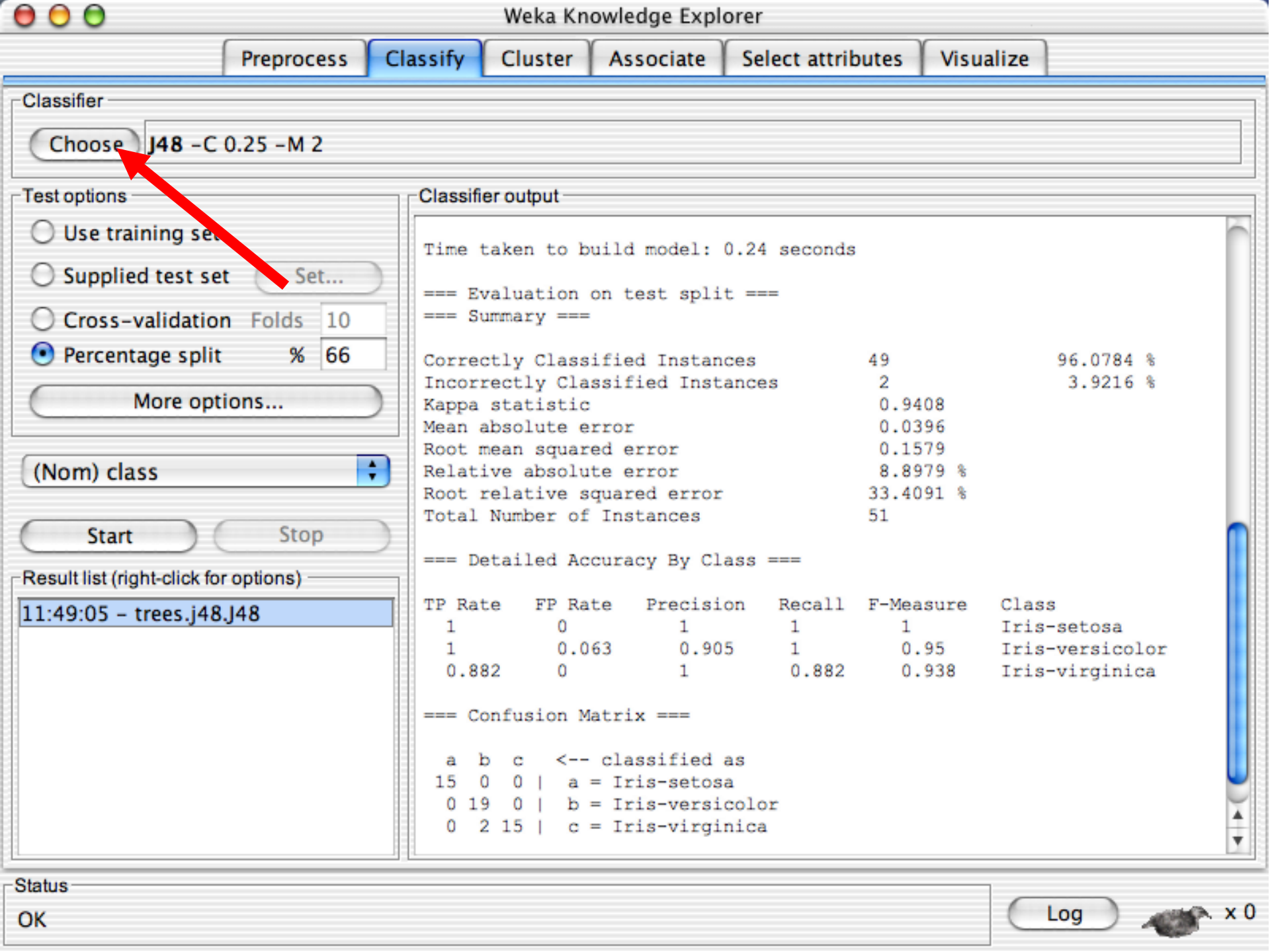
```
Time taken to build model: 0.24 seconds

=== 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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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
```

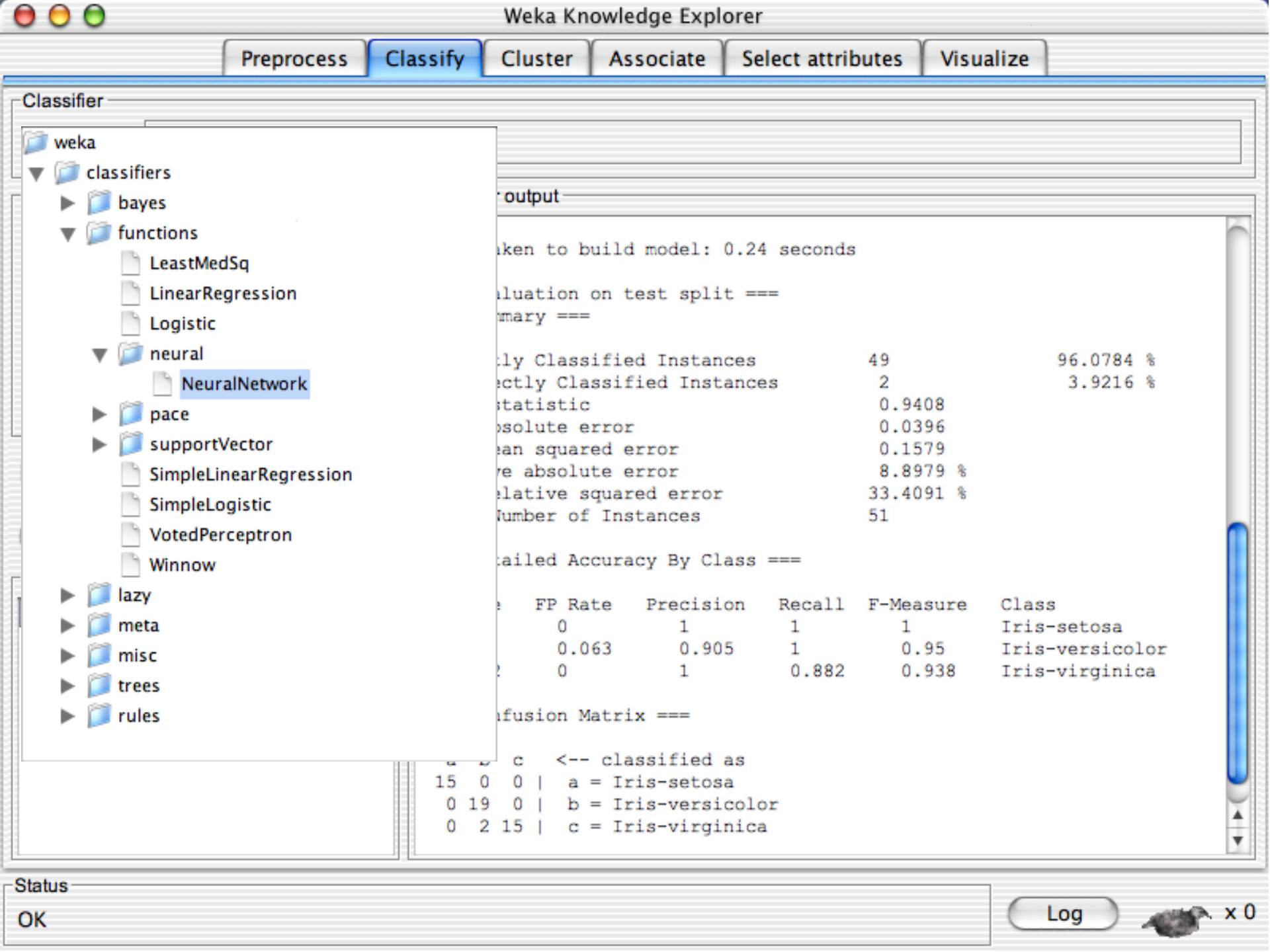
Status

OK

Log

x 0


Weka Knowledge 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
Time taken to build model: 0.24 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 Class
1 0 1 1 1 Iris-setosa
1 0.063 0.905 1 0.95 Iris-versicolor
0.882 0 1 0.882 0.938 Iris-virginica
=== 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
(Nom) class
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
LeastMedSq
LinearRegression
Logistic
neural
NeuralNetwork
pace
supportVector
SimpleLinearRegression
SimpleLogistic
VotedPerceptron
Winnow
lazy
meta
misc
trees
rules
output
ken to build model: 0.24 seconds
luation on test split ===
mary ===
ly Classified Instances 49 96.0784 %
ctly Classified Instances 2 3.9216 %
statistic 0.9408
solute error 0.0396
an squared error 0.1579
e absolute error 8.8979 %
lative squared error 33.4091 %
Number of Instances 51
ailed Accuracy By Class ===
FP Rate Precision Recall F-Measure Class
0 1 1 1 Iris-setosa
0.063 0.905 1 0.95 Iris-versicolor
0 1 0.882 0.938 Iris-virginica
fusion Matrix ===
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 Knowledge Explorer

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

Classifier

Choose **NeuralNetwork** -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)

11:49:05 - trees.j48.J48

Classifier output

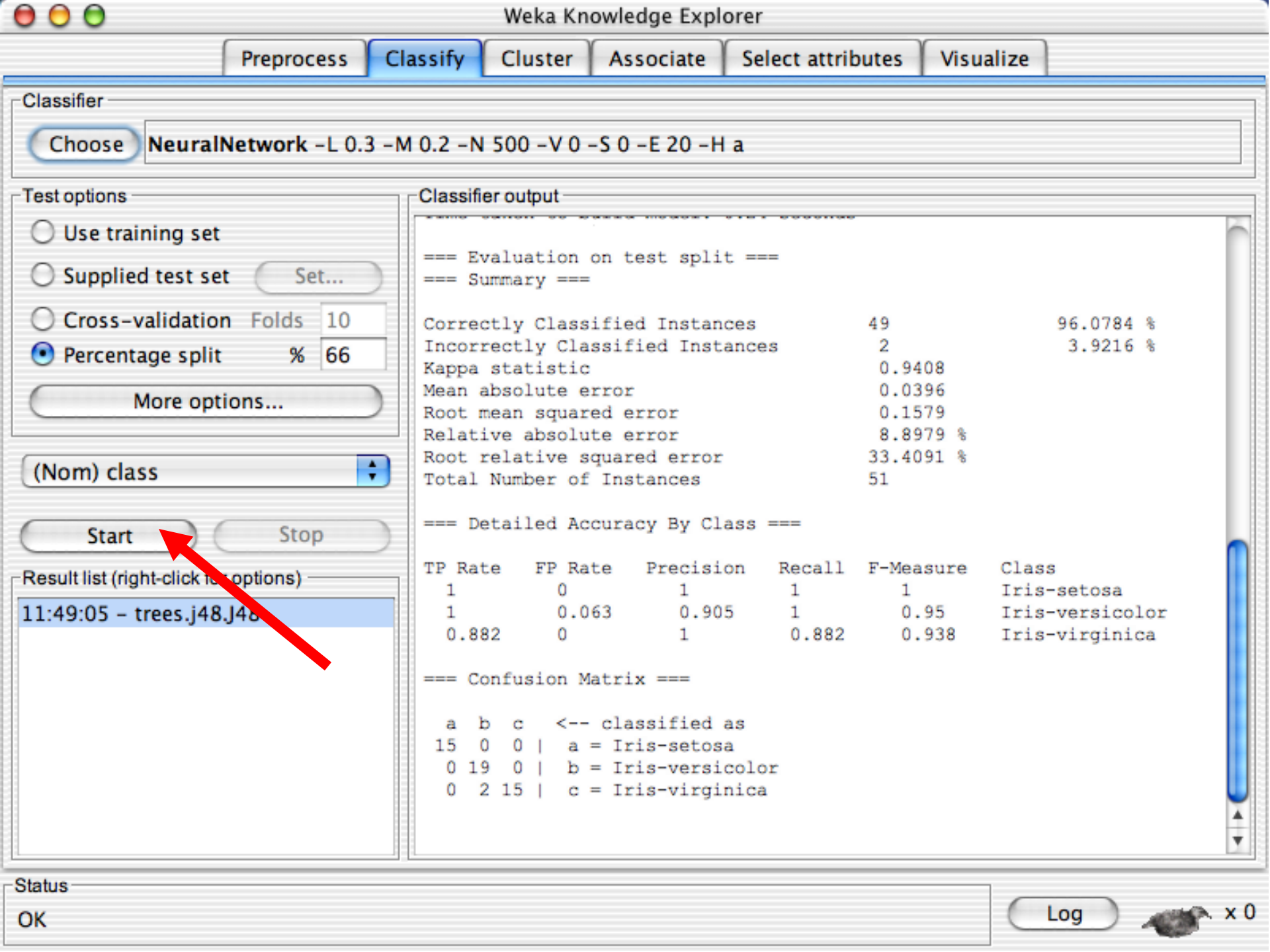
```
=== Evaluation on test split ===
=== Summary ===

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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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
```

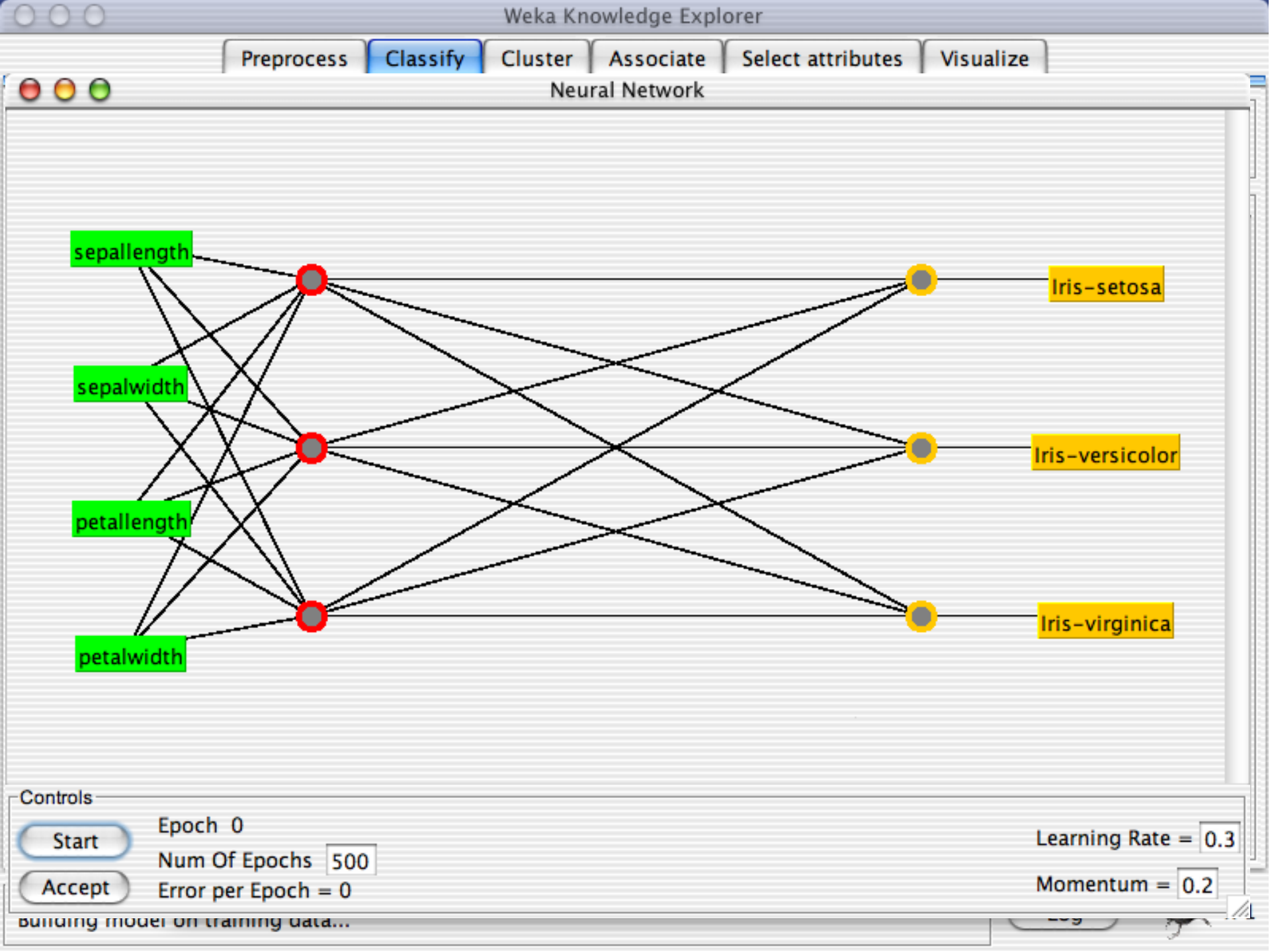
Status

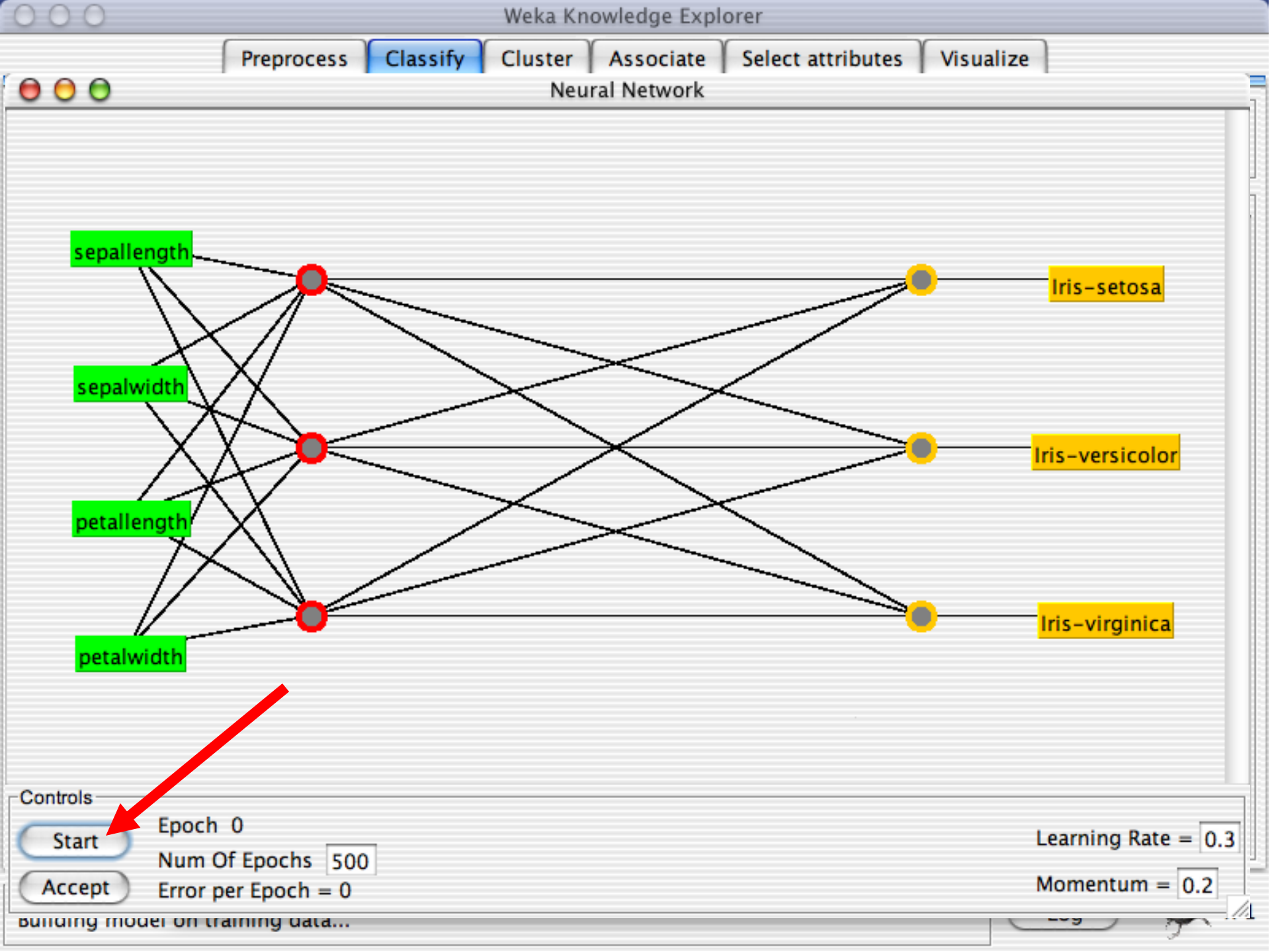
OK

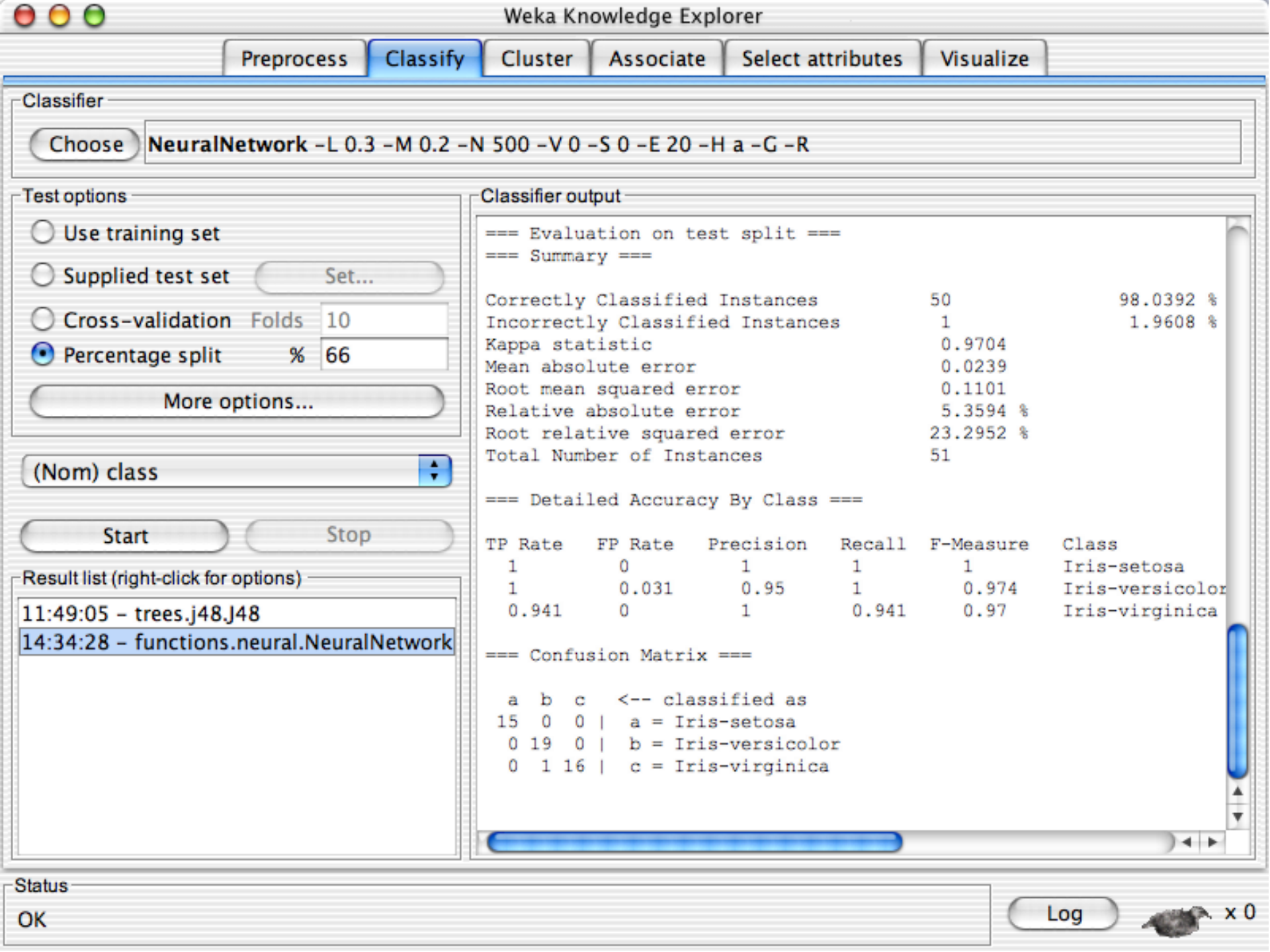
Log

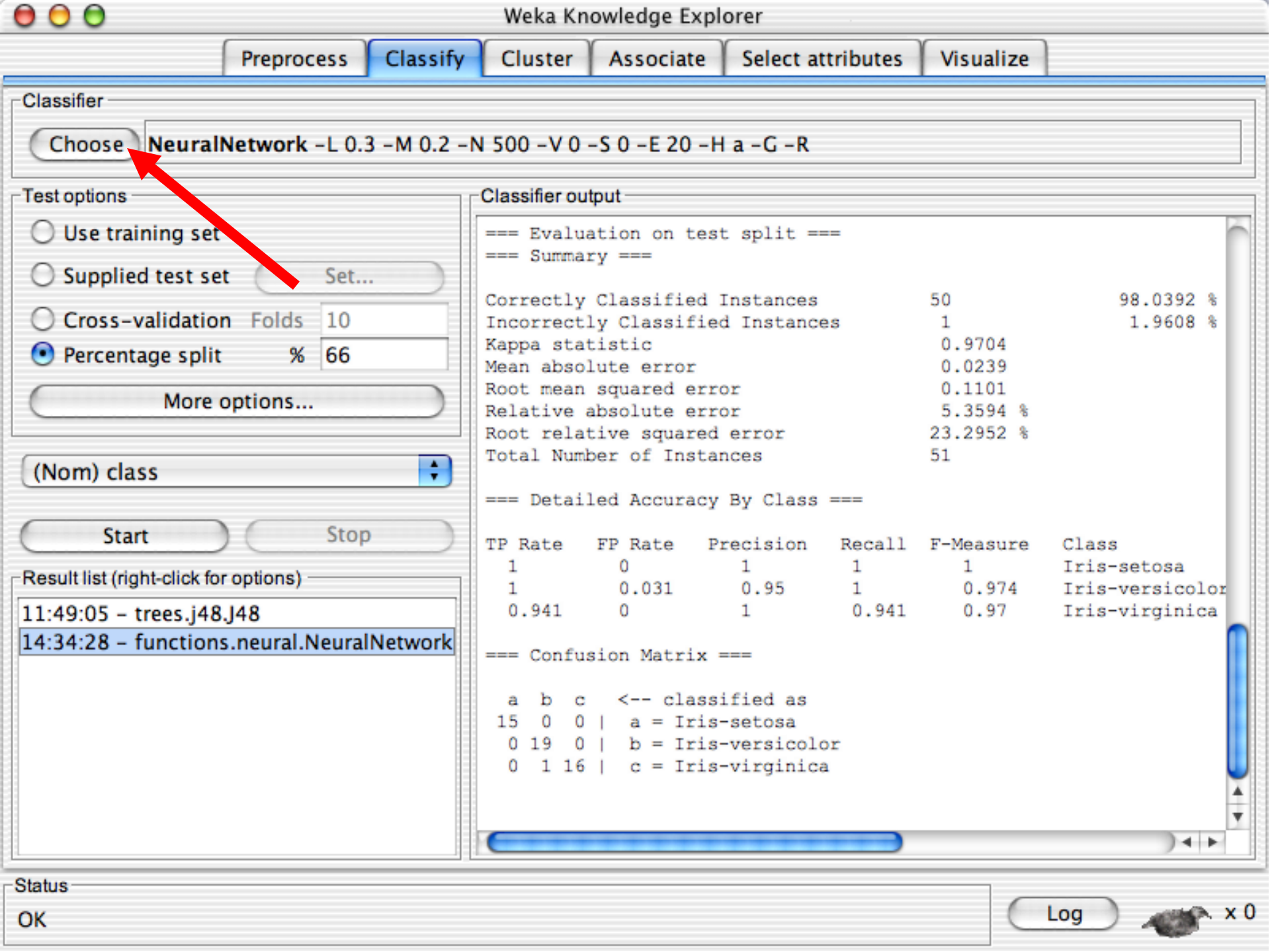
x 0

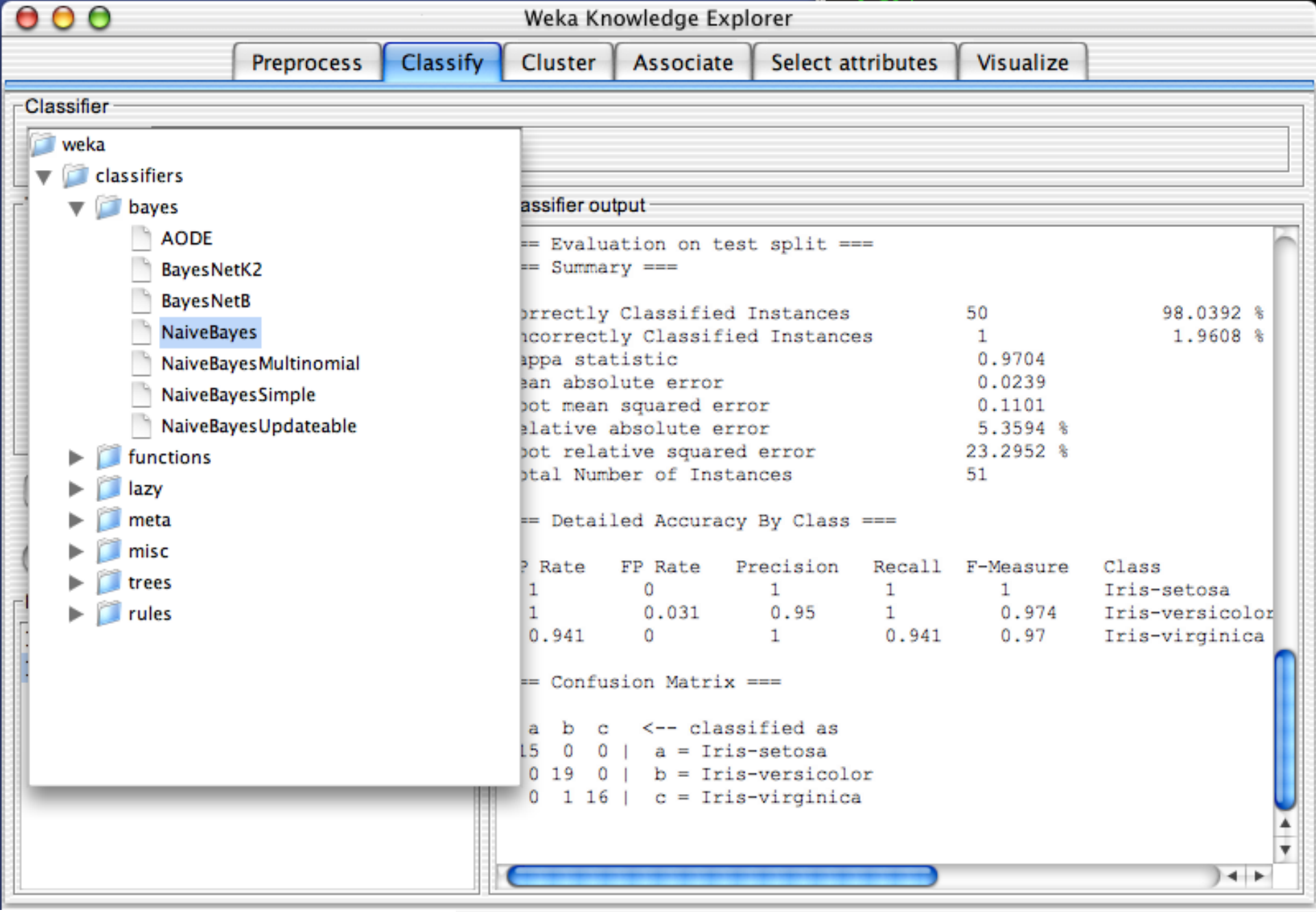

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
Classifier output
=== 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 Class
1 0 1 1 1 Iris-setosa
1 0.063 0.905 1 0.95 Iris-versicolor
0.882 0 1 0.882 0.938 Iris-virginica
=== 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 Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
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

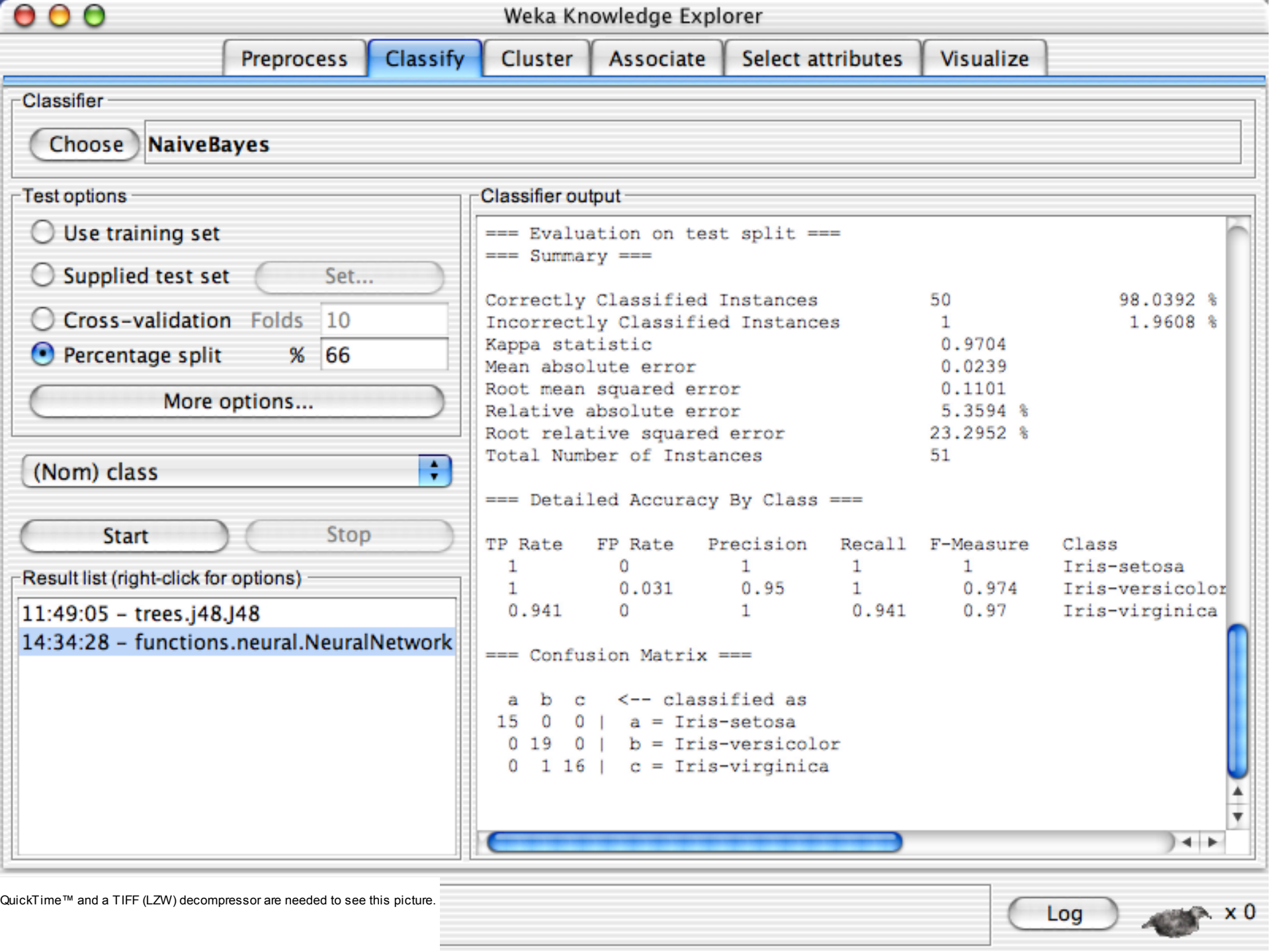

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Neural Network
sepallength
sepalwidth
petallength
petalwidth
Iris-setosa
Iris-versicolor
Iris-virginica
Controls
Start
Accept
Epoch 0
Num Of Epochs 500
Error per Epoch = 0
Learning Rate = 0.3
Momentum = 0.2


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NeuralNetwork -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)
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
Status
OK
Log
x 0

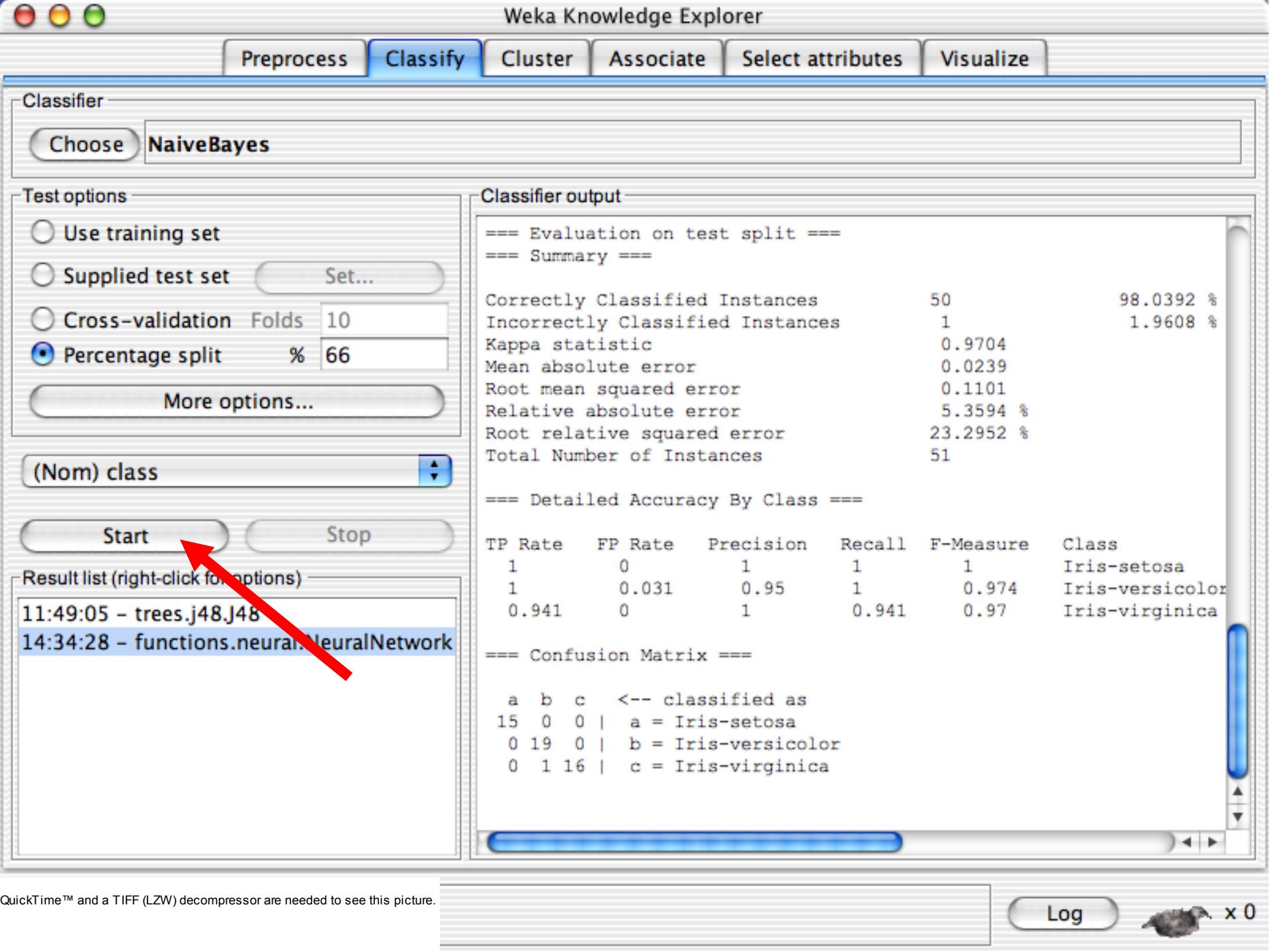

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a -G -R
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
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
Status
OK
Log
x 0



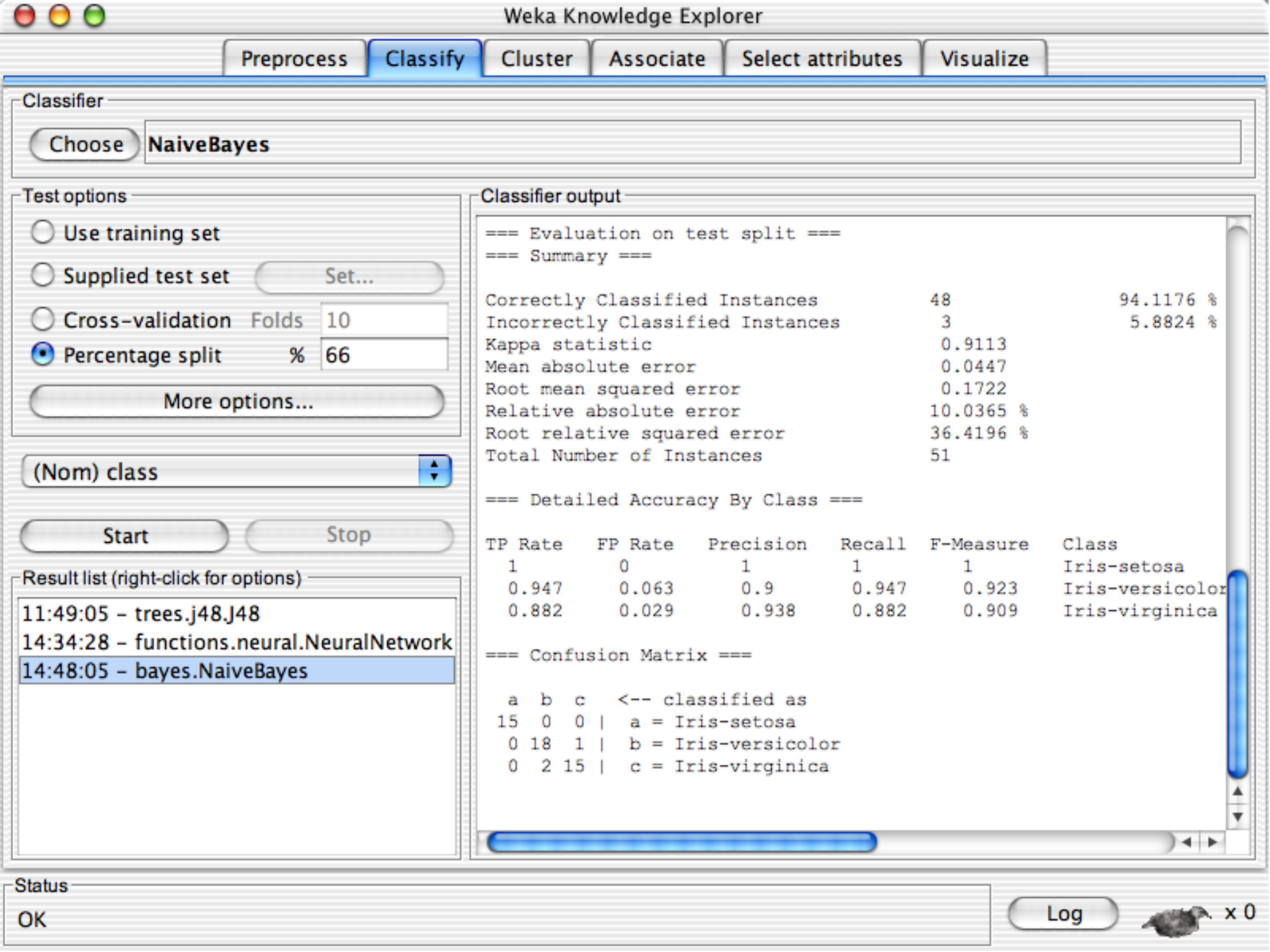

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

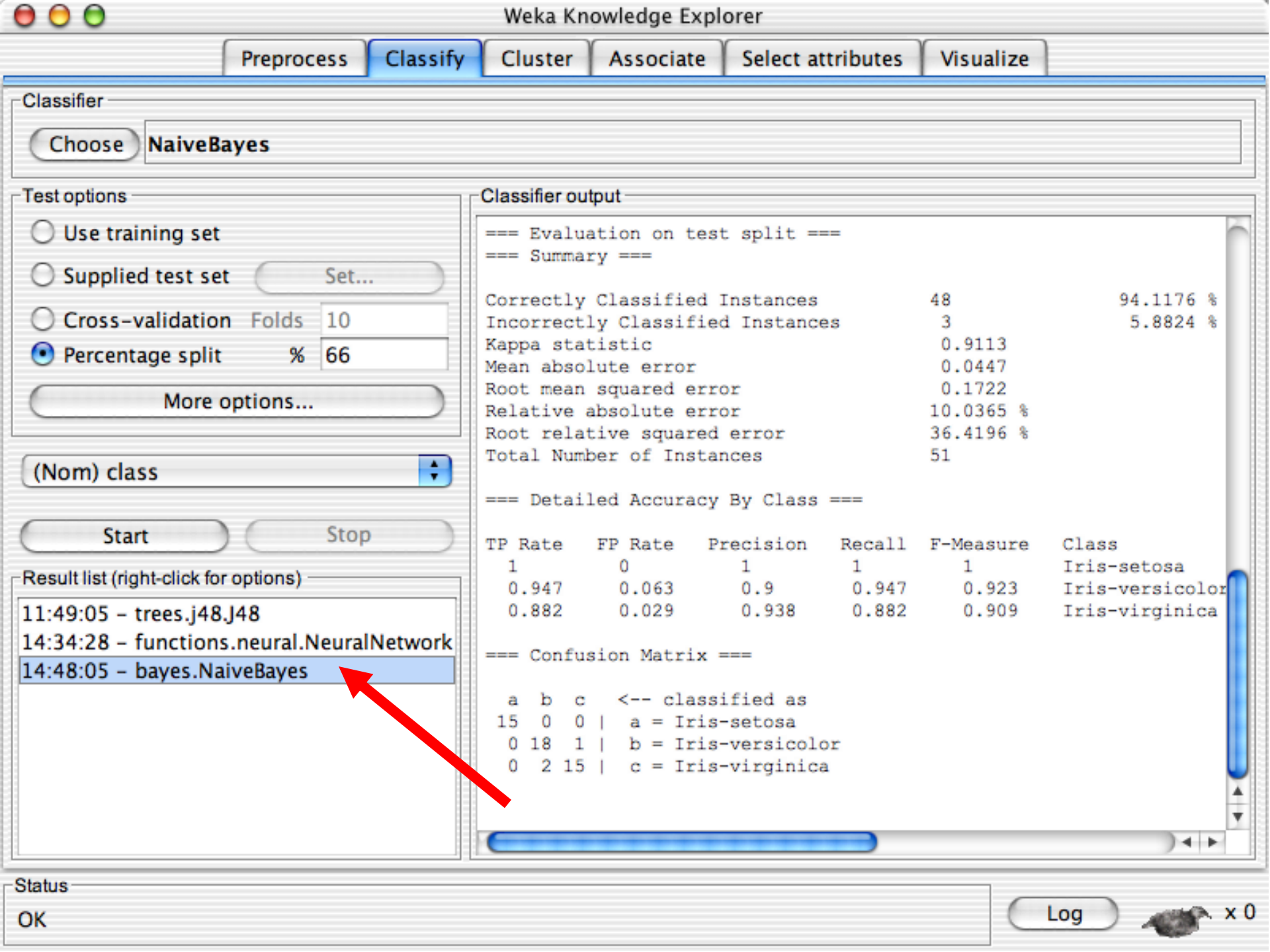


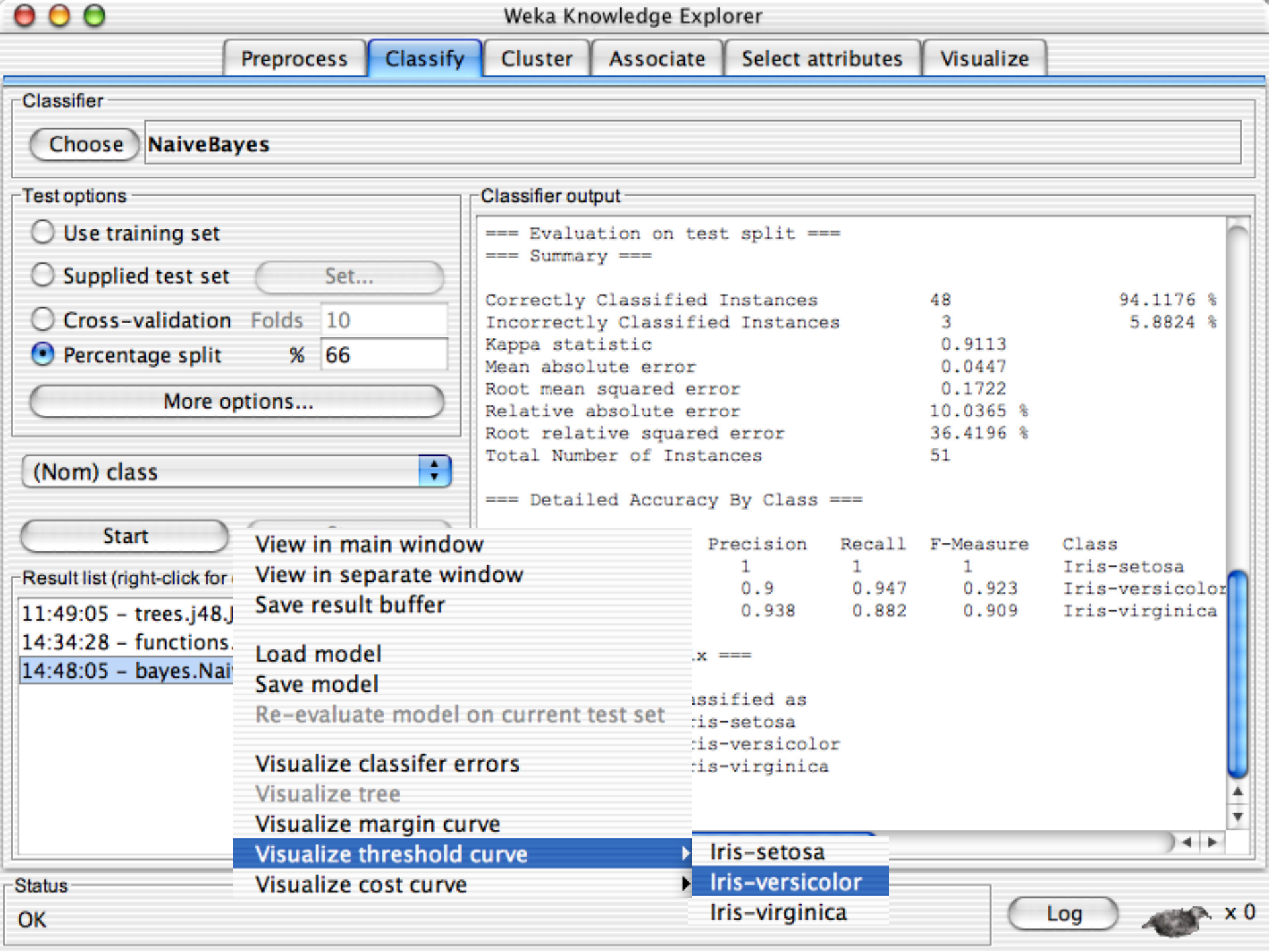

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

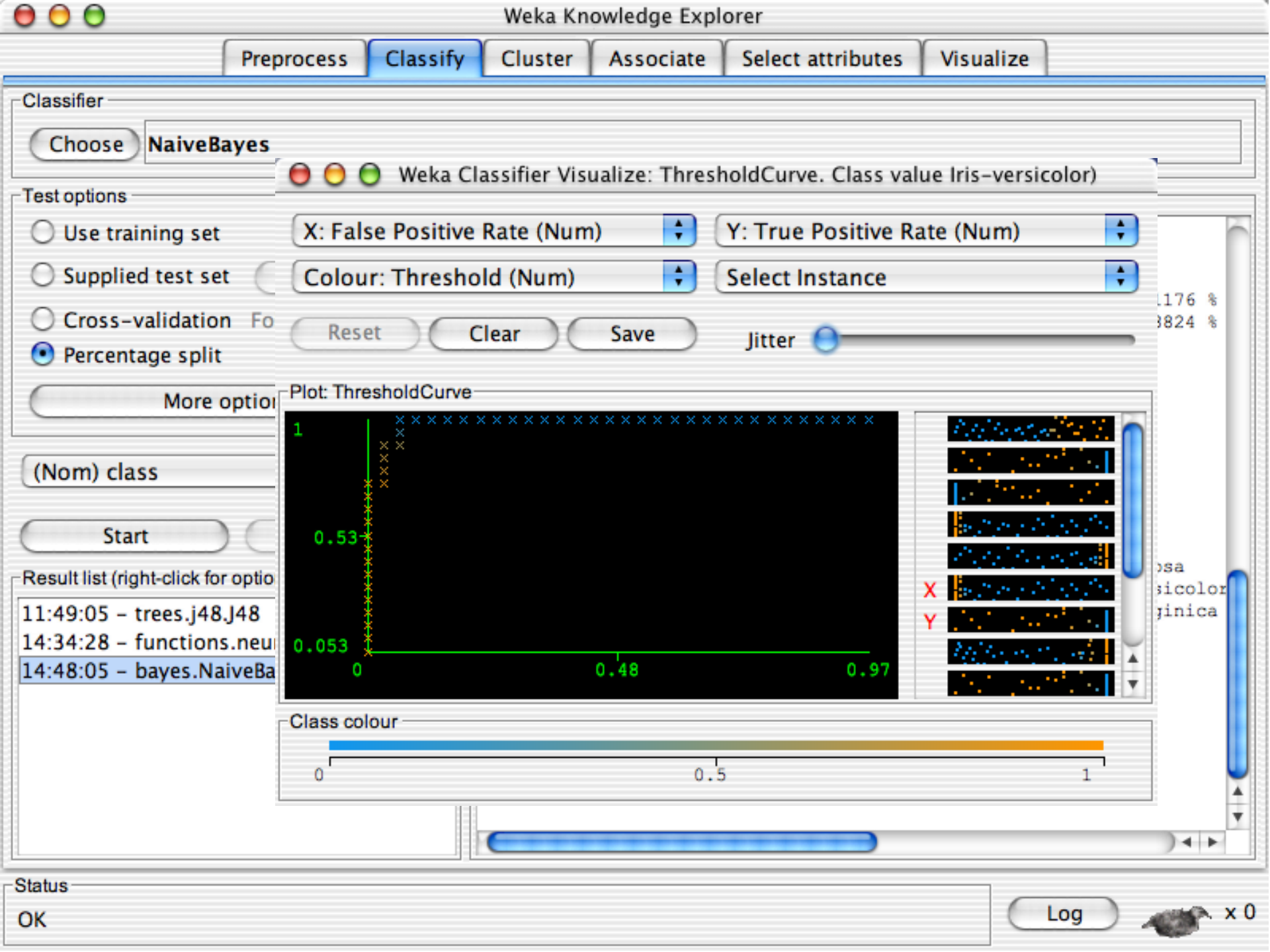


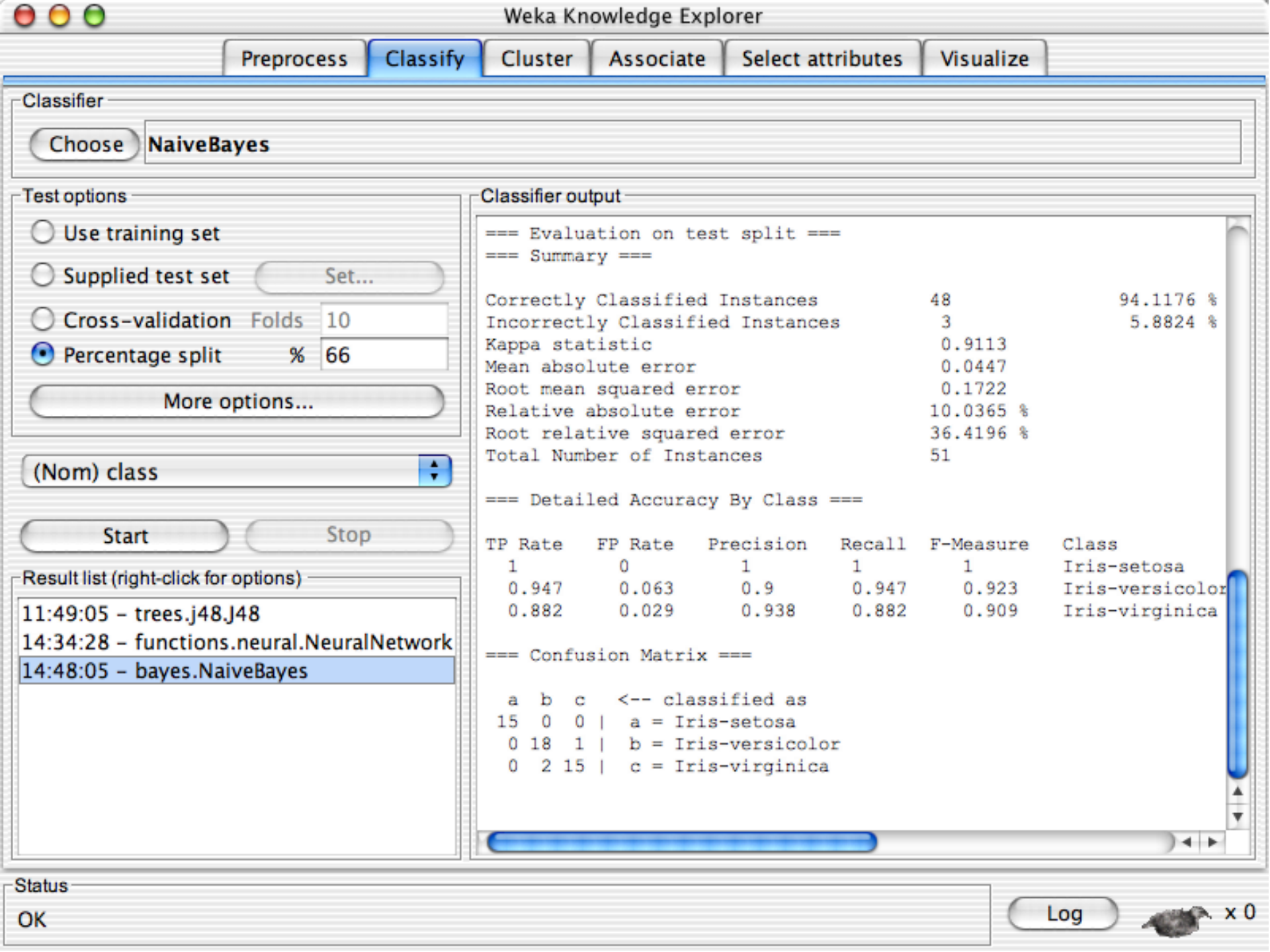

QuickTime™ and a TIFF (LZW) decompressor are needed to see this picture.

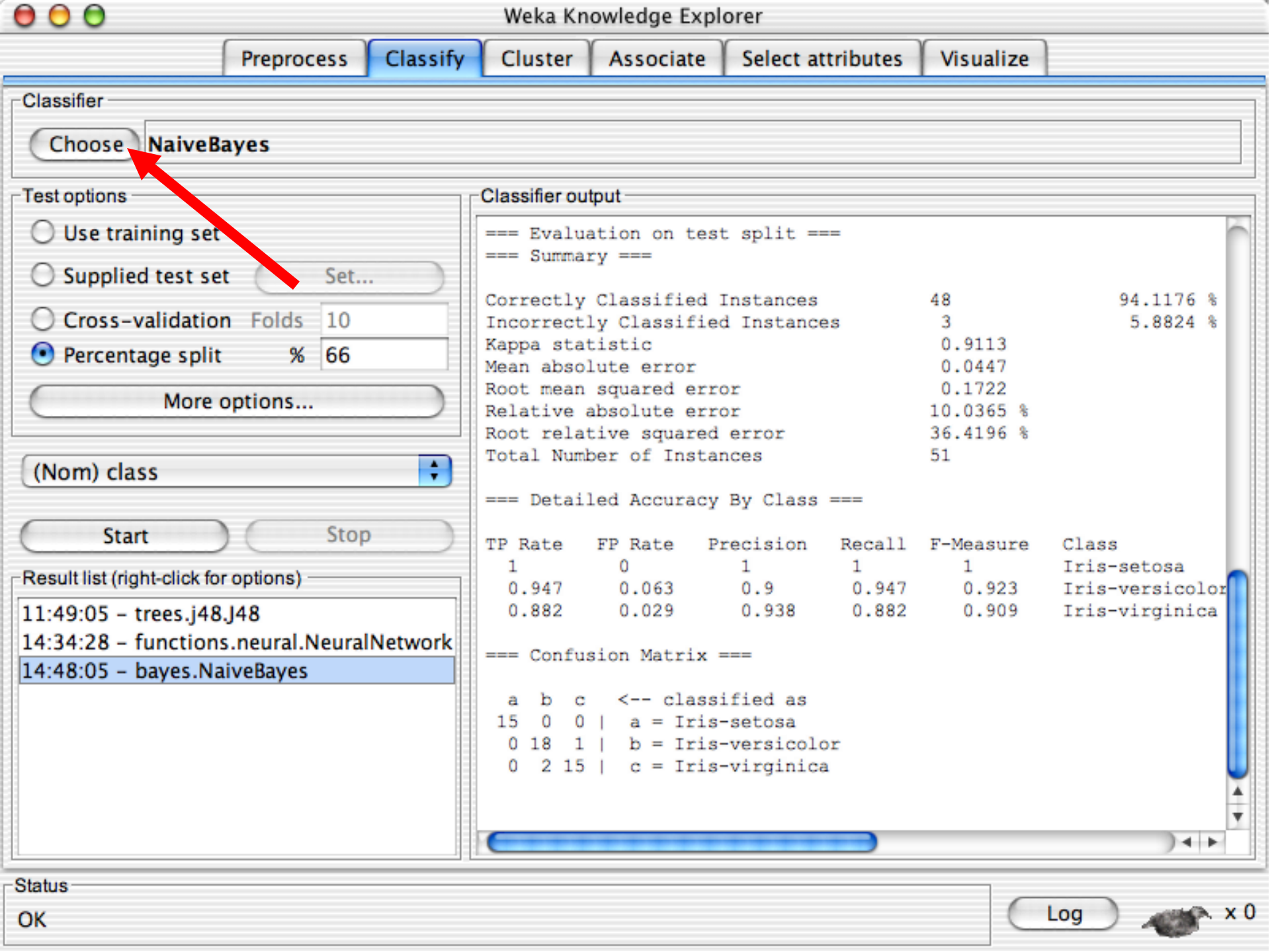

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
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

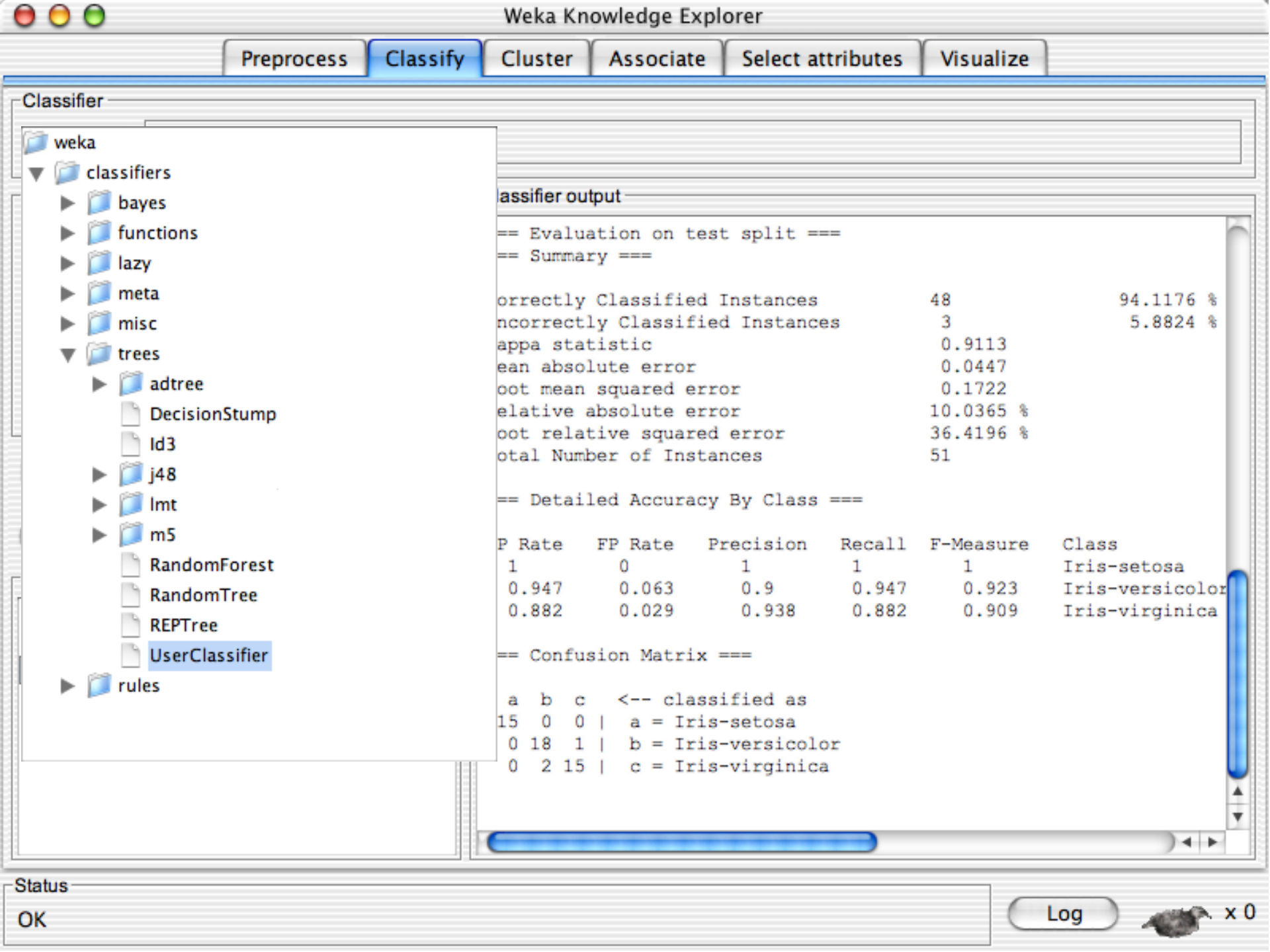

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
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

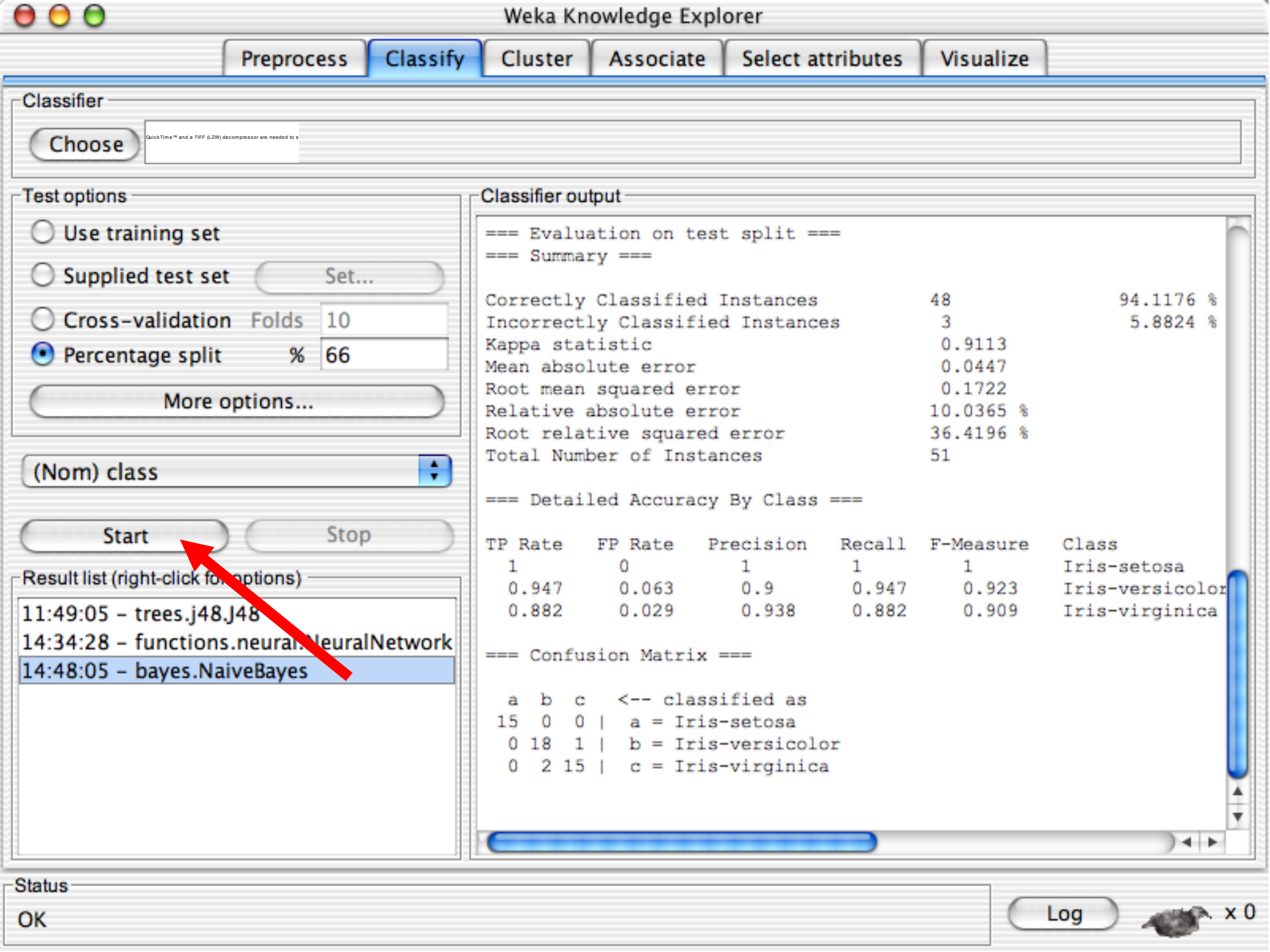

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
Classifier output
=== Evaluation on test split ===
=== Summary ===
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 ===
Precision Recall F-Measure Class
1 1 1 Iris-setosa
0.9 0.947 0.923 Iris-versicolor
0.938 0.882 0.909 Iris-virginica
(Nom) class
Start
Result list (right-click for
11:49:05 - trees.j48.J
14:34:28 - functions.
14:48:05 - bayes.Nai
View in main window
View in separate window
Save result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize classifer errors
Visualize tree
Visualize margin curve
Visualize threshold curve
Visualize cost curve
Iris-setosa
Iris-versicolor
Iris-virginica
Status
OK
Log
x 0

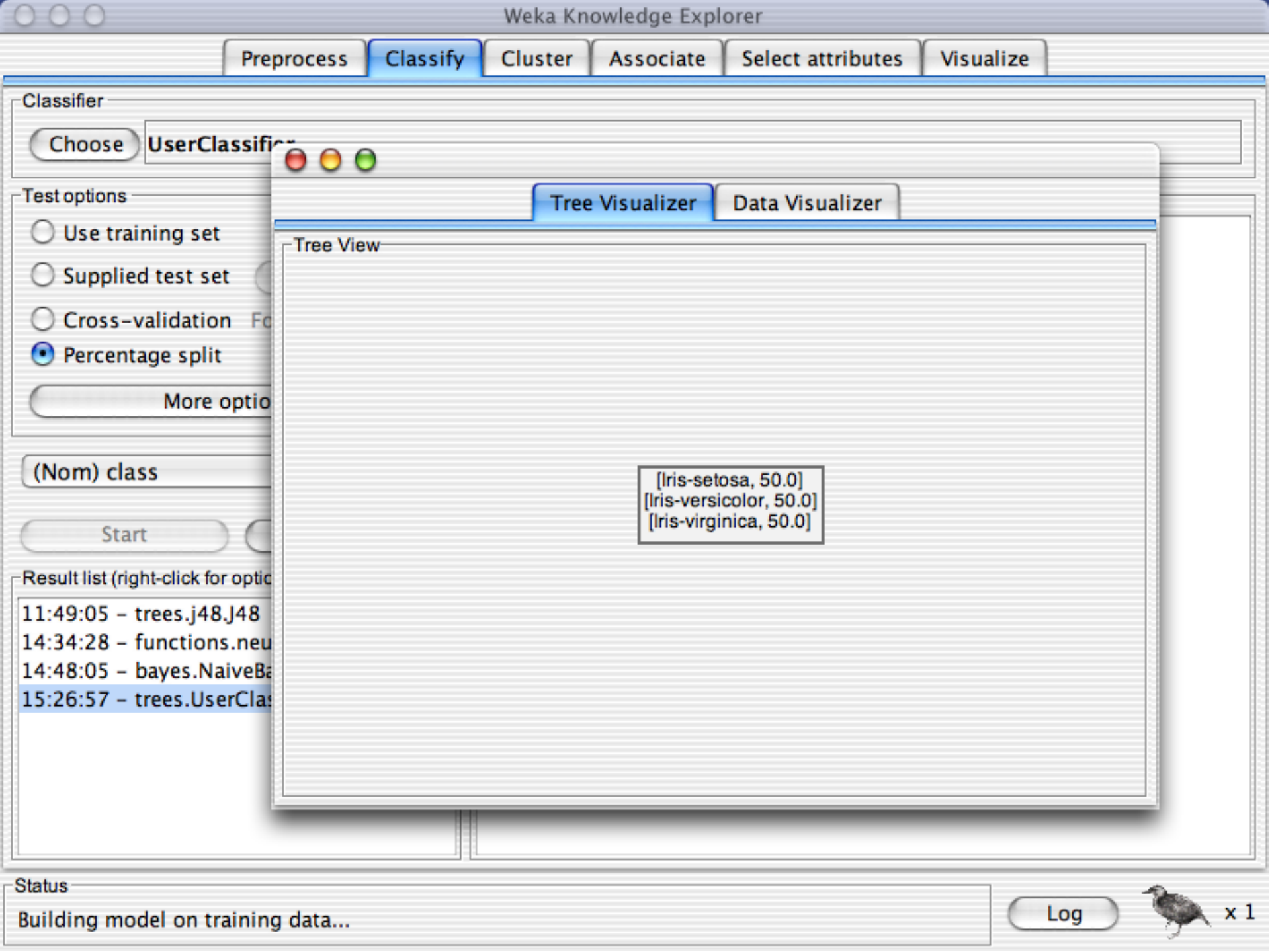

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optio
(Nom) class
Start
Result list (right-click for optio
11:49:05 - trees.j48.J48
14:34:28 - functions.neu
14:48:05 - bayes.NaiveBa
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
Save
Jitter
Plot: ThresholdCurve
1
0.53
0.053
0
0.48
0.97
X
Y
Class colour
0
0.5
1
Status
OK
Log
x 0

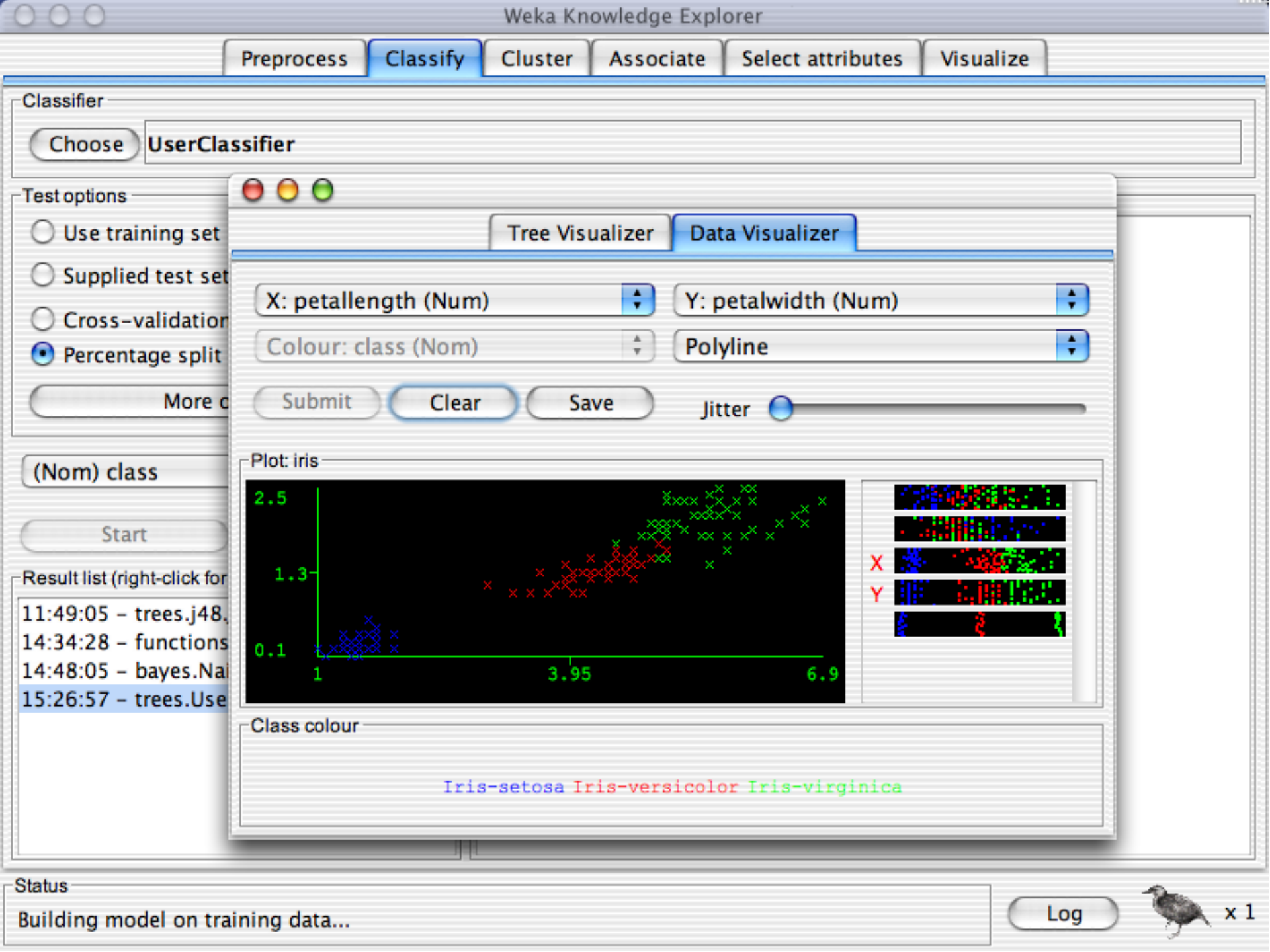

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
NaiveBayes
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Nom) class
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
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


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
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


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
misc
trees
adtree
DecisionStump
Id3
j48
lmt
m5
RandomForest
RandomTree
REPTree
UserClassifier
rules
lassifier output
== Evaluation on test split ===
== Summary ===
orrectly Classified Instances 48 94.1176 %
ncorrectly Classified Instances 3 5.8824 %
appa statistic 0.9113
ean absolute error 0.0447
oot mean squared error 0.1722
elative absolute error 10.0365 %
oot relative squared error 36.4196 %
otal Number of Instances 51
== Detailed Accuracy By Class ===
P 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


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


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More optio
(Nom) class
Start
Result list (right-click for optic
11:49:05 - trees.j48.J48
14:34:28 - functions.neu
14:48:05 - bayes.NaiveBa
15:26:57 - trees.UserCla
Tree Visualizer
Data Visualizer
Tree View
[Iris-setosa, 50.0]
[Iris-versicolor, 50.0]
[Iris-virginica, 50.0]
Status
Building model on training data...
Log
x 1


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More c
(Nom) class
Start
Result list (right-click for
11:49:05 - trees.j48.
14:34:28 - functions
14:48:05 - bayes.Nai
15:26:57 - trees.Use
Tree Visualizer
Data Visualizer
X: petallength (Num)
Y: petalwidth (Num)
Colour: class (Nom)
Polyline
Submit
Clear
Save
Jitter
Plot: iris
2.5
1.3
0.1
1
3.95
6.9
X
Y
Class colour
Iris-setosa Iris-versicolor Iris-virginica
Status
Building model on training data...
Log
x 1

Weka Knowledge Explorer

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

Classifier

Choose **UserClassifier**

Test options

Use training set

Supplied test set

Cross-validation

Percentage split

More c

(Nom) class

Start

Result list (right-click for

11:49:05 - trees.j48.

14:34:28 - functions

14:48:05 - bayes.Nai

15:26:57 - trees.Use

Tree Visualizer | Data Visualizer

X: petallength (Num)

Y: petalwidth (Num)

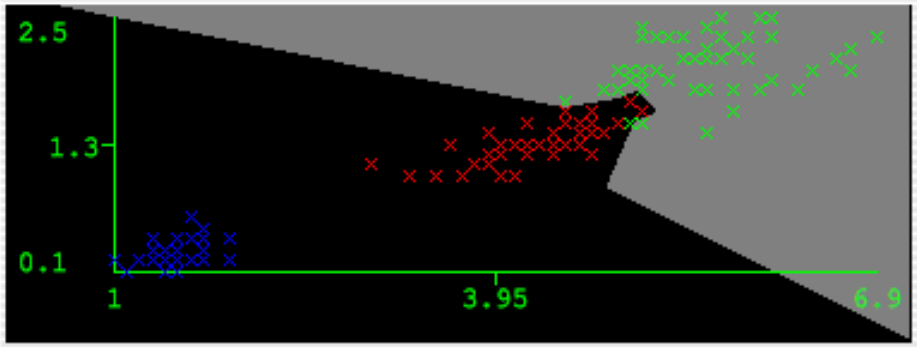
Colour: class (Nom)

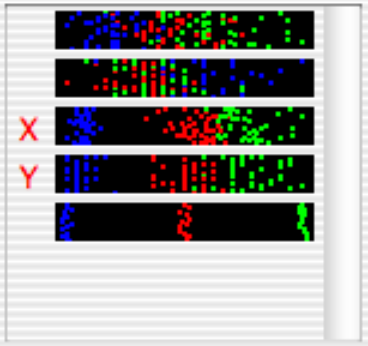
Polyline

Submit | Clear | Save

Jitter

Plot: iris




Class colour

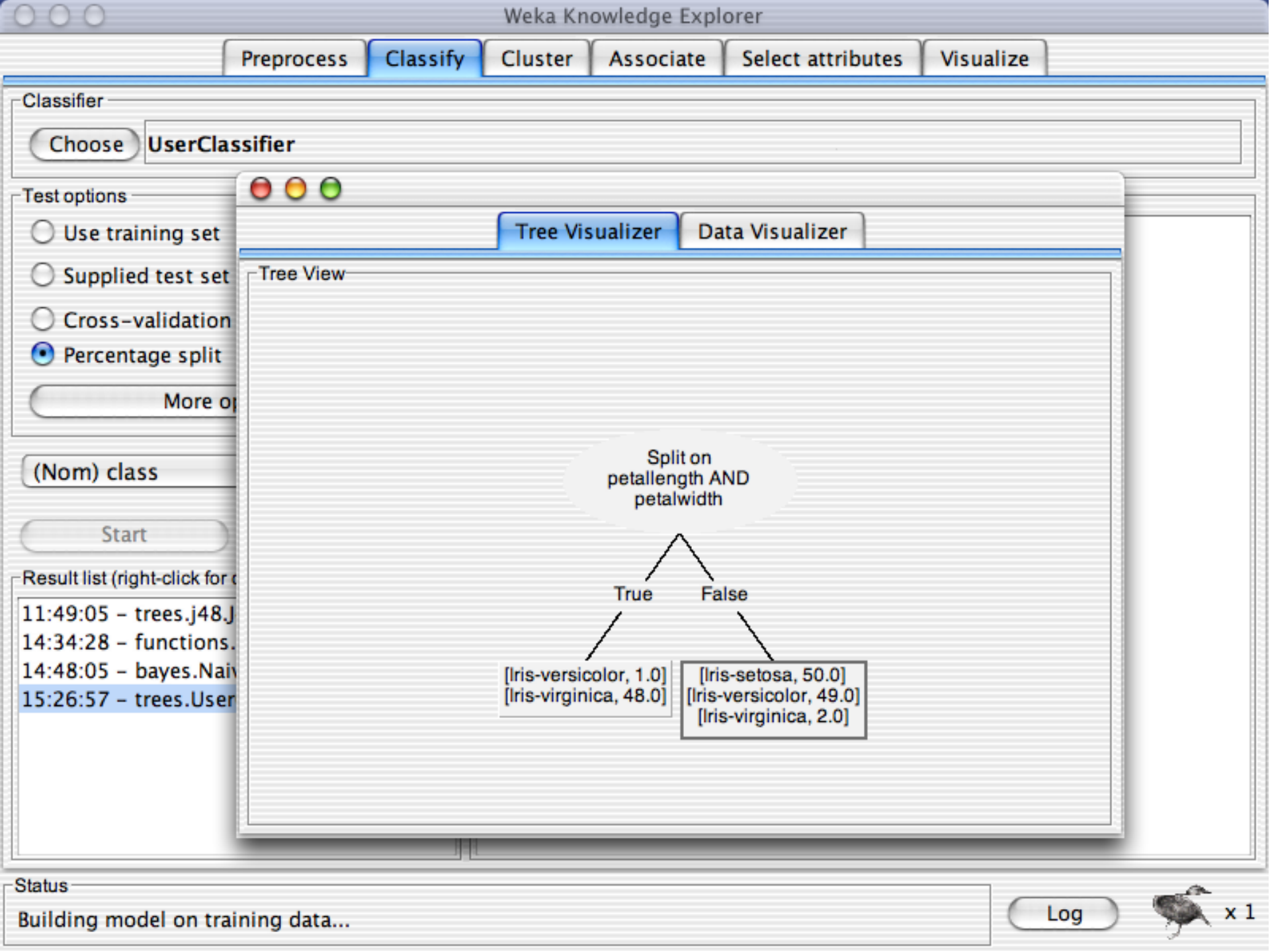
Iris-setosa Iris-versicolor Iris-virginica

Status

Building model on training data...

Log

x 1


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
UserClassifier
Test options
Use training set
Supplied test set
Cross-validation
Percentage split
More op
(Nom) class
Start
Result list (right-click for
11:49:05 - trees.j48.J
14:34:28 - functions.
14:48:05 - bayes.Naiv
15:26:57 - trees.User
Tree Visualizer
Data Visualizer
Tree View
Split on
petallength AND
petalwidth
True
False
[Iris-versicolor, 1.0]
[Iris-virginica, 48.0]
[Iris-setosa, 50.0]
[Iris-versicolor, 49.0]
[Iris-virginica, 2.0]
Status
Building model on training data...
Log
x 1

Weka Knowledge 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

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

11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier

Classifier output

```
=== 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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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 Knowledge 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

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

11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier

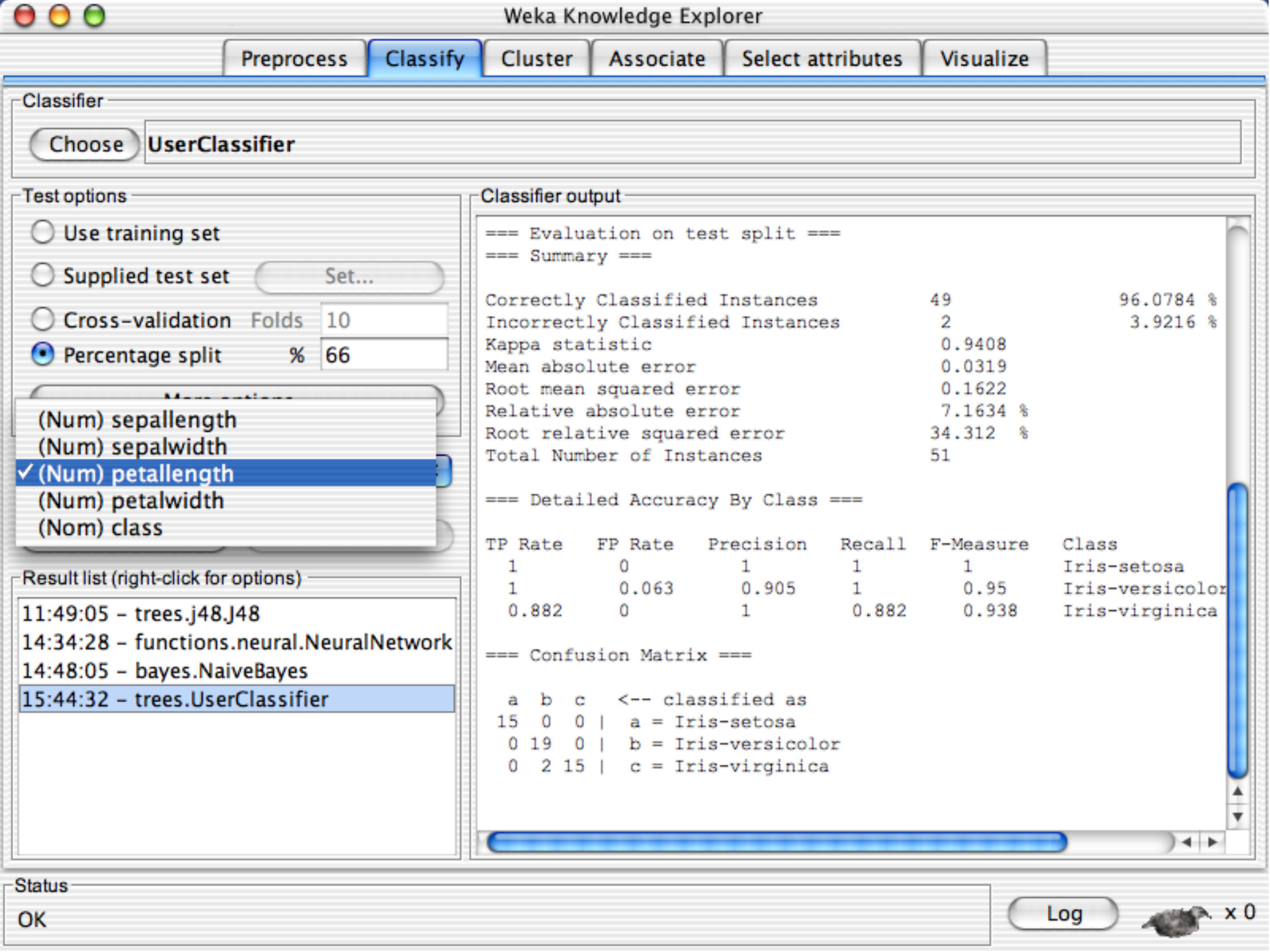
Classifier output

```
=== 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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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 Knowledge 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
(Num) sepallength
(Num) sepalwidth
(Num) petallength
(Num) petalwidth
(Nom) class
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier
Classifier output
=== 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 Class
1 0 1 1 1 Iris-setosa
1 0.063 0.905 1 0.95 Iris-versicolor
0.882 0 1 0.882 0.938 Iris-virginica
=== 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 Knowledge 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

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

11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier

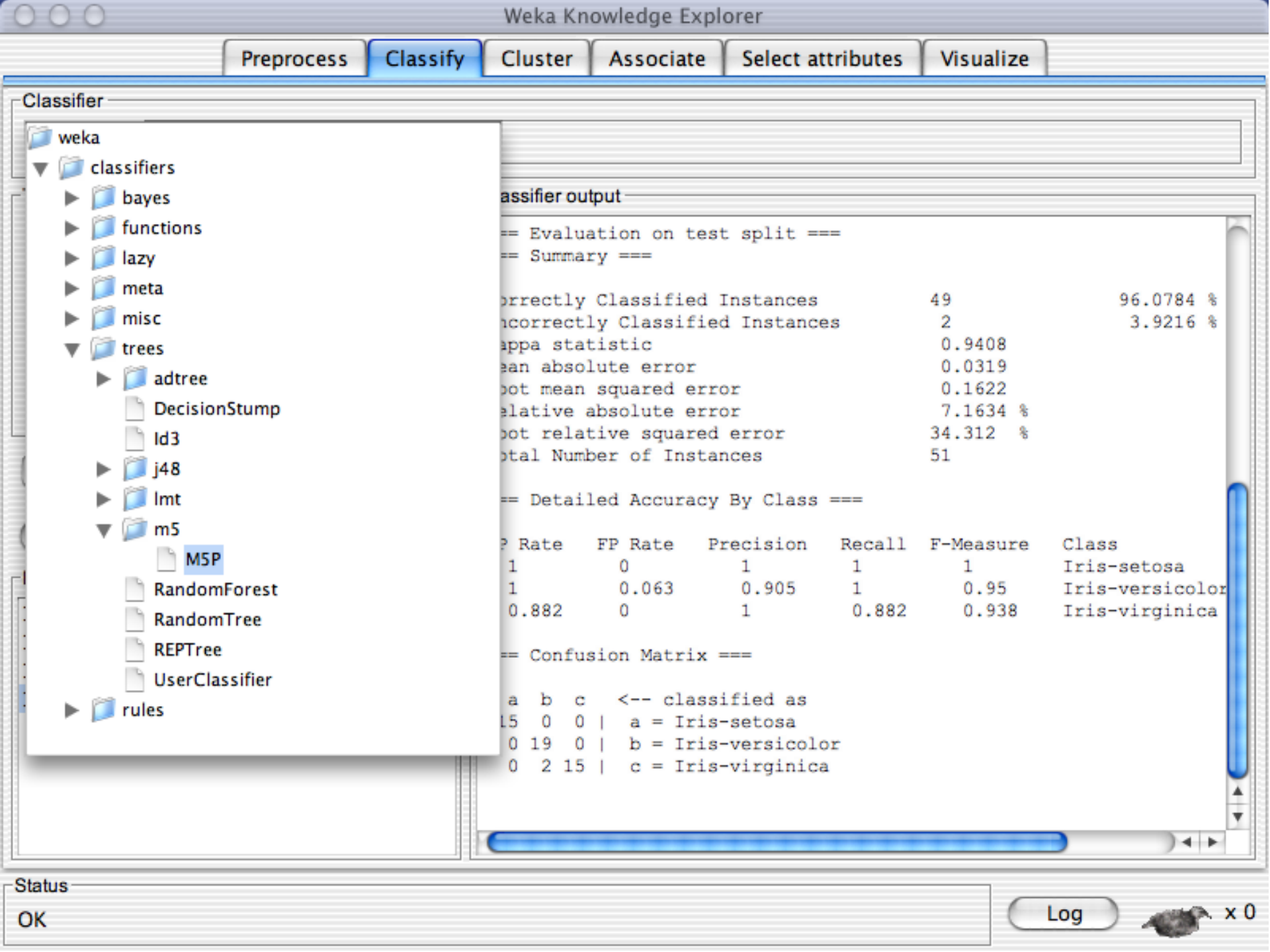
Classifier output

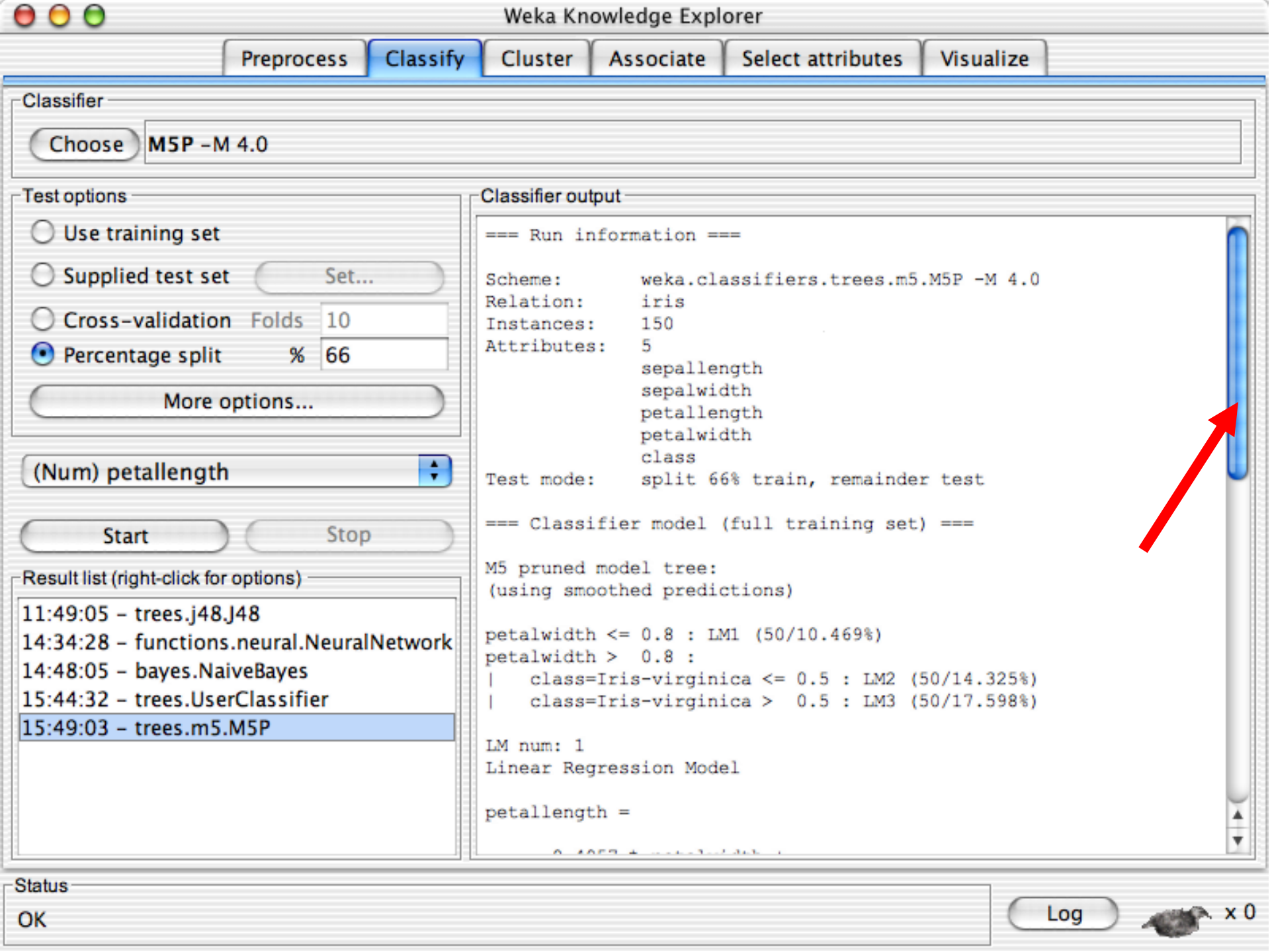
```
=== Evaluation on test split ===
=== Summary ===

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   Class
  1          0          1          1          1        Iris-setosa
  1          0.063      0.905      1          0.95     Iris-versicolor
  0.882      0          1          0.882      0.938    Iris-virginica

=== 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
```

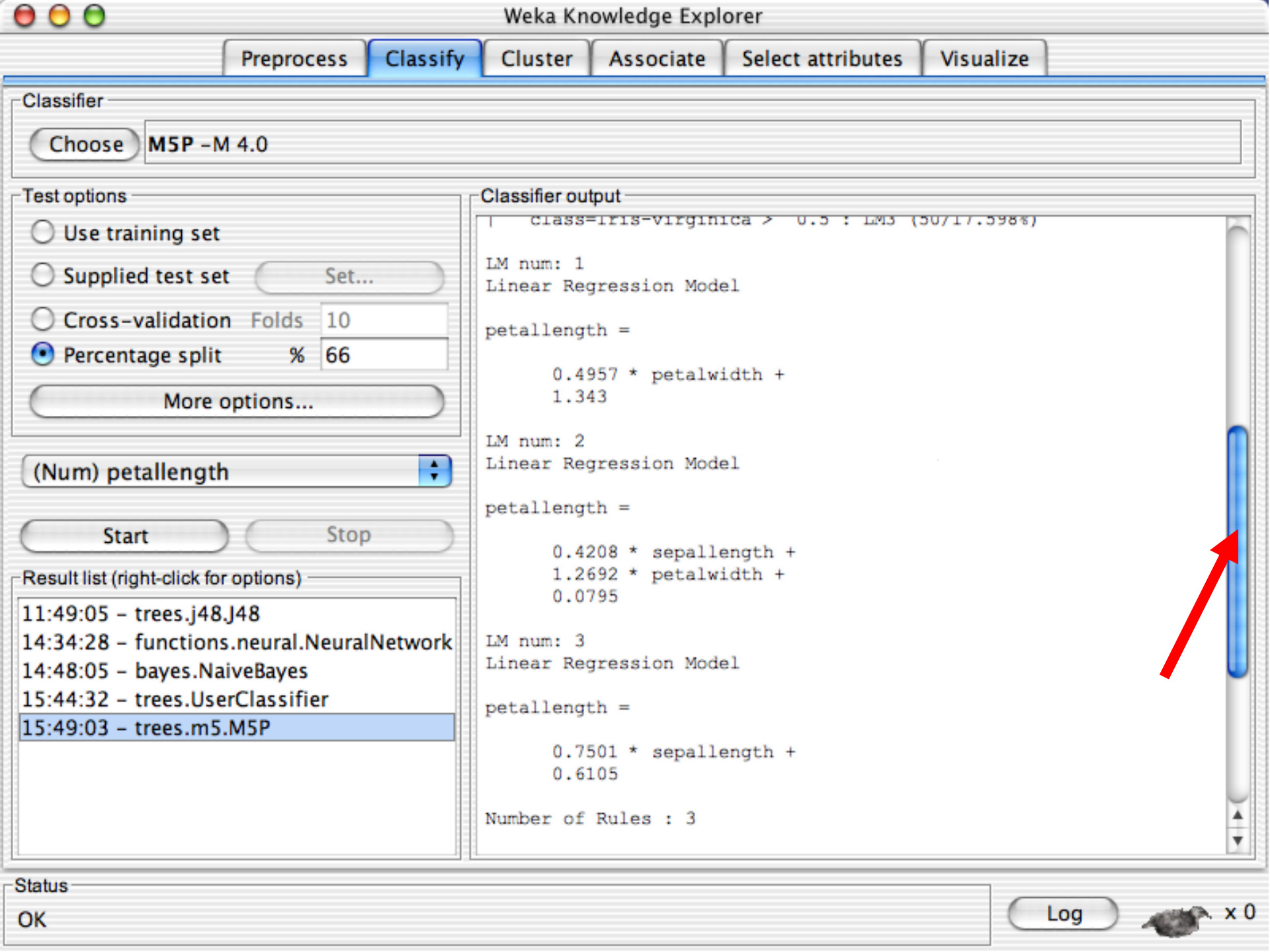
Status

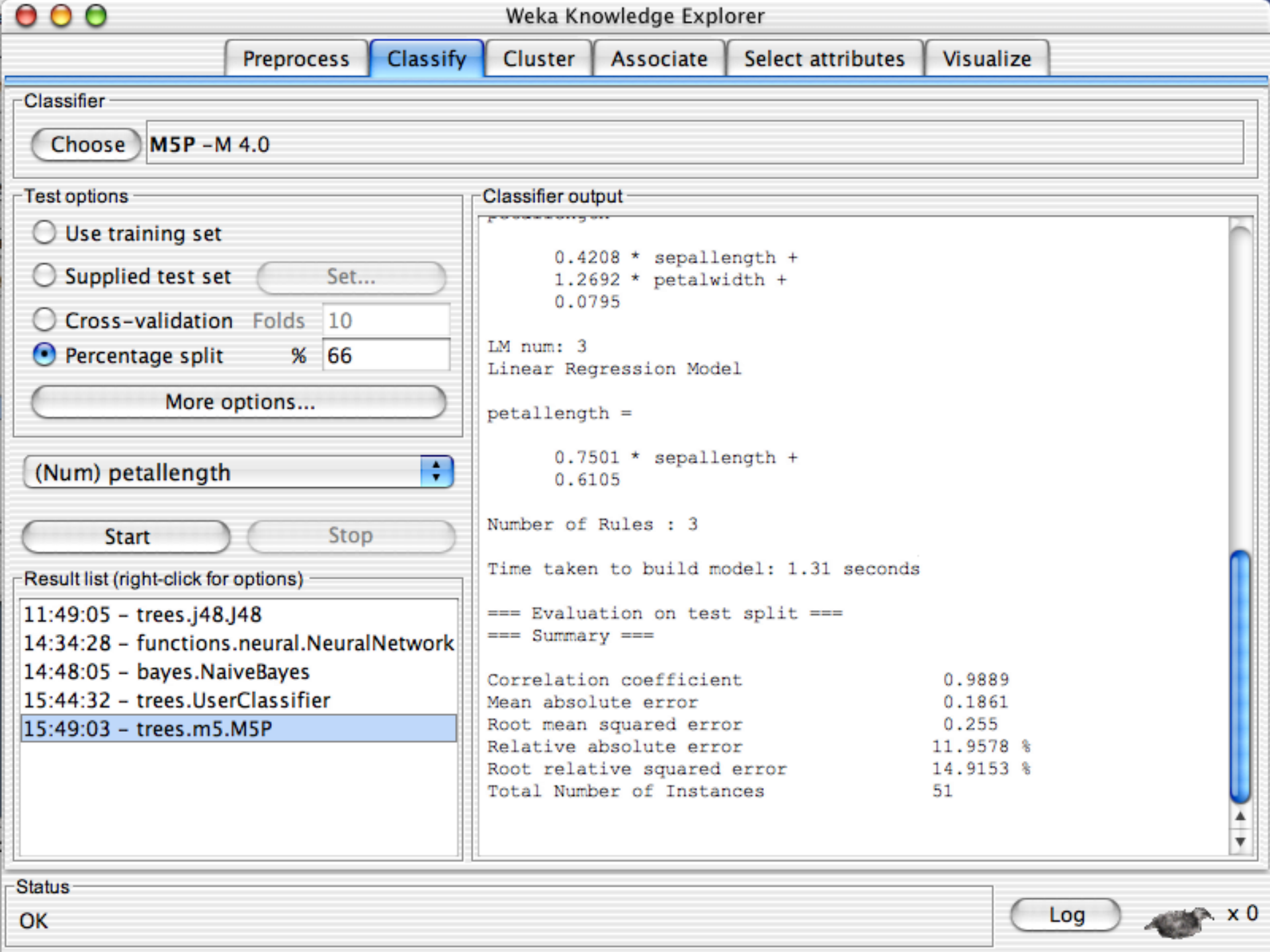
OK

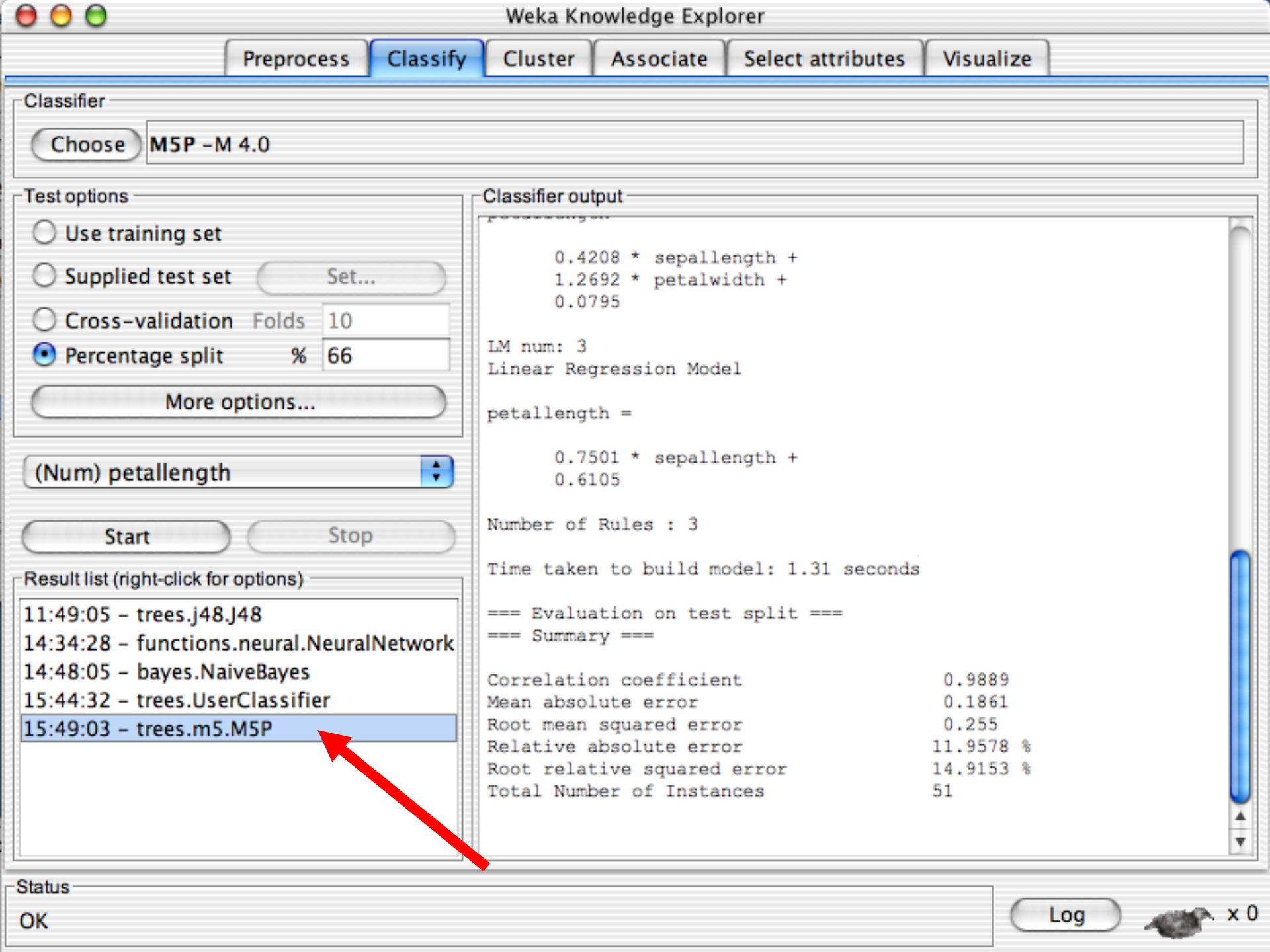
Log x 0

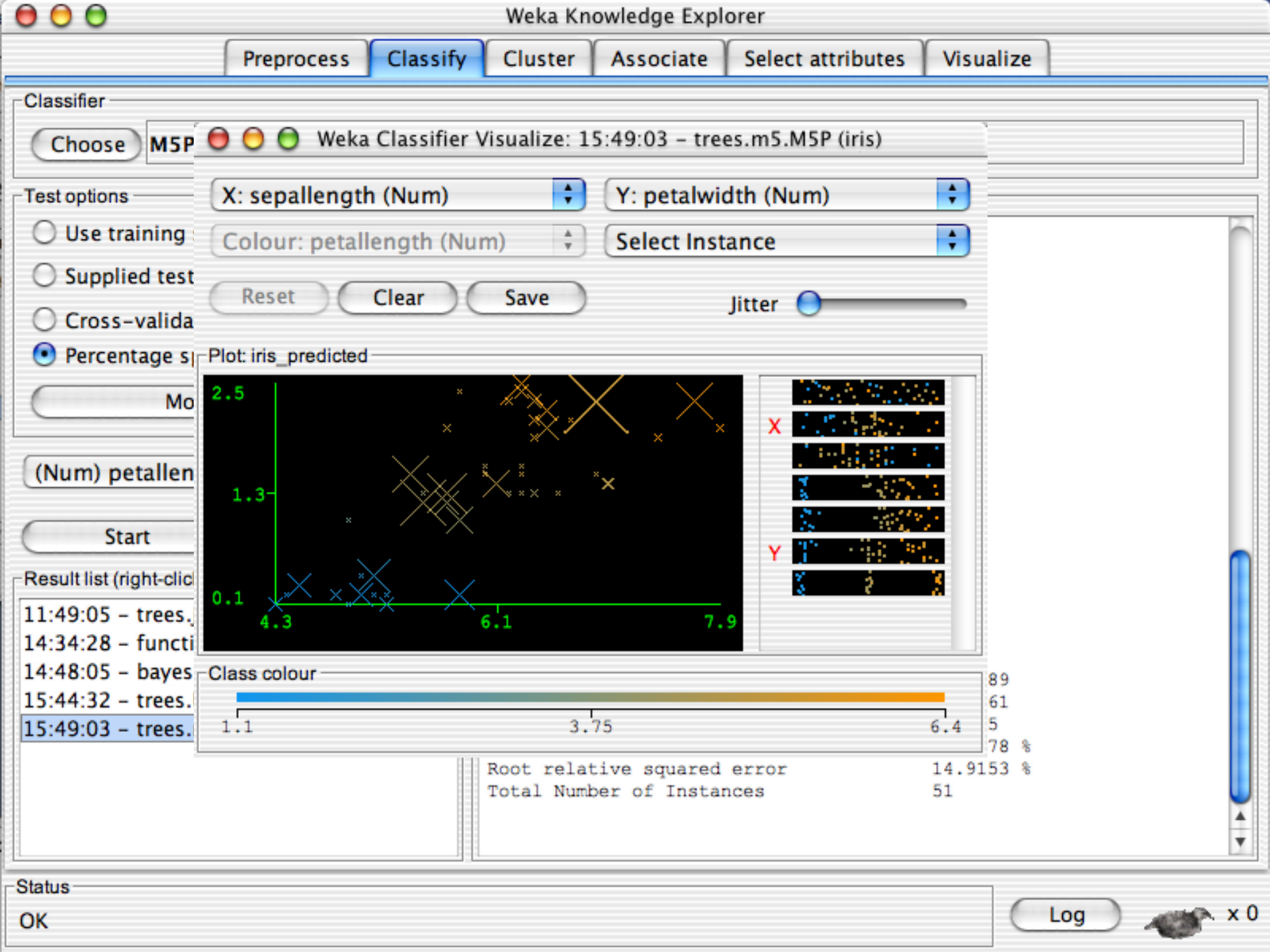

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
weka
classifiers
bayes
functions
lazy
meta
misc
trees
adtree
DecisionStump
Id3
j48
lmt
m5
M5P
RandomForest
RandomTree
REPTree
UserClassifier
rules
assifier output
== Evaluation on test split ===
== Summary ===
orrectly Classified Instances 49 96.0784 %
ncorrectly Classified Instances 2 3.9216 %
appa statistic 0.9408
ean absolute error 0.0319
oot mean squared error 0.1622
elative absolute error 7.1634 %
oot relative squared error 34.312 %
otal Number of Instances 51
== Detailed Accuracy By Class ===
P Rate FP Rate Precision Recall F-Measure Class
1 0 1 1 1 Iris-setosa
1 0.063 0.905 1 0.95 Iris-versicolor
0.882 0 1 0.882 0.938 Iris-virginica
== 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

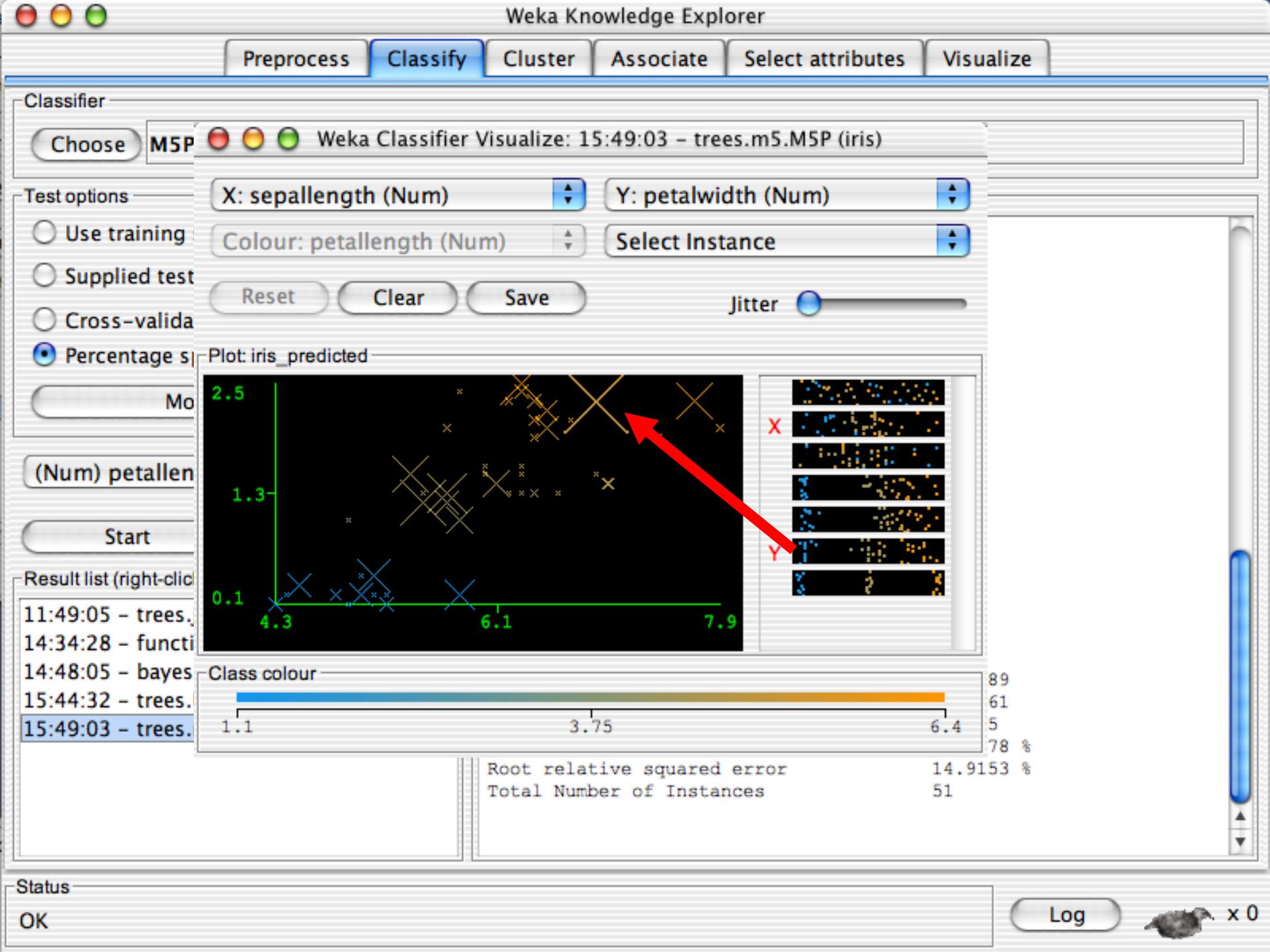

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P -M 4.0
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Num) petallength
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier
15:49:03 - trees.m5.M5P
Classifier output
=== Run information ===
Scheme: weka.classifiers.trees.m5.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 predictions)
petalwidth <= 0.8 : LM1 (50/10.469%)
petalwidth > 0.8 :
| class=Iris-virginica <= 0.5 : LM2 (50/14.325%)
| class=Iris-virginica > 0.5 : LM3 (50/17.598%)
LM num: 1
Linear Regression Model
petallength =
Status
OK
Log
x 0

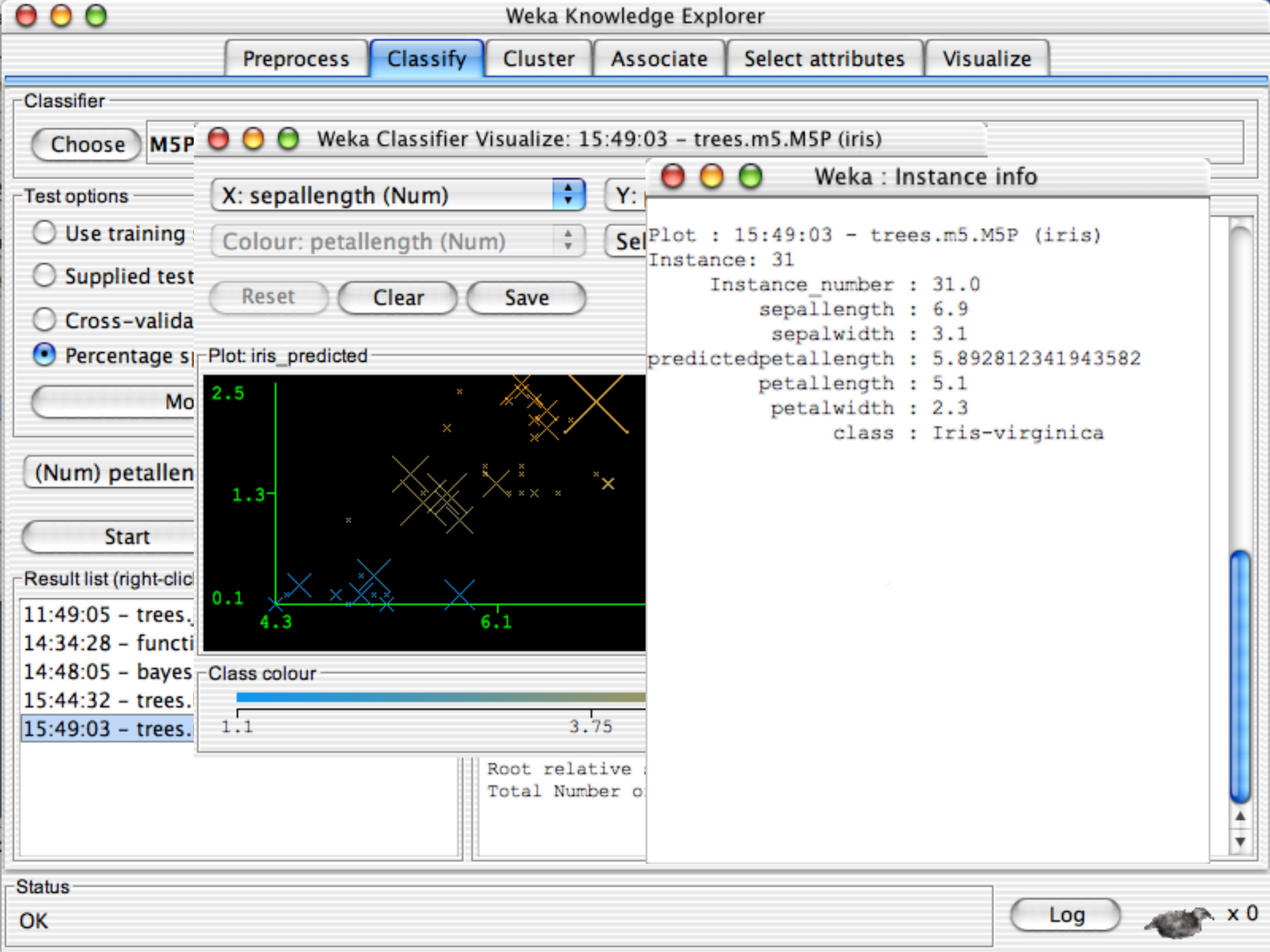

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P -M 4.0
Test options
Use training set
Supplied test set
Set...
Cross-validation
Folds
10
Percentage split
%
66
More options...
(Num) petallength
Start
Stop
Result list (right-click for options)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier
15:49:03 - trees.m5.M5P
Classifier output
LM num: 1
Linear Regression Model
petallength =
0.4957 * petalwidth +
1.343
LM num: 2
Linear Regression Model
petallength =
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795
LM num: 3
Linear Regression Model
petallength =
0.7501 * sepallength +
0.6105
Number of Rules : 3
Status
OK
Log
x 0


Weka Knowledge 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)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier
15:49:03 - trees.m5.M5P
Classifier output
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795
LM num: 3
Linear Regression Model
petallength =
0.7501 * sepallength +
0.6105
Number of Rules : 3
Time taken to build model: 1.31 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 Knowledge 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)
11:49:05 - trees.j48.J48
14:34:28 - functions.neural.NeuralNetwork
14:48:05 - bayes.NaiveBayes
15:44:32 - trees.UserClassifier
15:49:03 - trees.m5.M5P
Classifier output
0.4208 * sepallength +
1.2692 * petalwidth +
0.0795
LM num: 3
Linear Regression Model
petallength =
0.7501 * sepallength +
0.6105
Number of Rules : 3
Time taken to build model: 1.31 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

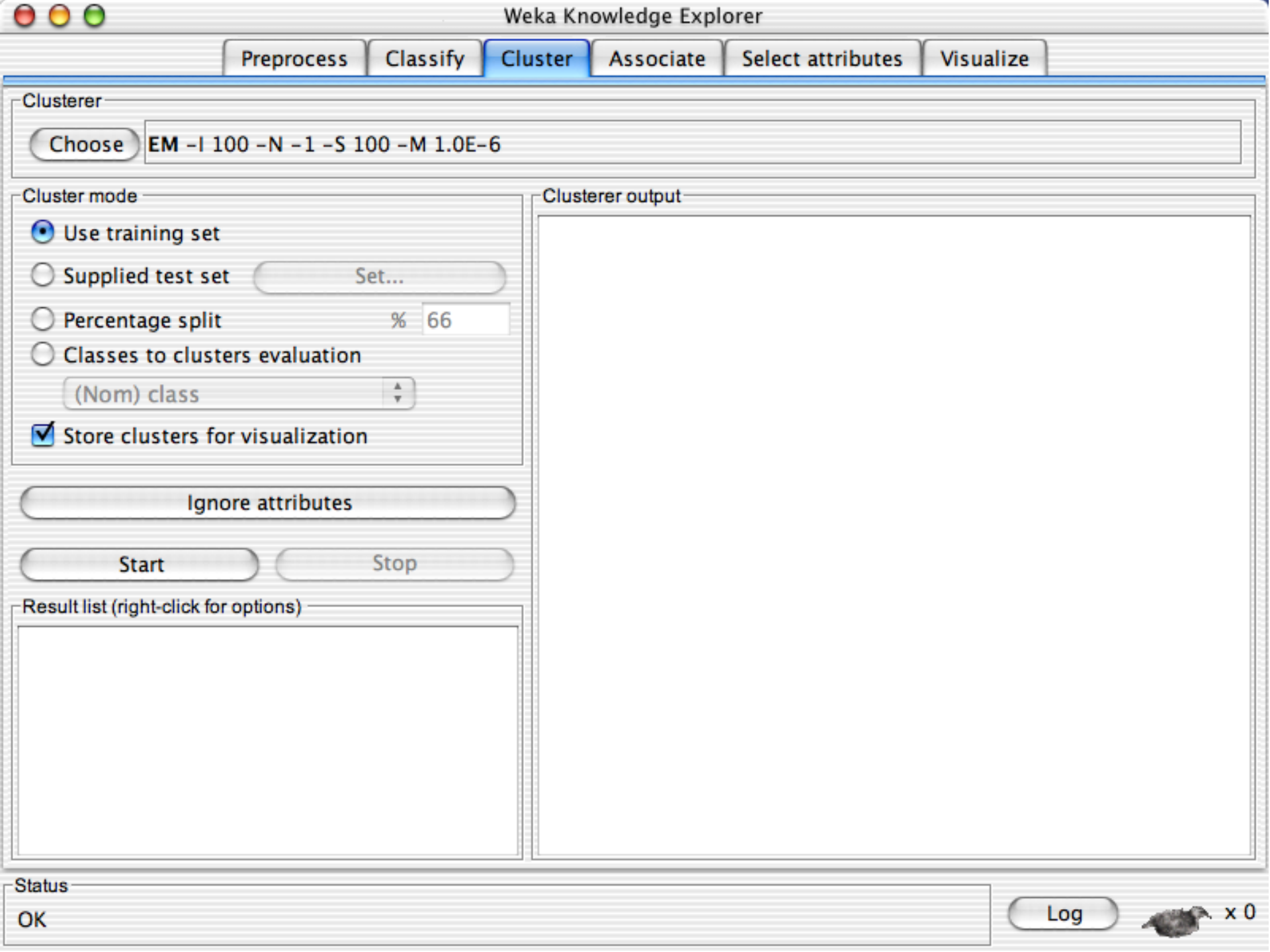

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P
Weka Classifier Visualize: 15:49:03 - trees.m5.M5P (iris)
X: sepallength (Num)
Y: petalwidth (Num)
Colour: petallength (Num)
Select Instance
Reset
Clear
Save
Jitter
Test options
Use training
Supplied test
Cross-valida
Percentage s
Mo
(Num) petallen
Start
Result list (right-clic
11:49:05 - trees.
14:34:28 - functi
14:48:05 - bayes
15:44:32 - trees.
15:49:03 - trees.
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
89
61
5
78 %
Root relative squared error 14.9153 %
Total Number of Instances 51
Status
OK
Log
x 0

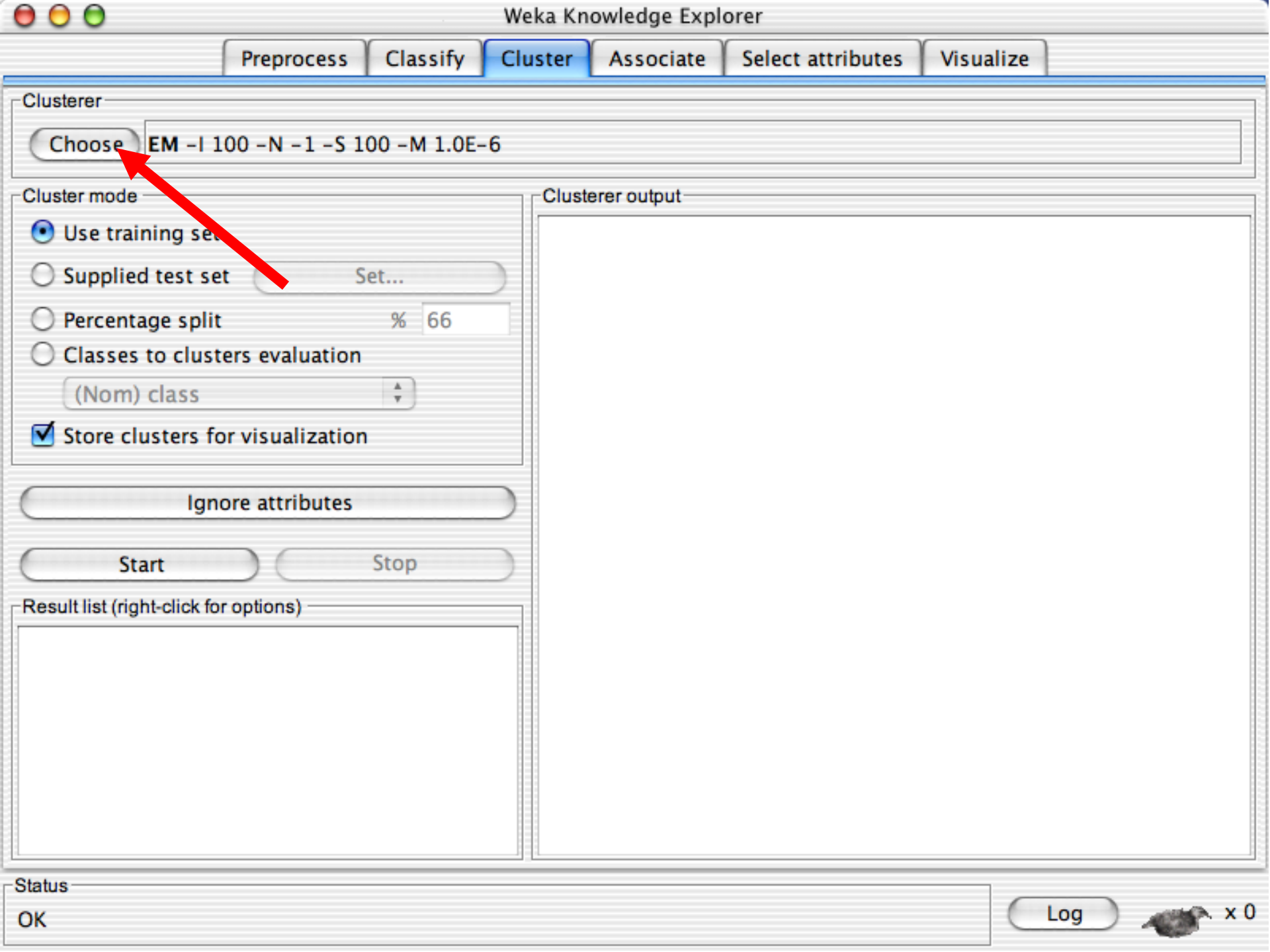

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P
Weka Classifier Visualize: 15:49:03 - trees.m5.M5P (iris)
X: sepallength (Num)
Y: petalwidth (Num)
Colour: petallength (Num)
Select Instance
Reset
Clear
Save
Jitter
Test options
Use training
Supplied test
Cross-valida
Percentage s
Mo
(Num) petallen
Start
Result list (right-clic
11:49:05 - trees.
14:34:28 - functi
14:48:05 - bayes
15:44:32 - trees.
15:49:03 - trees.
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
89
61
5
78 %
Root relative squared error 14.9153 %
Total Number of Instances 51
Status
OK
Log
x 0

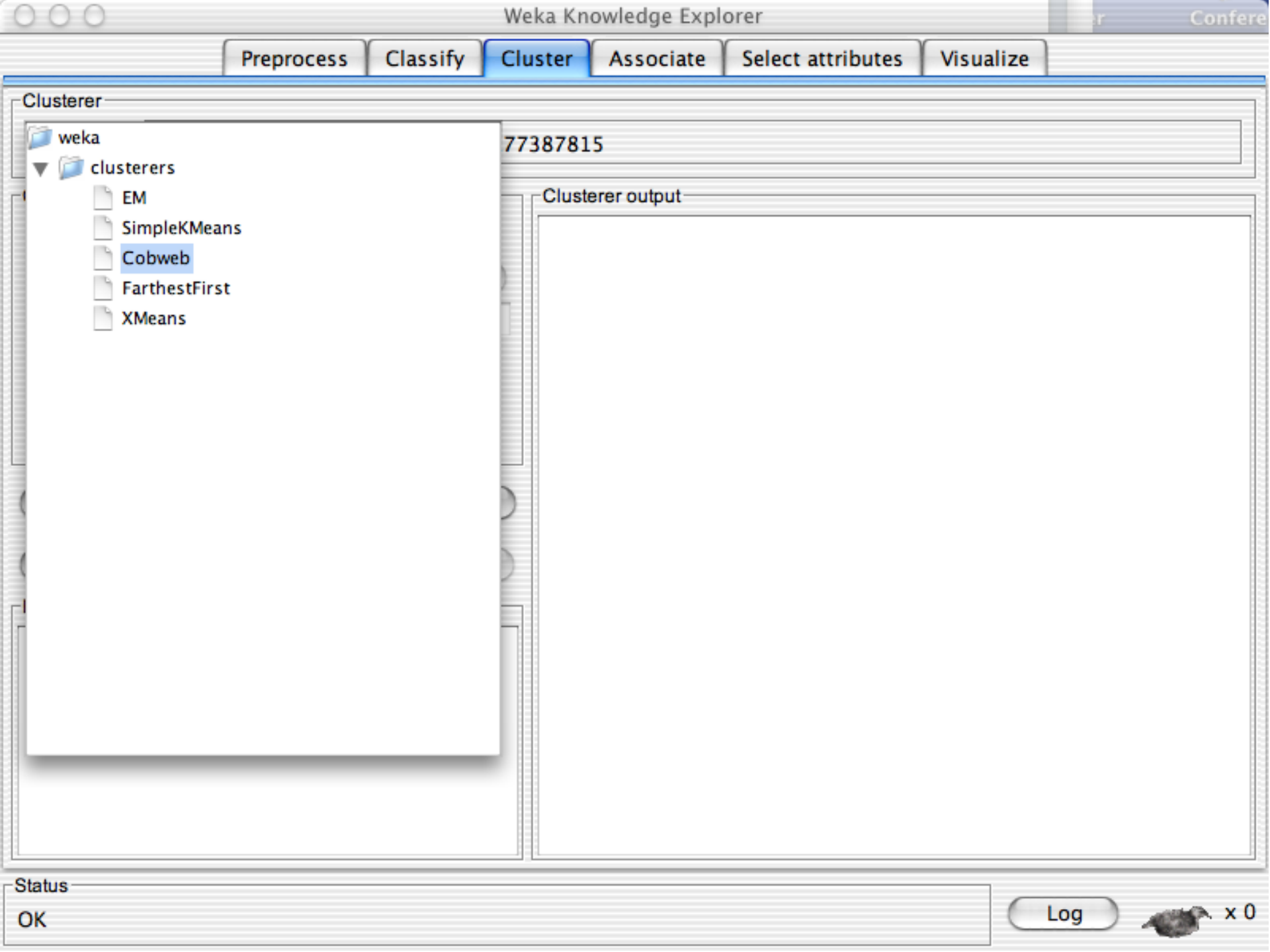

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Classifier
Choose
M5P
Weka Classifier Visualize: 15:49:03 - trees.m5.M5P (iris)
Weka : Instance info
Test options
Use training
Supplied test
Cross-valida
Percentage s
Mo
X: sepallength (Num)
Y:
Colour: petallength (Num)
Sel
Reset
Clear
Save
Plot: iris_predicted
2.5
1.3
0.1
4.3
6.1
(Num) petallen
Start
Result list (right-clic
11:49:05 - trees.
14:34:28 - functi
14:48:05 - bayes
15:44:32 - trees.
15:49:03 - trees.
Class colour
1.1
3.75
Root relative
Total Number o
Plot : 15:49:03 - trees.m5.M5P (iris)
Instance: 31
Instance_number : 31.0
sepallength : 6.9
sepalwidth : 3.1
predictedpetallength : 5.892812341943582
petallength : 5.1
petalwidth : 2.3
class : Iris-virginica
Status
OK
Log
x 0

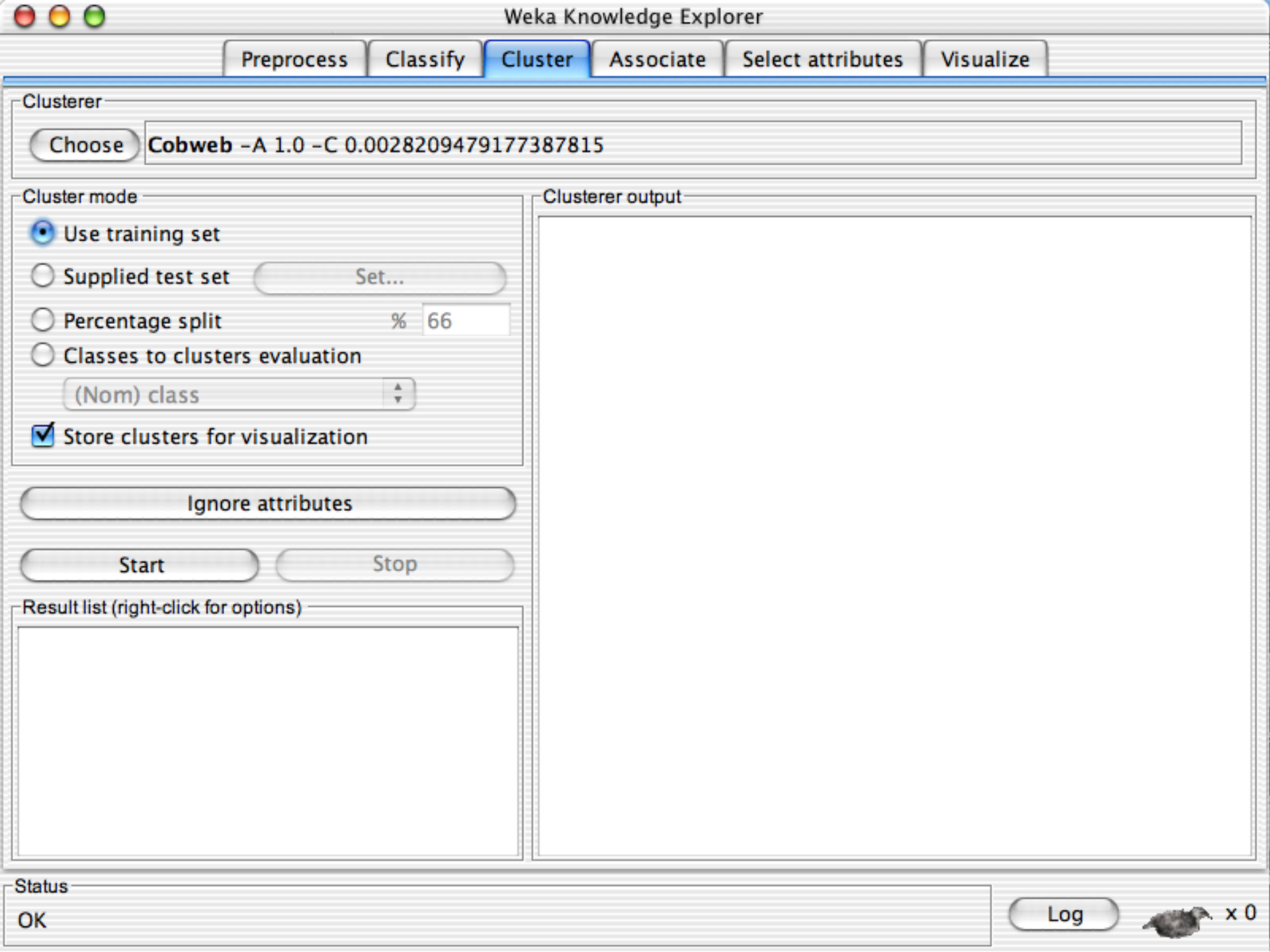
# Explorer: clustering data

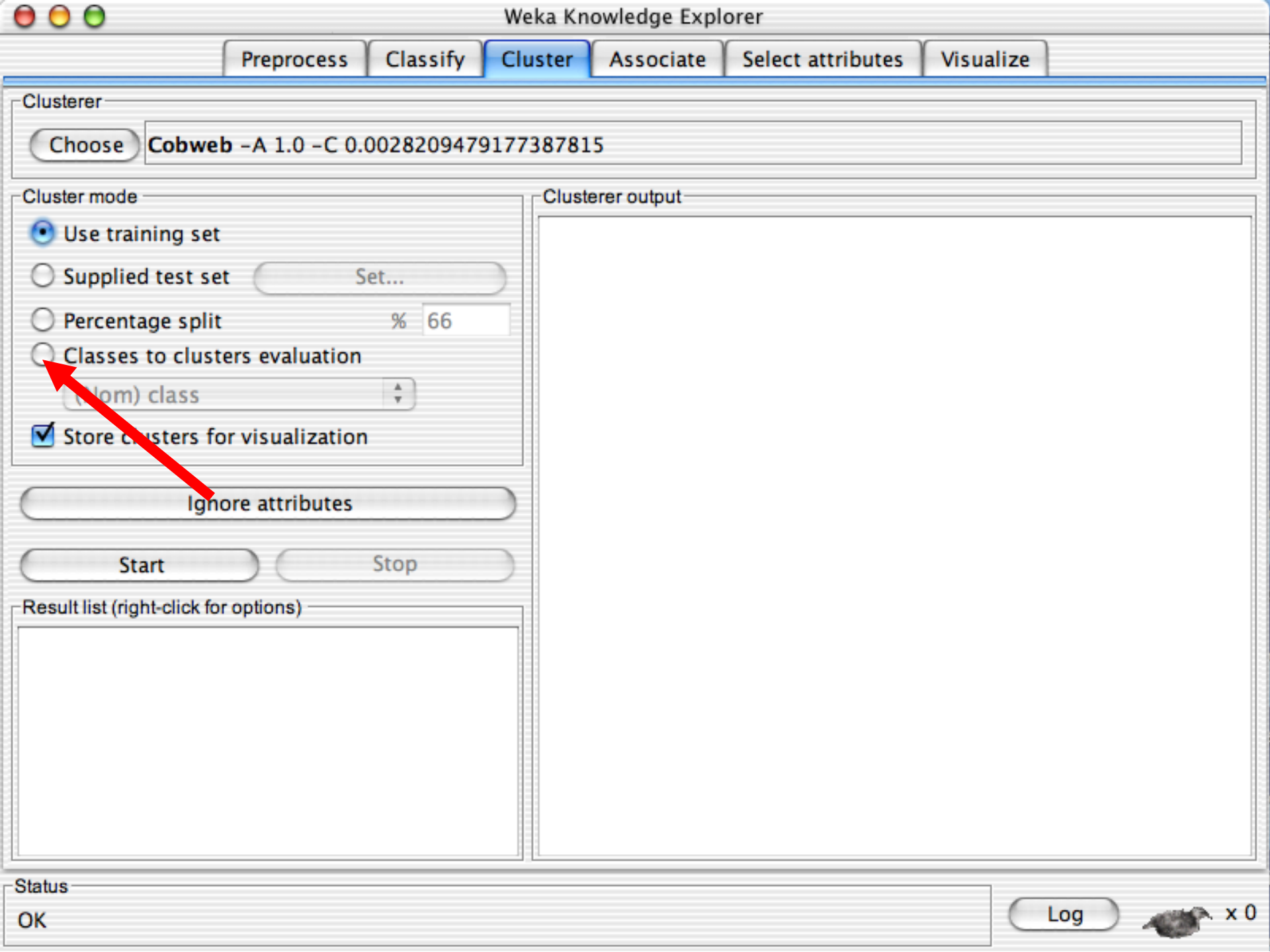
- WEKA contains “clusterers” for finding groups of similar instances in a dataset
- Implemented schemes are:
  - *k*-Means, EM, Cobweb, *X*-means, FarthestFirst
- Clusters can be visualized and compared to “true” clusters (if given)
- Evaluation based on loglikelihood if clustering scheme produces a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

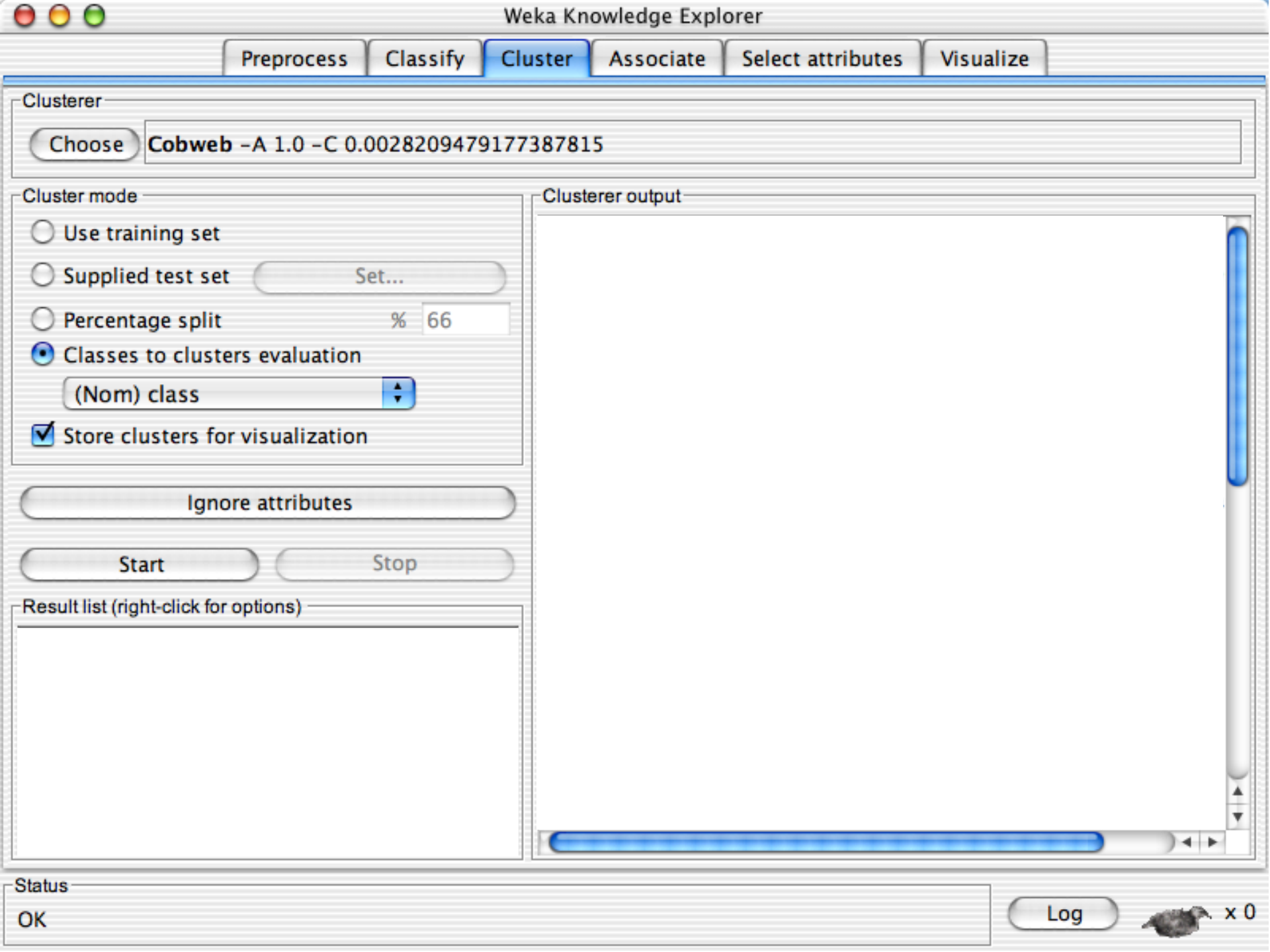

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -S 100 -M 1.0E-6
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes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

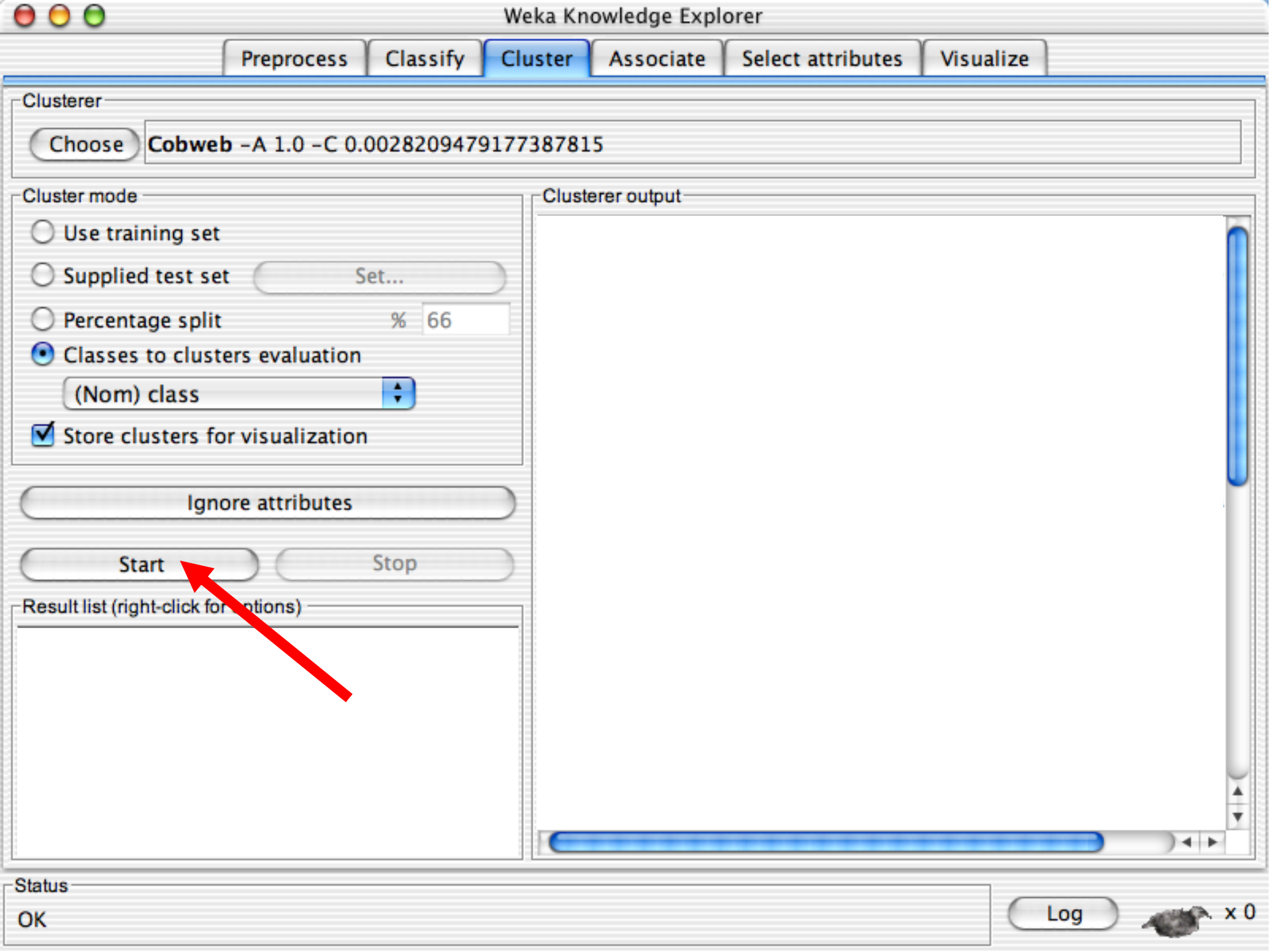

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
EM -I 100 -N -1 -S 100 -M 1.0E-6
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes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

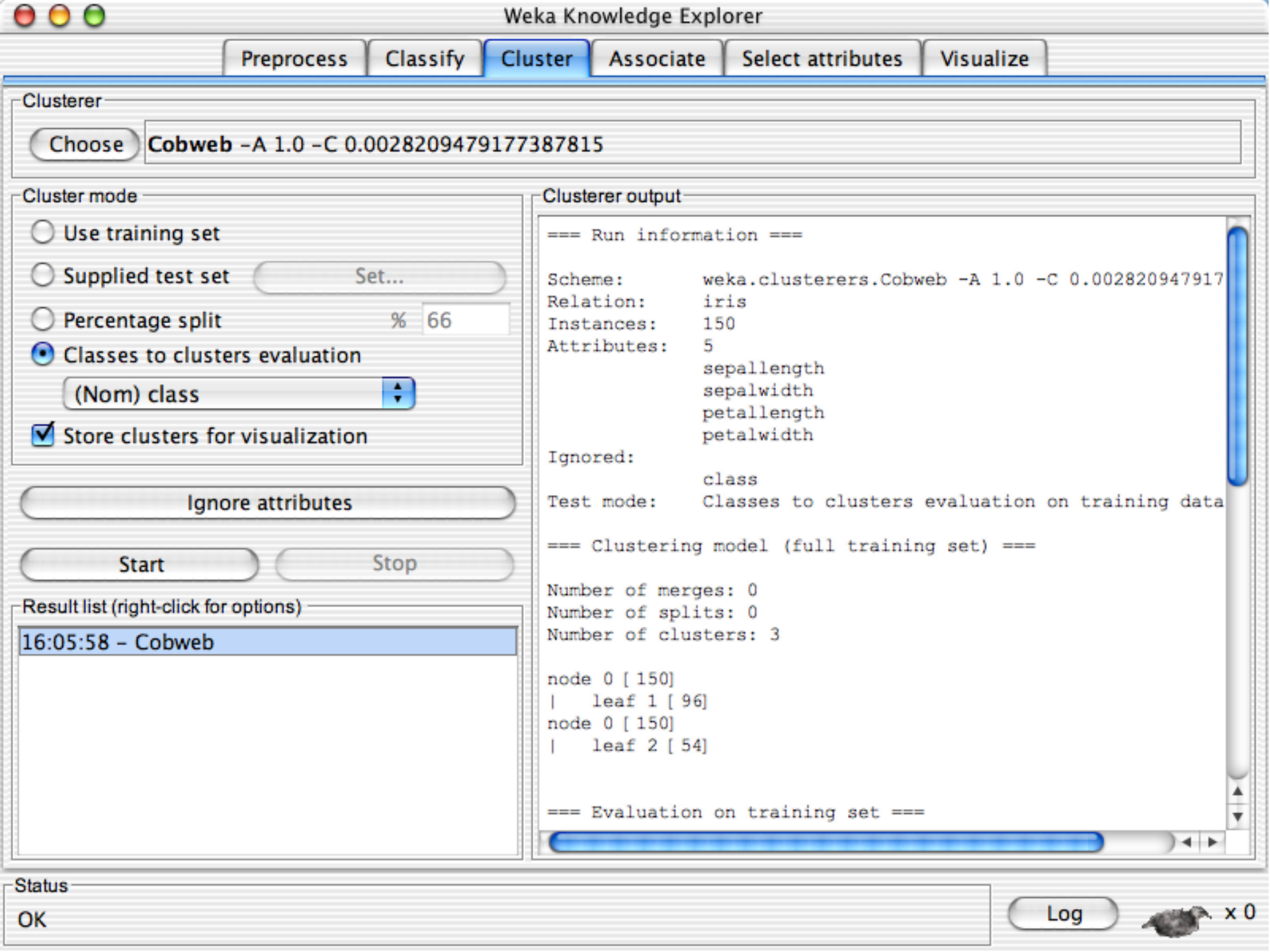

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
77387815
weka
clusterers
EM
SimpleKMeans
Cobweb
FarthestFirst
XMeans
Clusterer output
Status
OK
Log
x 0

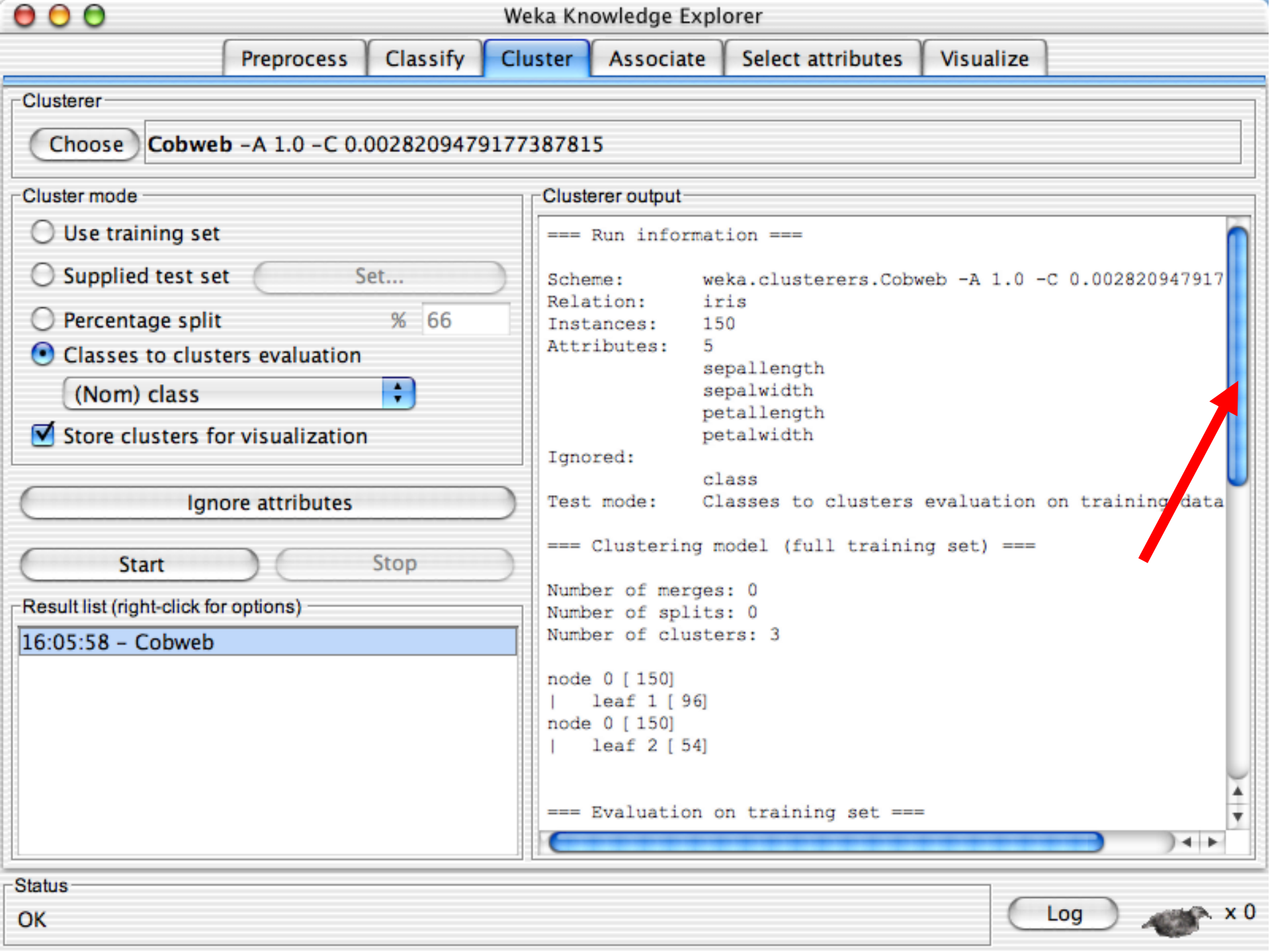

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
%
66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore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


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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 Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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 Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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 Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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)
16:05:58 - Cobweb
Clusterer output
=== Run information ===
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
Ignored:
class
Test mode: Classes to clusters evaluation on training data
=== Clustering model (full 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]
=== Evaluation on training set ===
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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)
16:05:58 - Cobweb
Clusterer output
=== Run information ===
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
Ignored:
class
Test mode: Classes to clusters evaluation on training data
=== Clustering model (full 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]
=== Evaluation on training set ===
Status
OK
Log
x 0

Weka Knowledge Explorer

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

Clusterer

Choose **Cobweb** -A 1.0 -C 0.0028209479177387815

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)

16:05:58 - Cobweb

Clusterer output

```
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 Knowledge Explorer

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

Clusterer

Choose **Cobweb** -A 1.0 -C 0.0028209479177387815

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)

16:05:58 - Cobweb

Clusterer output

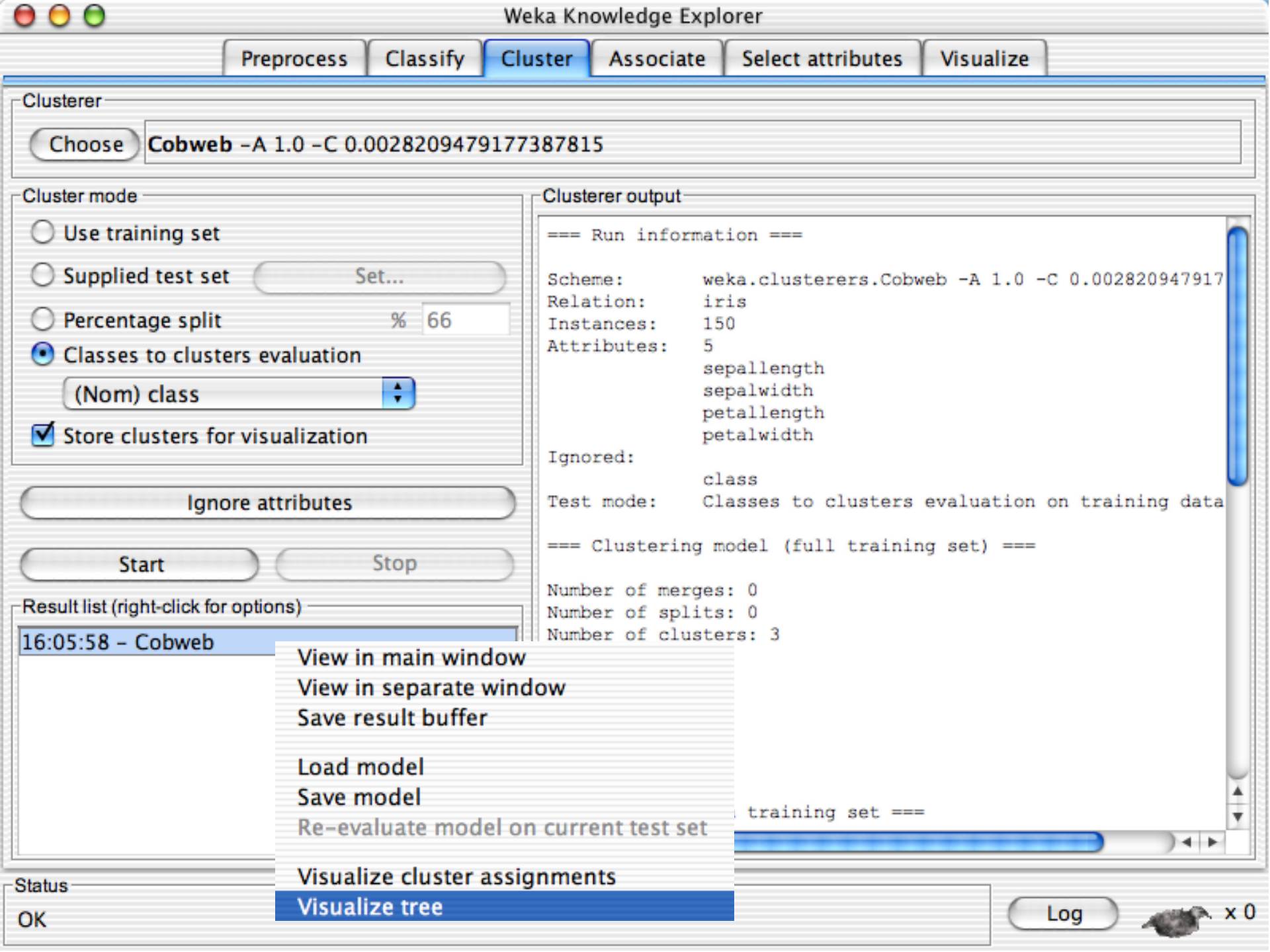
```
Number of clusters: 3

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 %
```

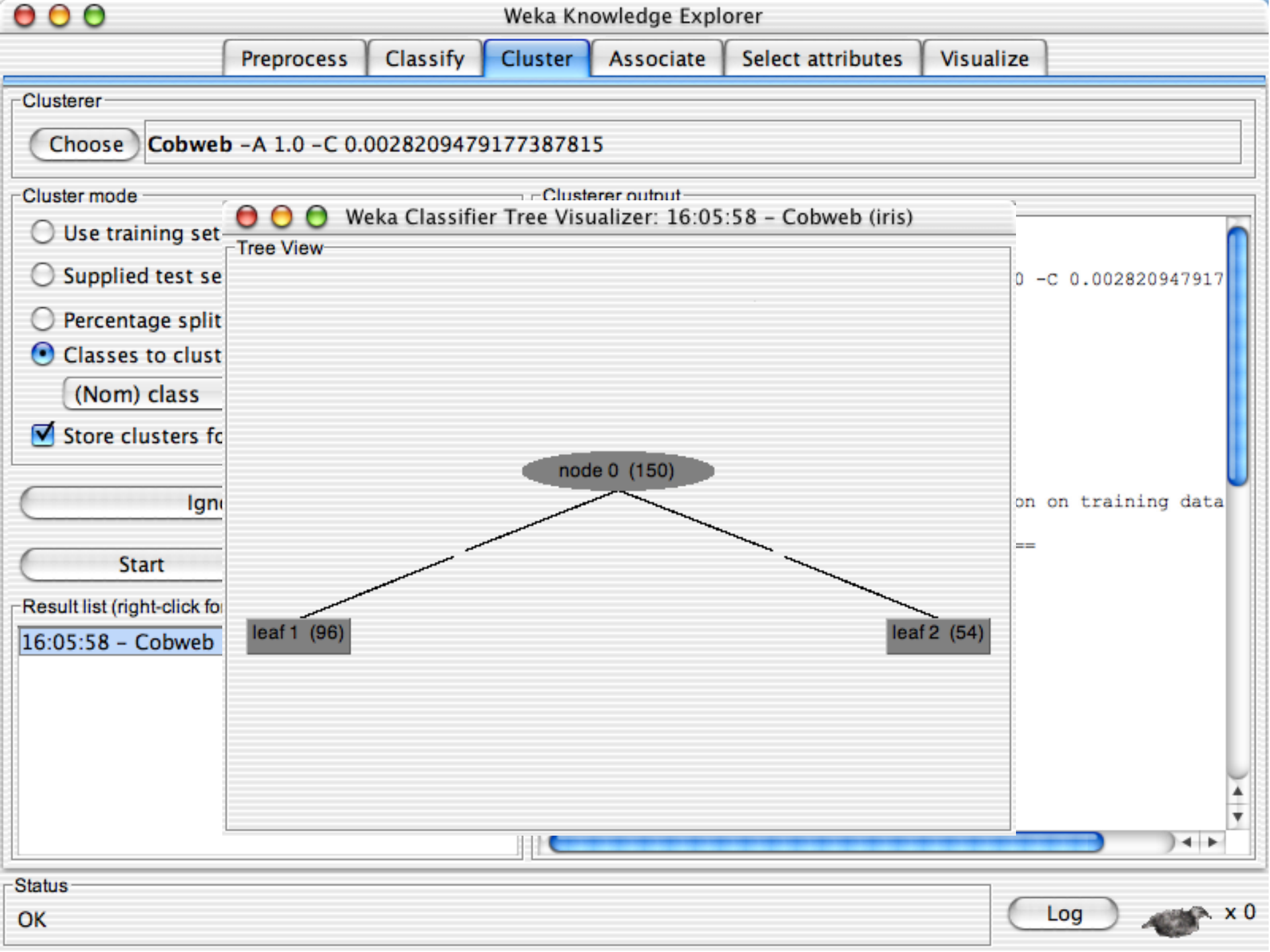
Status

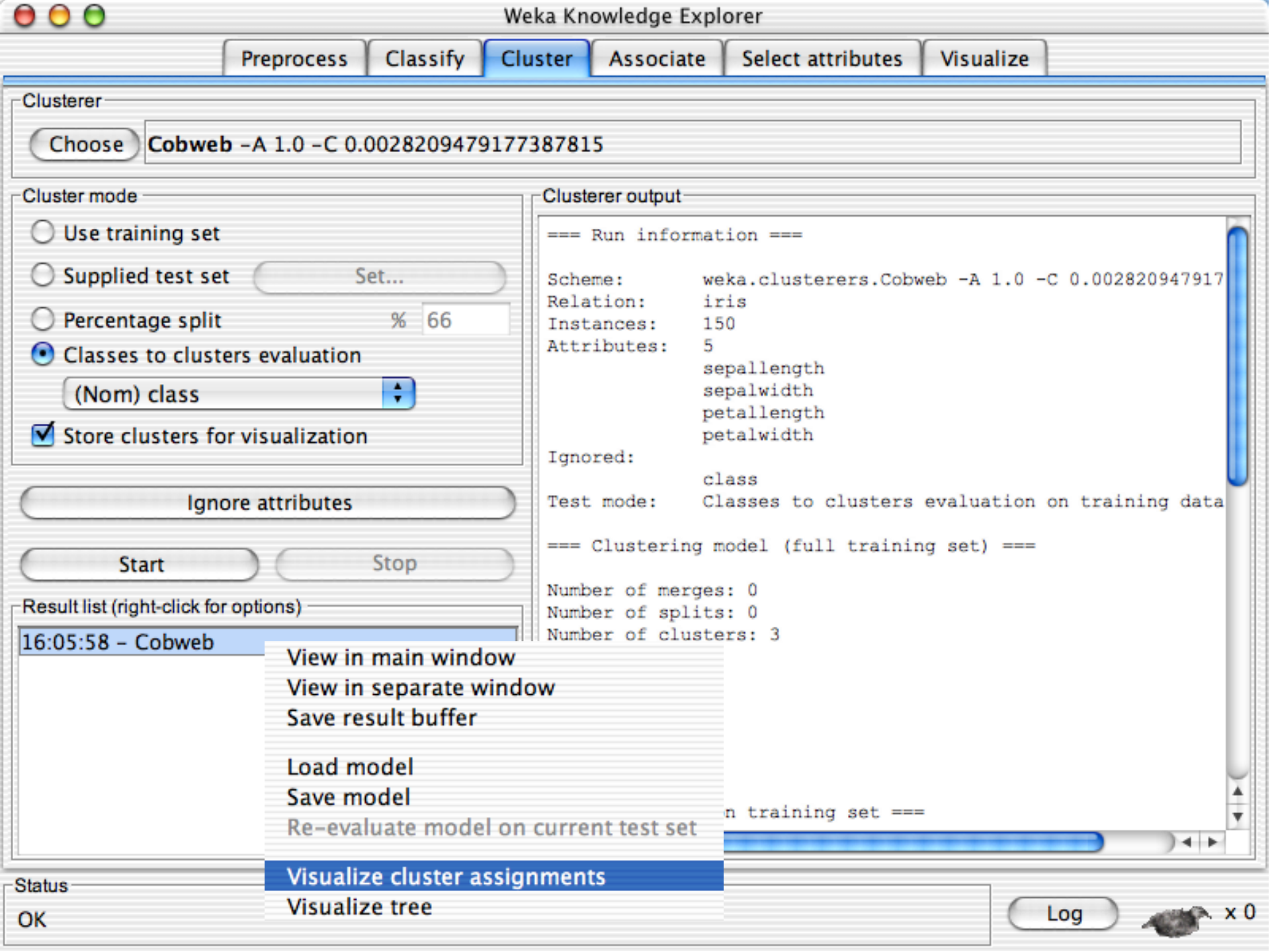
OK

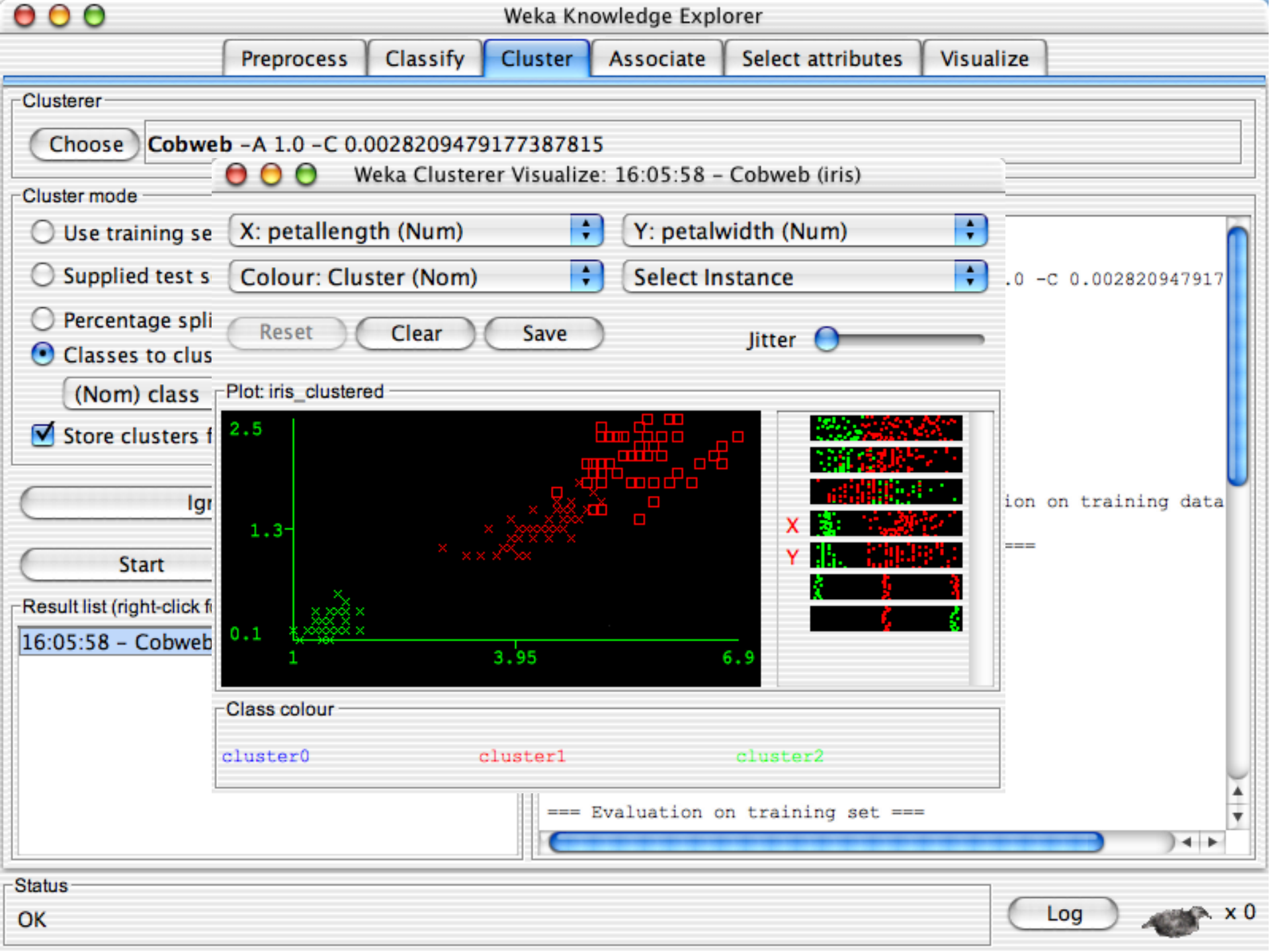
Log

x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
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)
16:05:58 - Cobweb
Clusterer output
=== Run information ===
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
Ignored:
class
Test mode: Classes to clusters evaluation on training data
=== Clustering model (full training set) ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
training set ===
View in main window
View in separate window
Save result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize cluster assignments
Visualize tree
Status
OK
Log
x 0

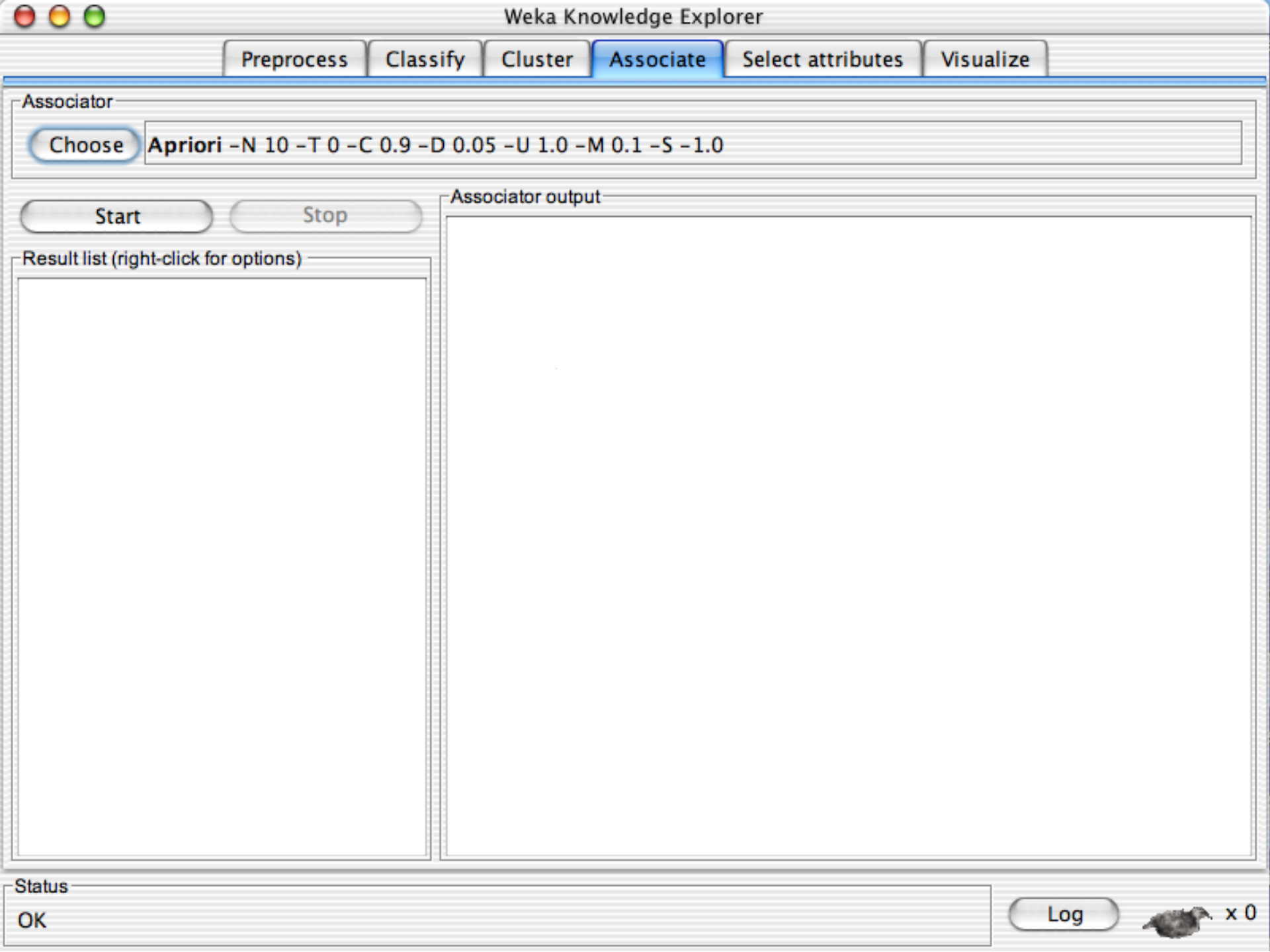

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
Cluster mode
Use training set
Supplied test se
Percentage split
Classes to clust
(Nom) class
Store clusters fo
Ignore
Start
Result list (right-click for
16:05:58 - Cobweb
Clusterer output
0 -C 0.002820947917
on on training data
==
Weka Classifier Tree Visualizer: 16:05:58 - Cobweb (iris)
Tree View
node 0 (150)
leaf 1 (96)
leaf 2 (54)
Status
OK
Log
x 0

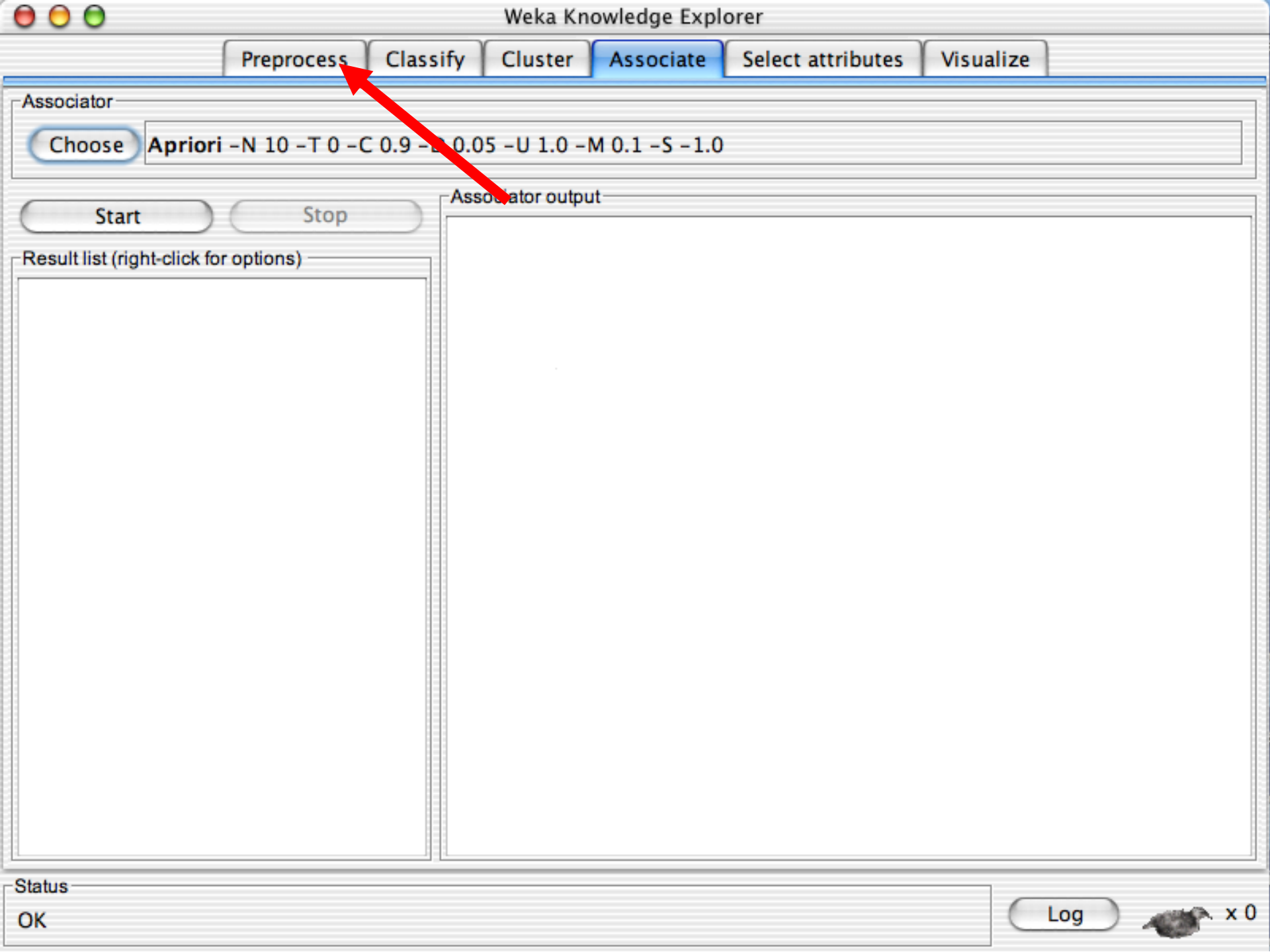

Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
Cluster mode
Use training set
Supplied test set
Set...
Percentage split
% 66
Classes to clusters evaluation
(Nom) class
Store clusters for visualization
Ignore attributes
Start
Stop
Result list (right-click for options)
16:05:58 - Cobweb
Clusterer output
=== Run information ===
Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.002820947917
Relation: iris
Instances: 150
Attributes: 5
sepallength
sepalwidth
petallength
petalwidth
Ignored:
class
Test mode: Classes to clusters evaluation on training data
=== Clustering model (full training set) ===
Number of merges: 0
Number of splits: 0
Number of clusters: 3
n training set ===
View in main window
View in separate window
Save result buffer
Load model
Save model
Re-evaluate model on current test set
Visualize cluster assignments
Visualize tree
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Clusterer
Choose
Cobweb -A 1.0 -C 0.0028209479177387815
Weka Clusterer Visualize: 16:05:58 - Cobweb (iris)
Cluster mode
Use training se
Supplied test s
Percentage spli
Classes to clus
(Nom) class
Store clusters f
X: petallength (Num)
Y: petalwidth (Num)
Colour: Cluster (Nom)
Select Instance
Reset
Clear
Save
Jitter
Plot: iris_clustered
2.5
1.3
0.1
1
3.95
6.9
X
Y
Class colour
cluster0
cluster1
cluster2
Start
Result list (right-click f
16:05:58 - Cobweb
.0 -C 0.002820947917
ion on training data
===
=== Evaluation on training set ===
Status
OK
Log
x 0

# Explorer: finding associations

- WEKA contains an implementation of the Apriori algorithm 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 Knowledge 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
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0


Weka Knowledge 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
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0

Weka Knowledge Explorer

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

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

Filter

Choose **None** | Apply

Current relation

Relation: vote
Instances: 435 Attributes: 17

Attributes

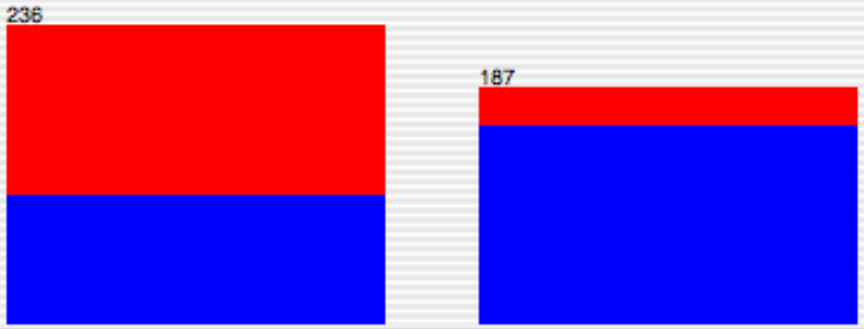
| No. | Name |
|---|---|
| 1 | handicapped-infants |
| 2 | water-project-cost-sharing |
| 3 | adoption-of-the-budget-resolution |
| 4 | physician-fee-freeze |
| 5 | el-salvador-aid |
| 6 | religious-groups-in-schools |
| 7 | anti-satellite-test-ban |
| 8 | aid-to-nicaraguan-contras |
| 9 | mx-missile |
| 10 | immigration |
| 11 | synfuels-corporation-cutback |
| 12 | education-spending |
| 13 | superfund-right-to-sue |
| 14 | crime |
| 15 | duty-free-exports |
| 16 | export-administration-act-south-africa |
| 17 | Class |

Selected attribute

Name: handicapped-infants Type: Nominal
Missing: 12 (3%) Distinct: 2 Unique: 0 (0%)

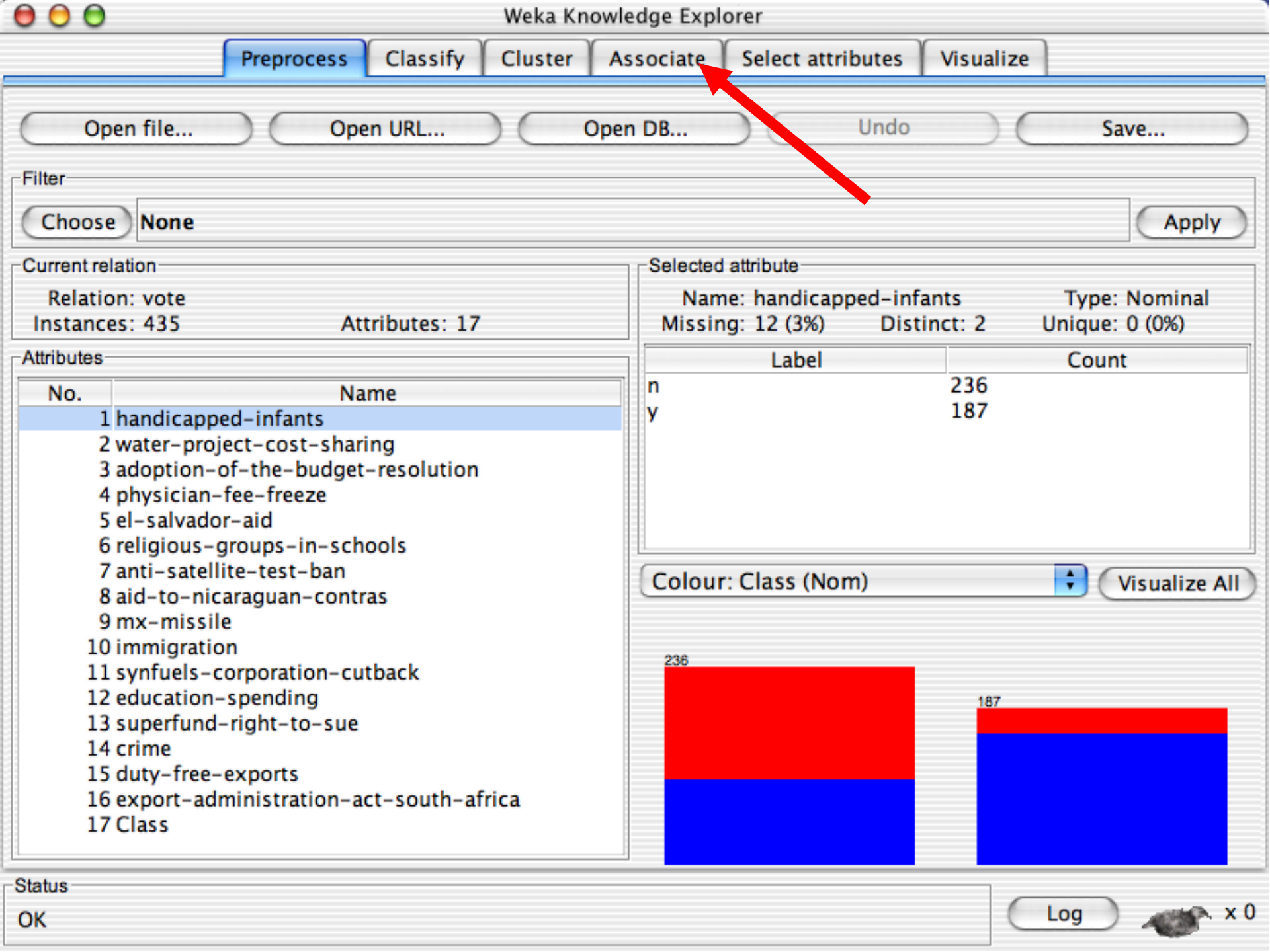
| Label | Count |
|---|---|
| n | 236 |
| y | 187 |

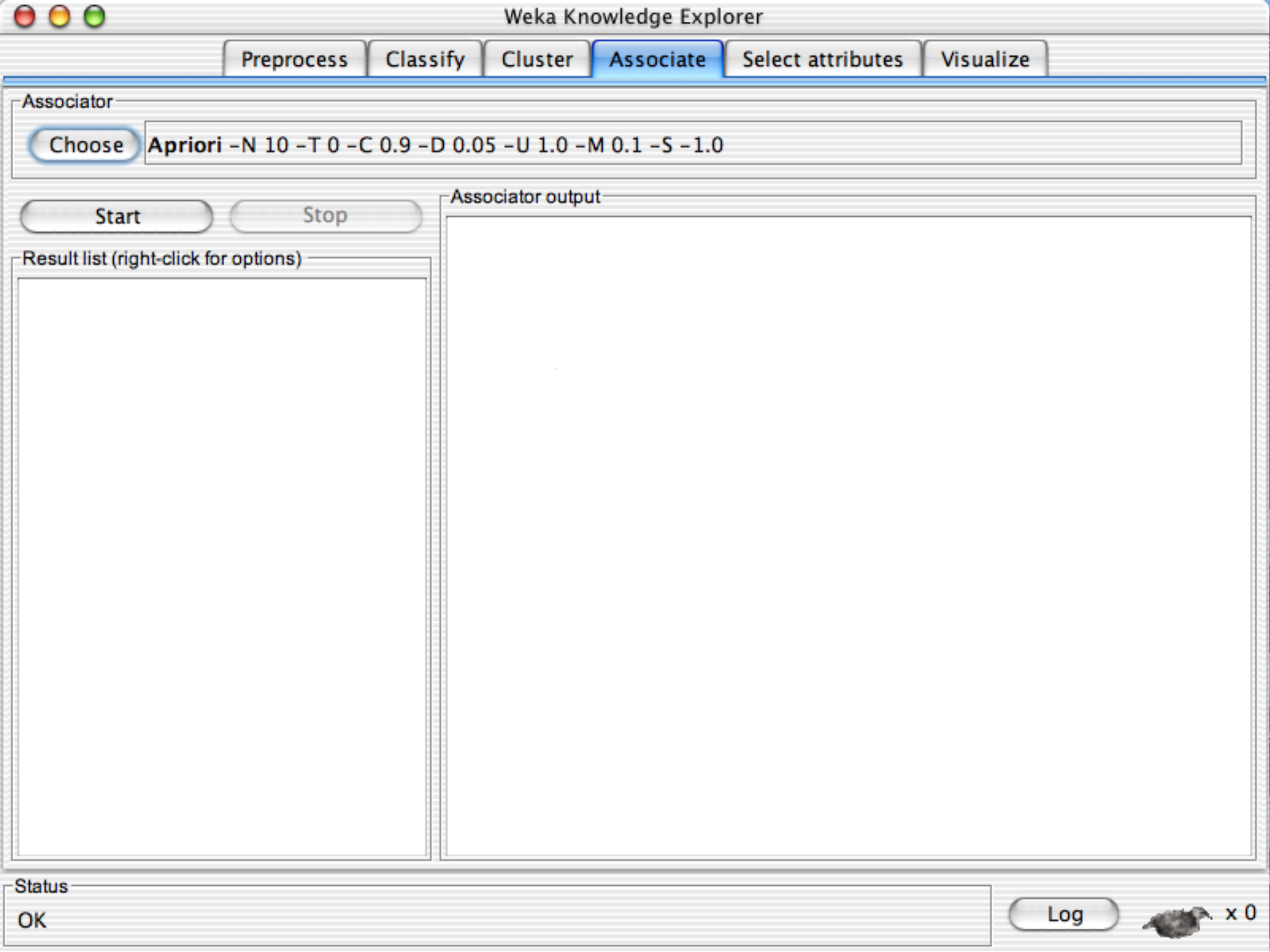
Colour: Class (Nom) | Visualize All

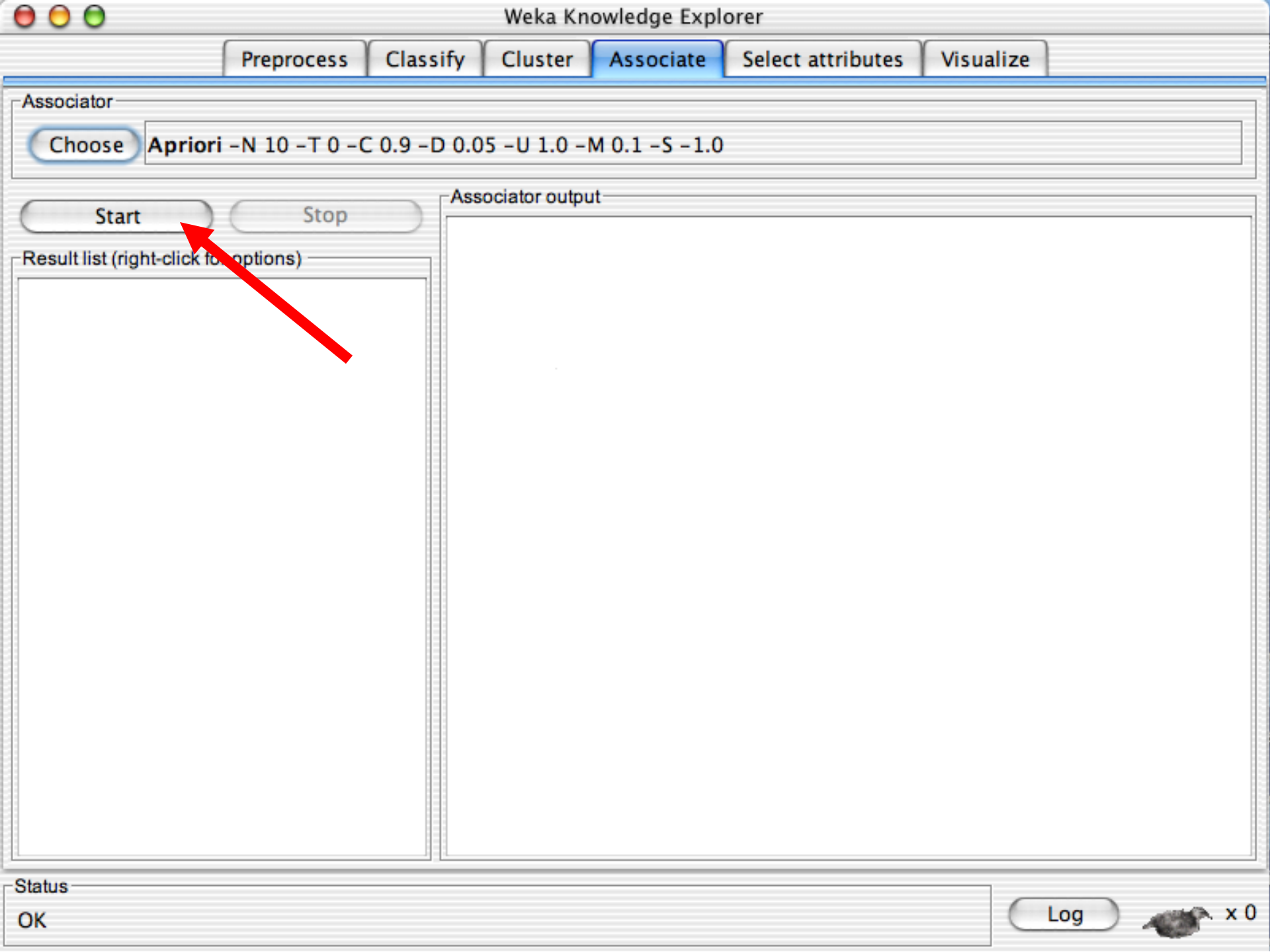



Status

OK | Log | x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: vote
Instances: 435
Attributes: 17
Attributes
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
12 education-spending
13 superfund-right-to-sue
14 crime
15 duty-free-exports
16 export-administration-act-south-africa
17 Class
Selected attribute
Name: handicapped-infants
Type: Nominal
Missing: 12 (3%)
Distinct: 2
Unique: 0 (0%)
Label
Count
n
236
y
187
Colour: Class (Nom)
Visualize All
236
187
Status
OK
Log
x 0


Weka Knowledge 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
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0


Weka Knowledge 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
Start
Stop
Result list (right-click for options)
Associator output
Status
OK
Log
x 0

Weka Knowledge 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

Start | Stop

Result list (right-click for optic

16:29:37 - Apriori

Associator output

```
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=democrat
 2. adoption-of-the-budget-resolution=y physician-fee-freeze=n aid-to-nicaraguan-cont
 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    conf
 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 2
 9. el-salvador-aid=n aid-to-nicaraguan-contras=y 204 ==> Class=democrat 197    conf
10. aid-to-nicaraguan-contras=y Class=democrat 218 ==> physician-fee-freeze=n 210
```

Status

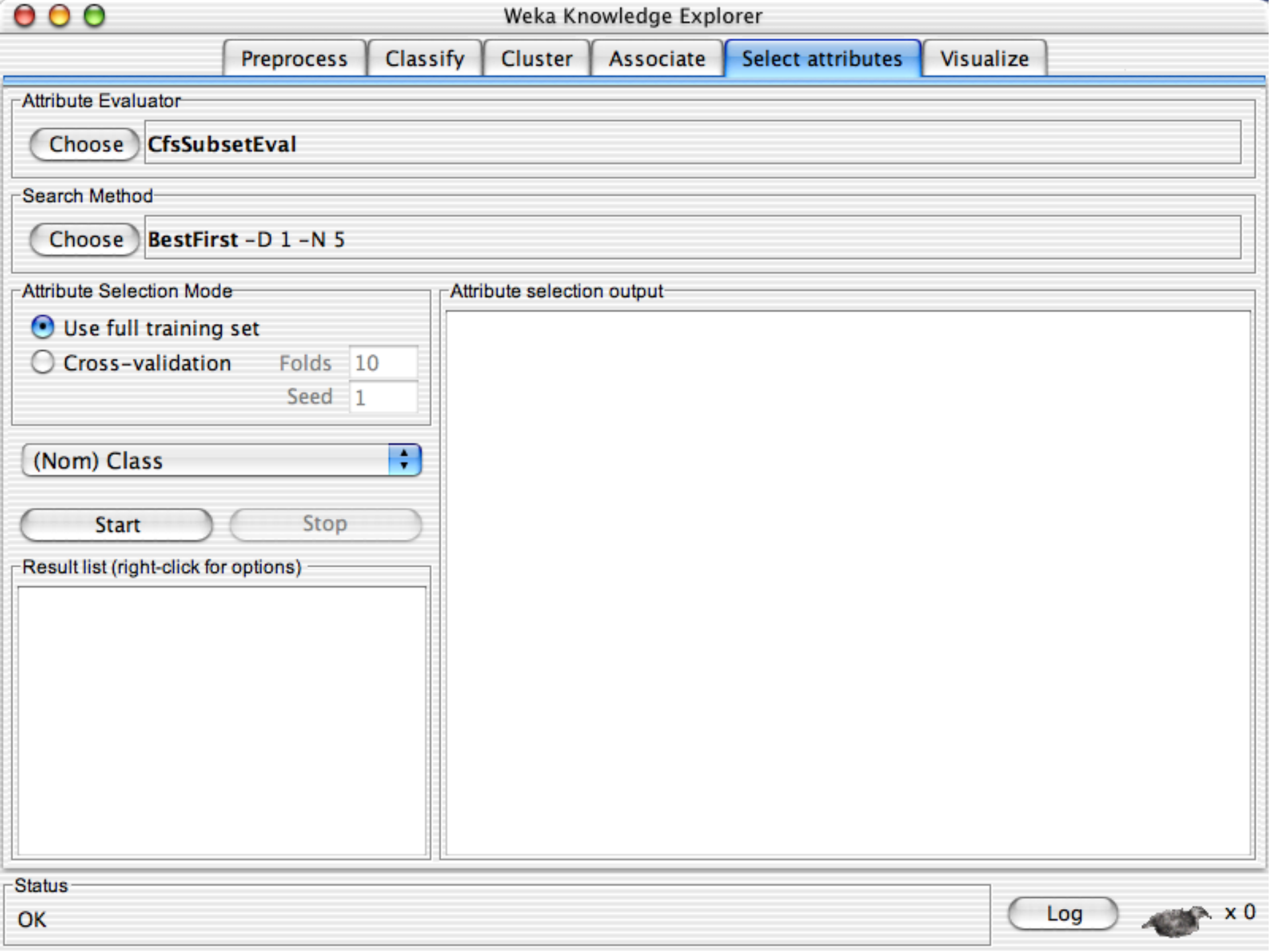
OK

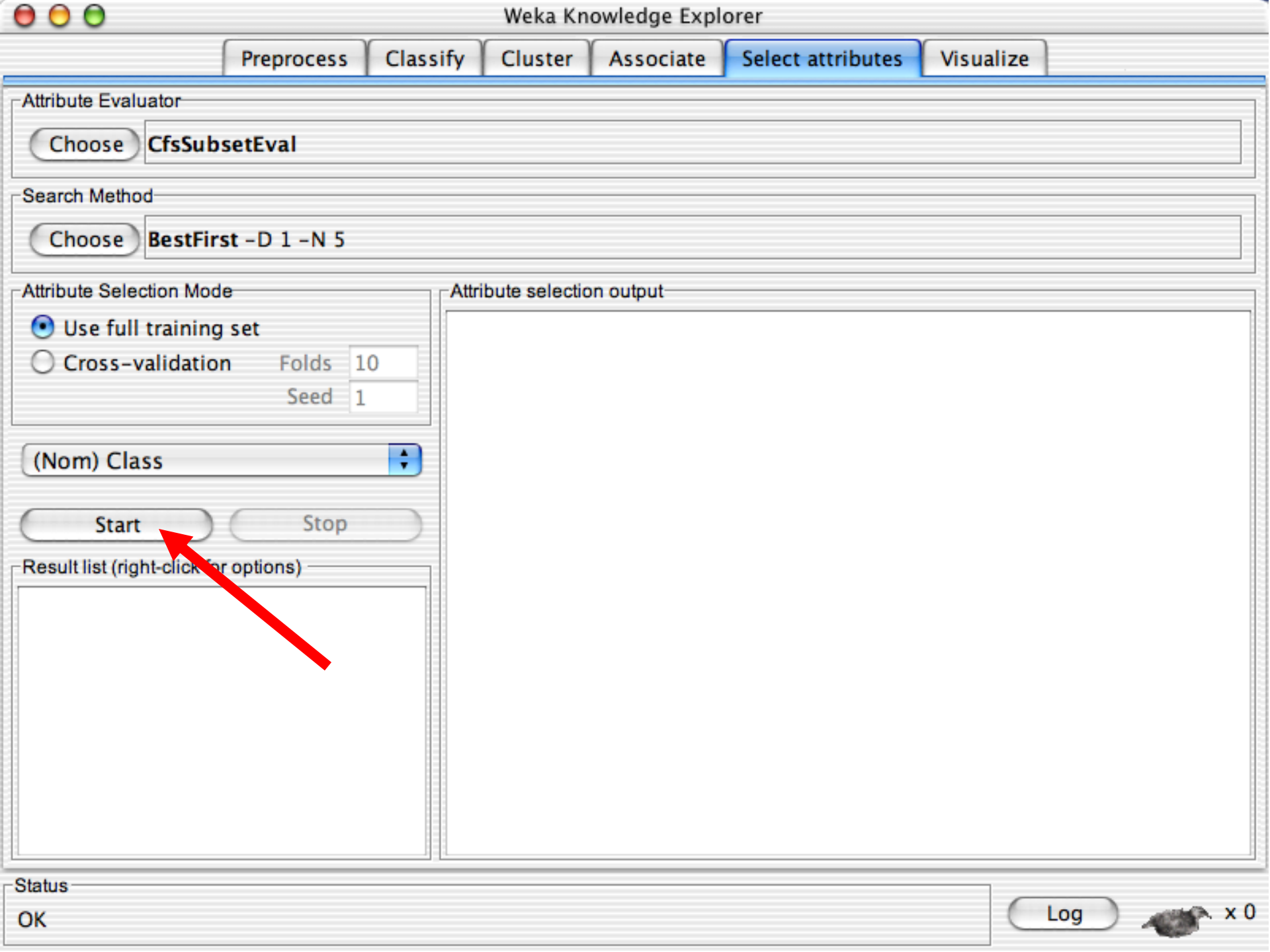
Log

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 Knowledge 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 Knowledge 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 Knowledge 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)

16:39:40 - BestFirst + CfsSubsetEval

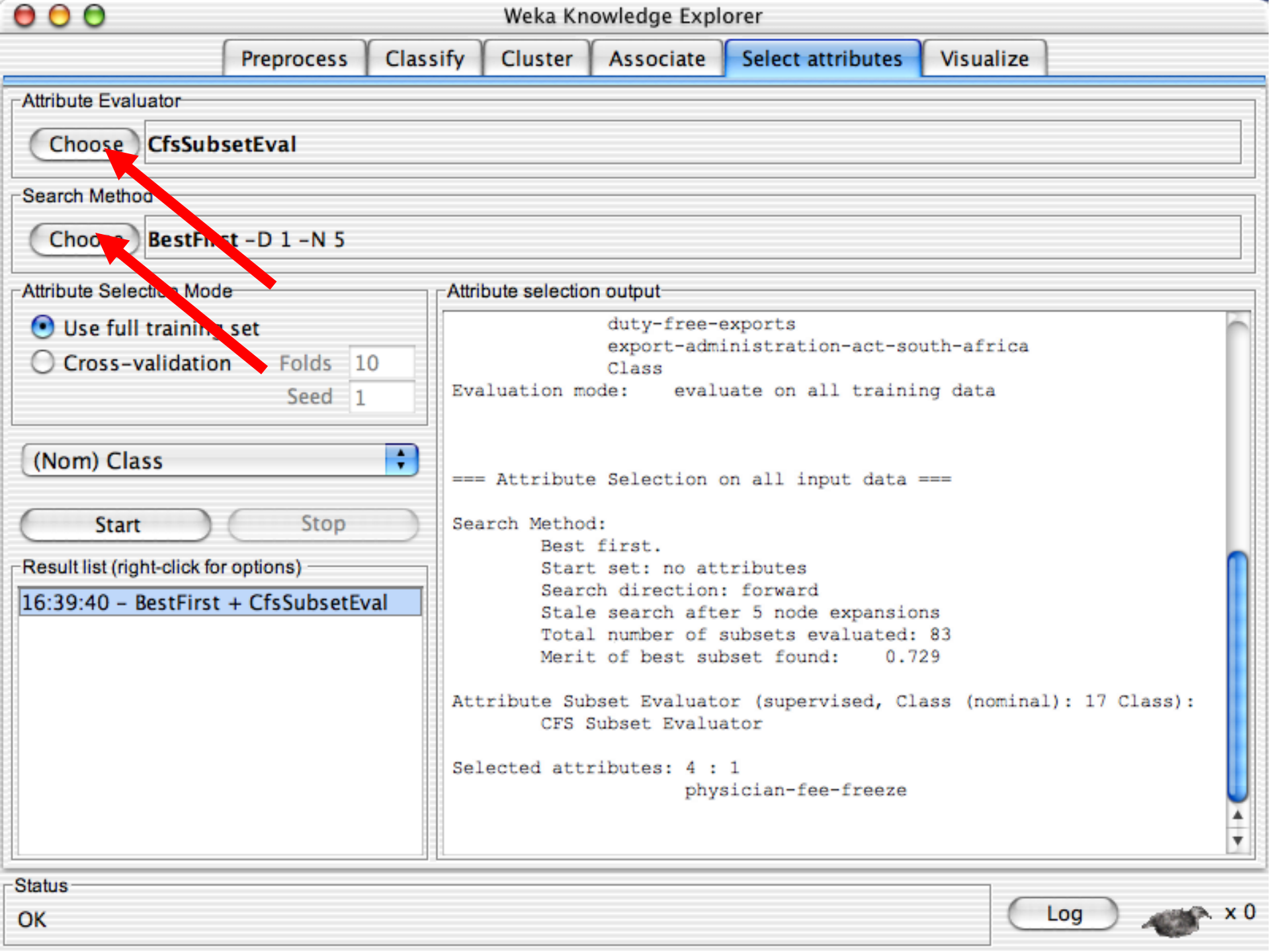
Attribute selection output

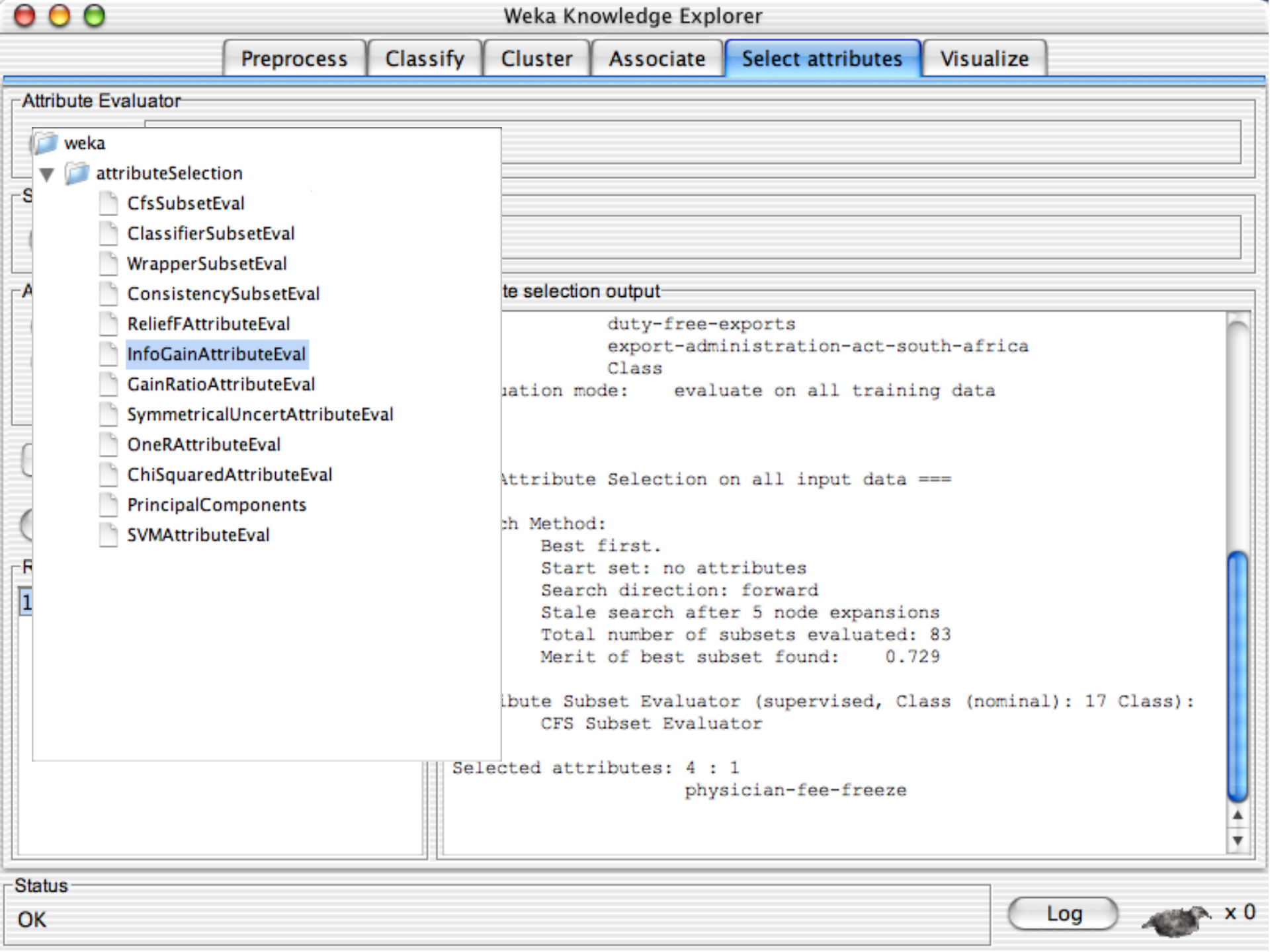
```
                duty-free-exports
                export-administration-act-south-africa
                Class
Evaluation mode:    evaluate on all training data



=== 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: 83
        Merit of best subset found:    0.729

Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
        CFS Subset Evaluator

Selected attributes: 4 : 1
                     physician-fee-freeze
```

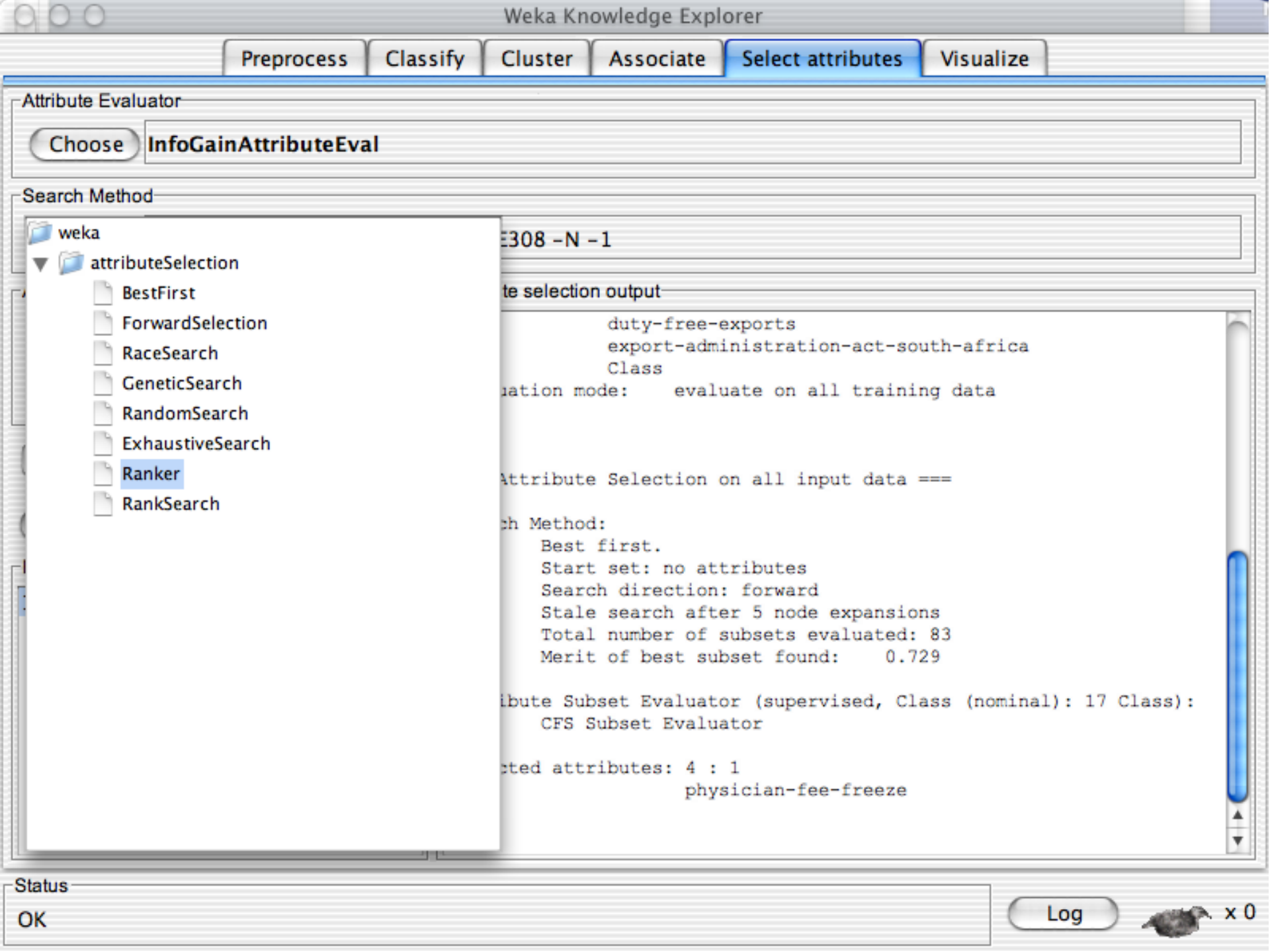
Status

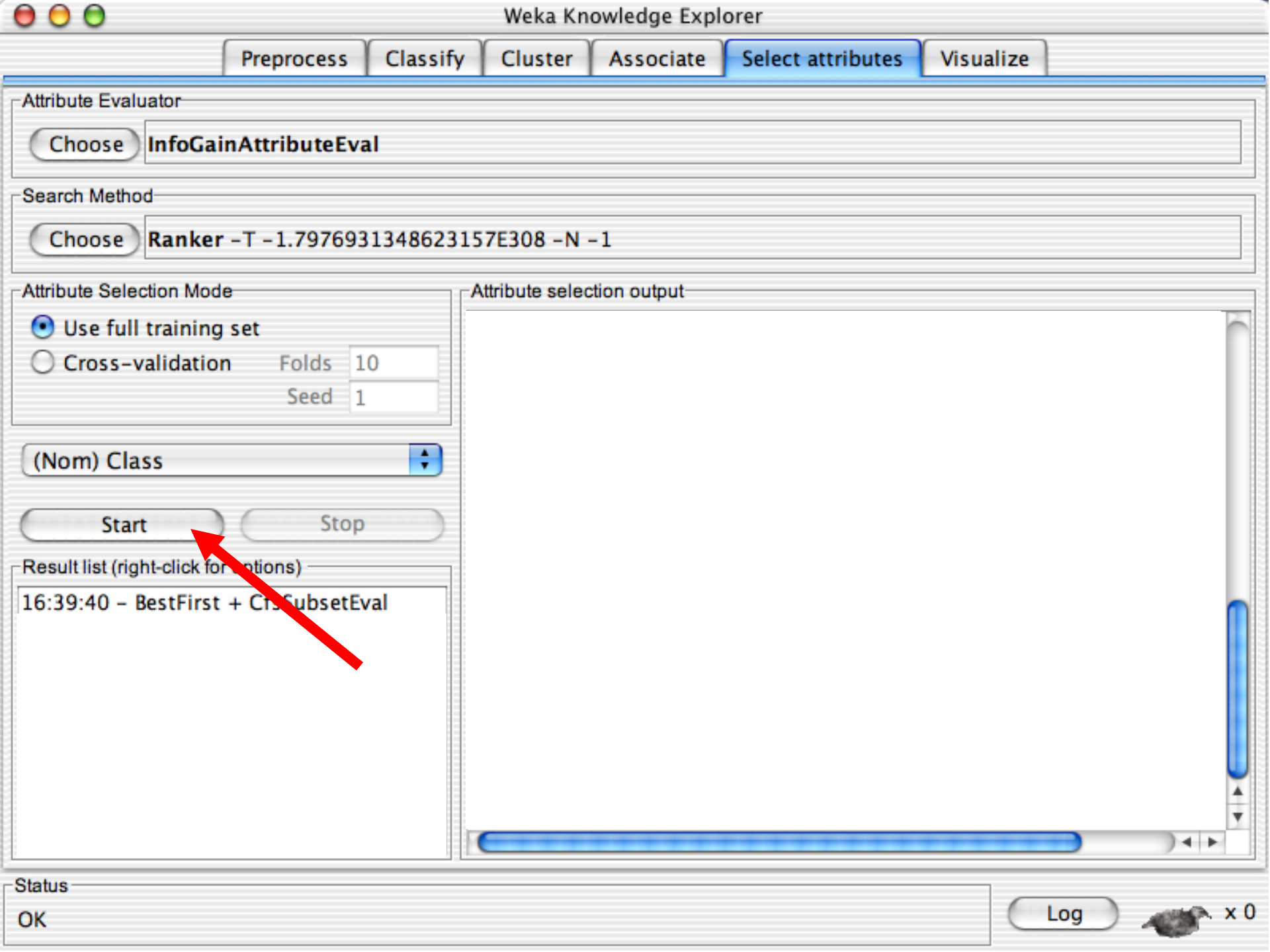
OK

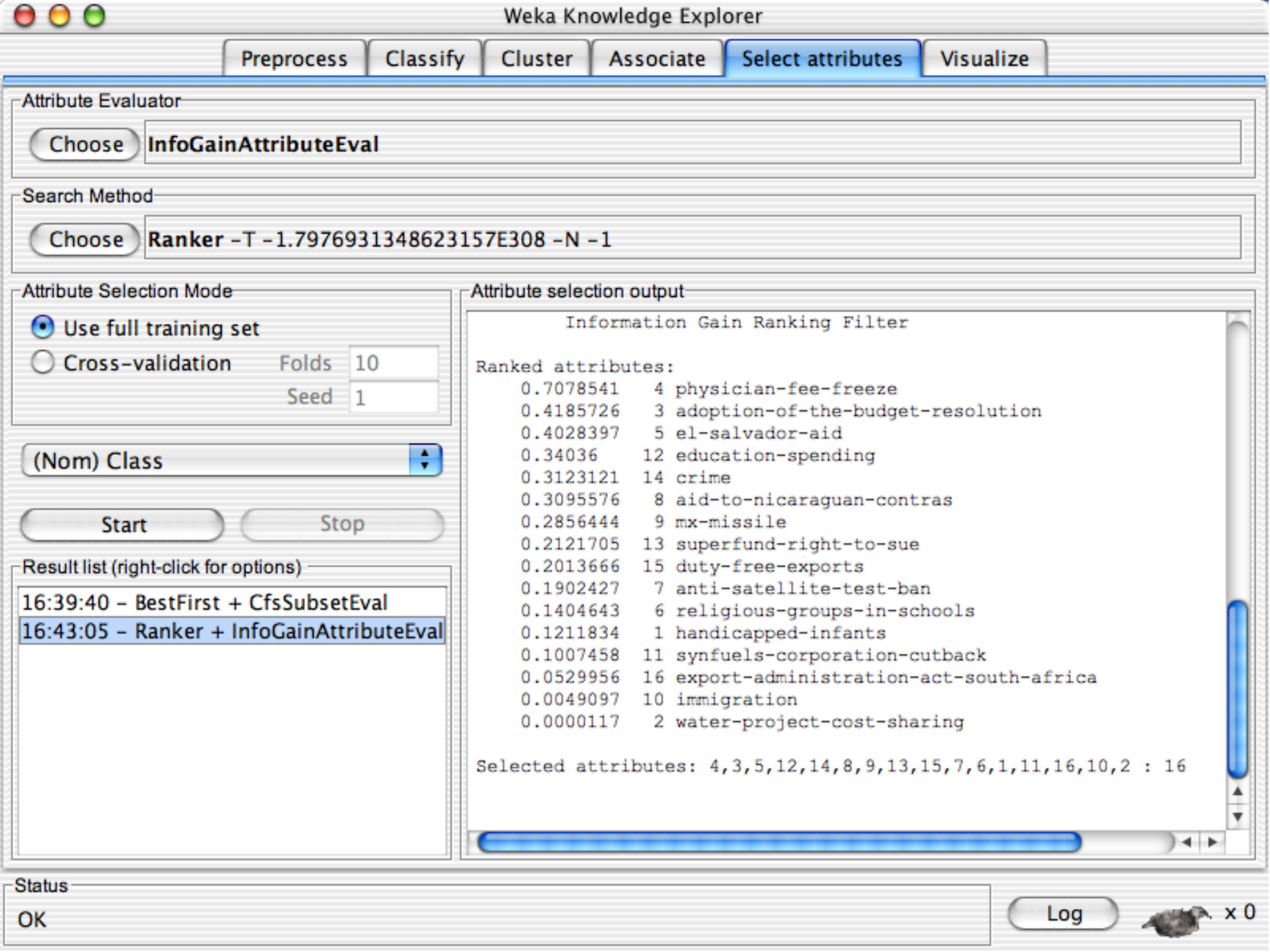
Log  x 0


Weka Knowledge 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)
16:39:40 - BestFirst + CfsSubsetEval
Attribute selection output
duty-free-exports
export-administration-act-south-africa
Class
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: 83
Merit of best subset found: 0.729
Attribute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Selected attributes: 4 : 1
physician-fee-freeze
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
weka
attributeSelection
CfsSubsetEval
ClassifierSubsetEval
WrapperSubsetEval
ConsistencySubsetEval
ReliefFAttributeEval
InfoGainAttributeEval
GainRatioAttributeEval
SymmetricalUncertAttributeEval
OneRAttributeEval
ChiSquaredAttributeEval
PrincipalComponents
SVMAttributeEval
te selection output
duty-free-exports
export-administration-act-south-africa
Class
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: 83
Merit of best subset found: 0.729
ibute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
Selected attributes: 4 : 1
physician-fee-freeze
Status
OK
Log
x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Attribute Evaluator
Choose
InfoGainAttributeEval
Search Method
weka
attributeSelection
BestFirst
ForwardSelection
RaceSearch
GeneticSearch
RandomSearch
ExhaustiveSearch
Ranker
RankSearch
E308 -N -1
te selection output
duty-free-exports
export-administration-act-south-africa
Class
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: 83
Merit of best subset found: 0.729
ibute Subset Evaluator (supervised, Class (nominal): 17 Class):
CFS Subset Evaluator
cted attributes: 4 : 1
physician-fee-freeze
Status
OK
Log
x 0


Weka Knowledge 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
Attribute selection output
(Nom) Class
Start
Stop
Result list (right-click for options)
16:39:40 - BestFirst + CfsSubsetEval
Status
OK
Log
x 0


Weka Knowledge 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)
16:39:40 - BestFirst + CfsSubsetEval
16:43:05 - Ranker + InfoGainAttributeEval
Attribute selection output
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 Knowledge Explorer

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

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

Filter

Choose **None** | Apply

Current relation

Relation: Glass
Instances: 214 Attributes: 10

Attributes

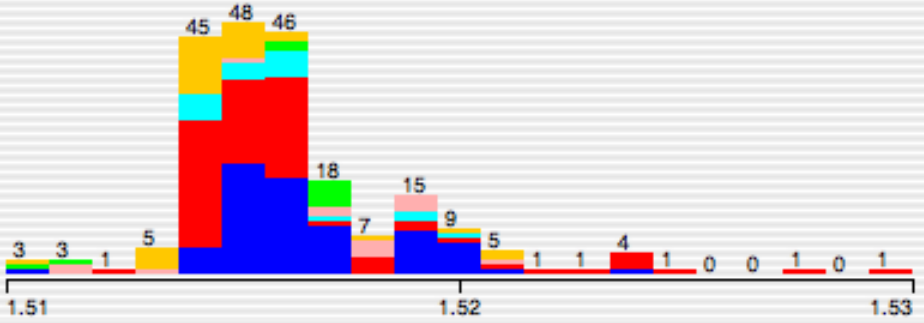
| No. | Name |
|---|---|
| 1 | RI |
| 2 | Na |
| 3 | Mg |
| 4 | Al |
| 5 | Si |
| 6 | K |
| 7 | Ca |
| 8 | Ba |
| 9 | Fe |
| 10 | Type |

Selected attribute

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

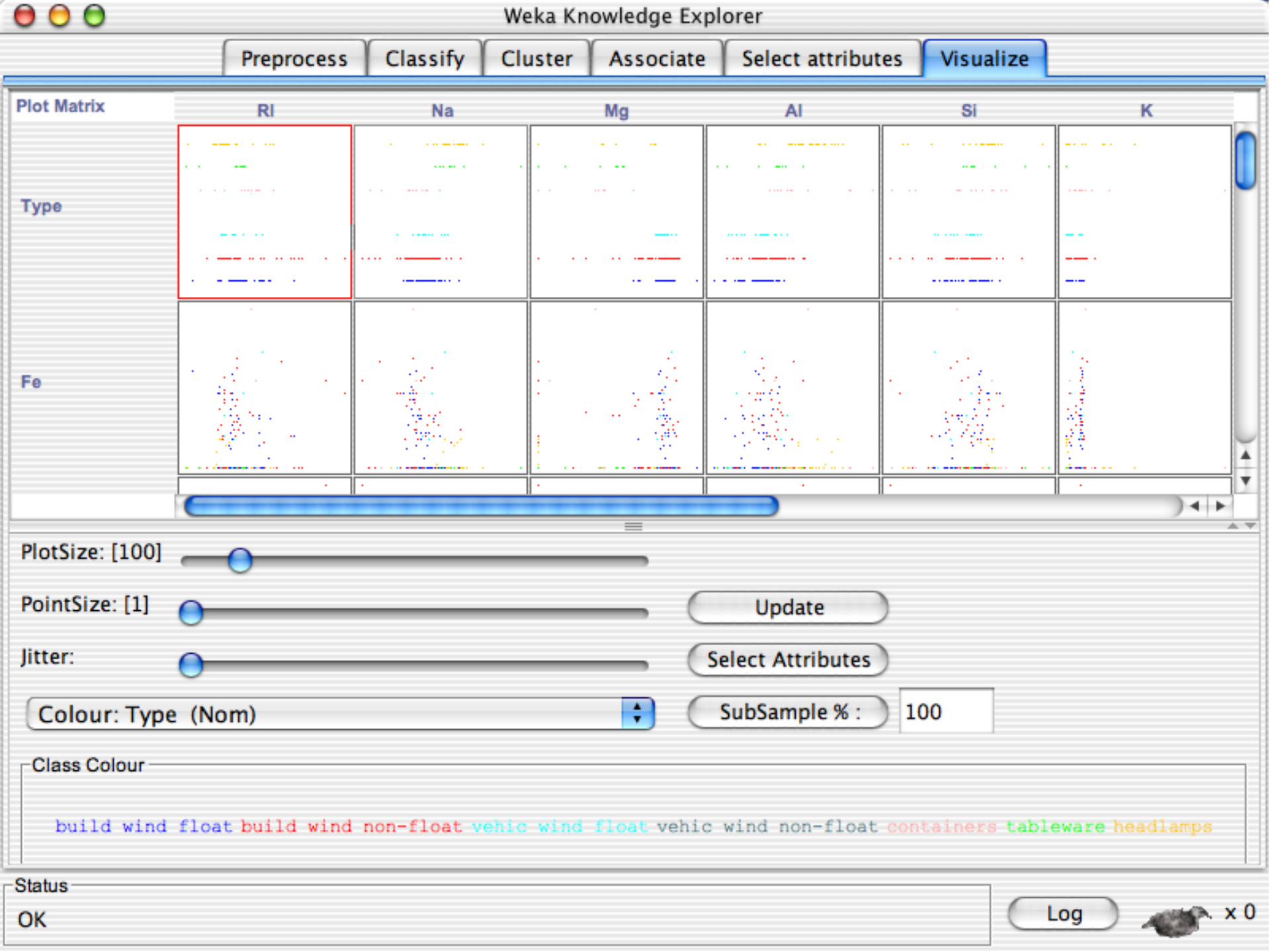
| Statistic | Value |
|---|---|
| Minimum | 1.511 |
| Maximum | 1.534 |
| Mean | 1.518 |
| StdDev | 0.003 |

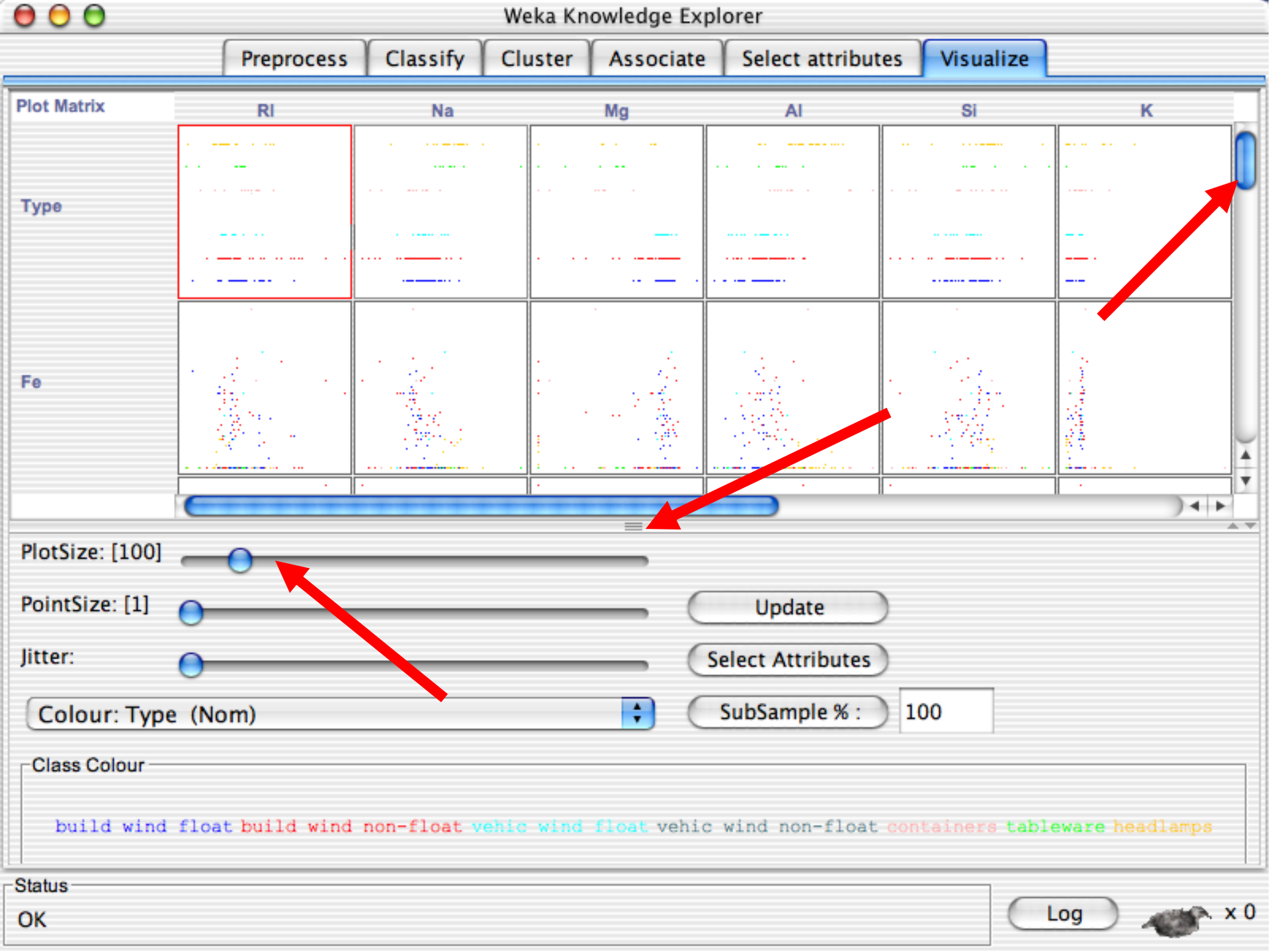
Colour: Type (Nom) | Visualize All




Status

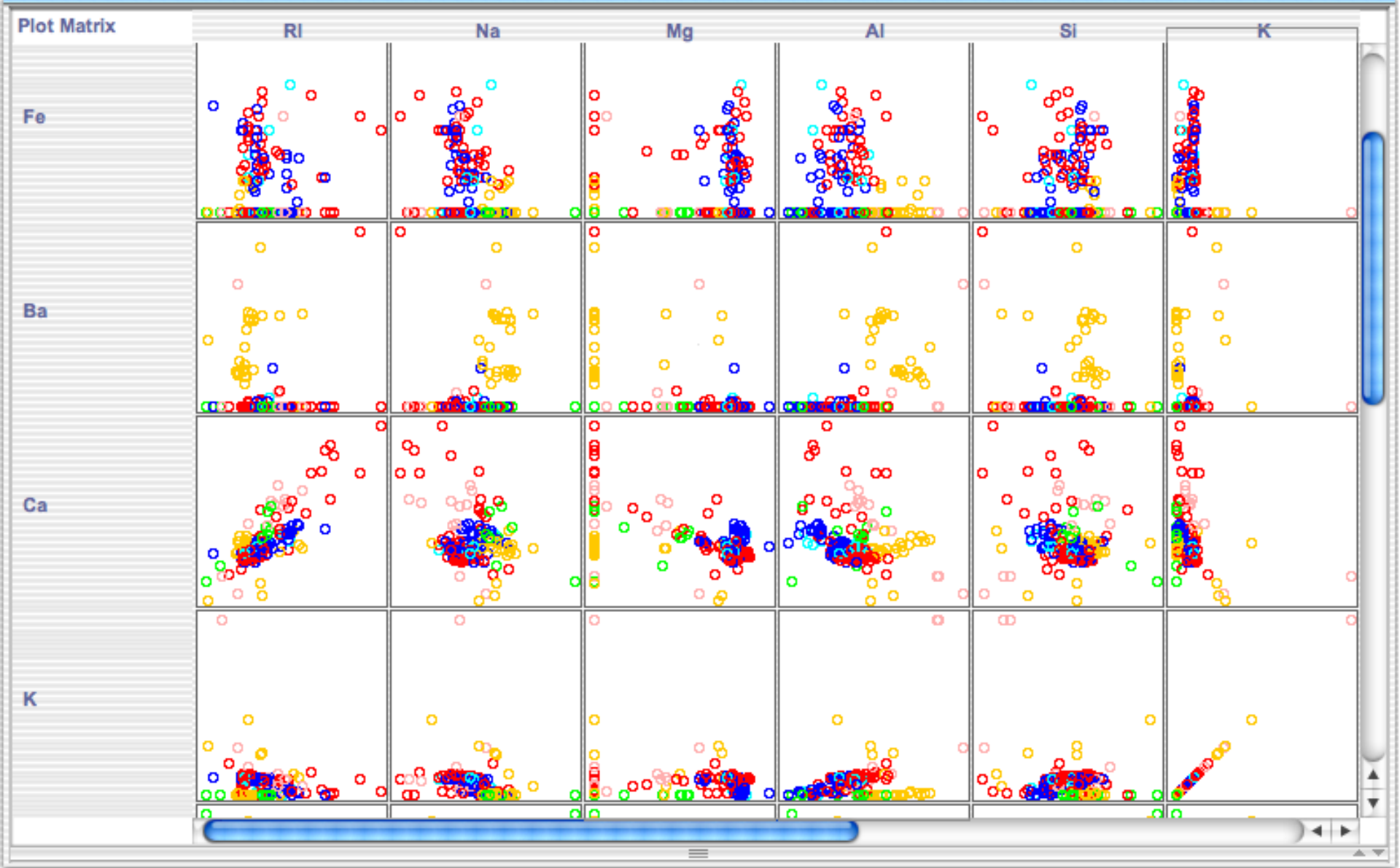
OK | Log | x 0


Weka Knowledge 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
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

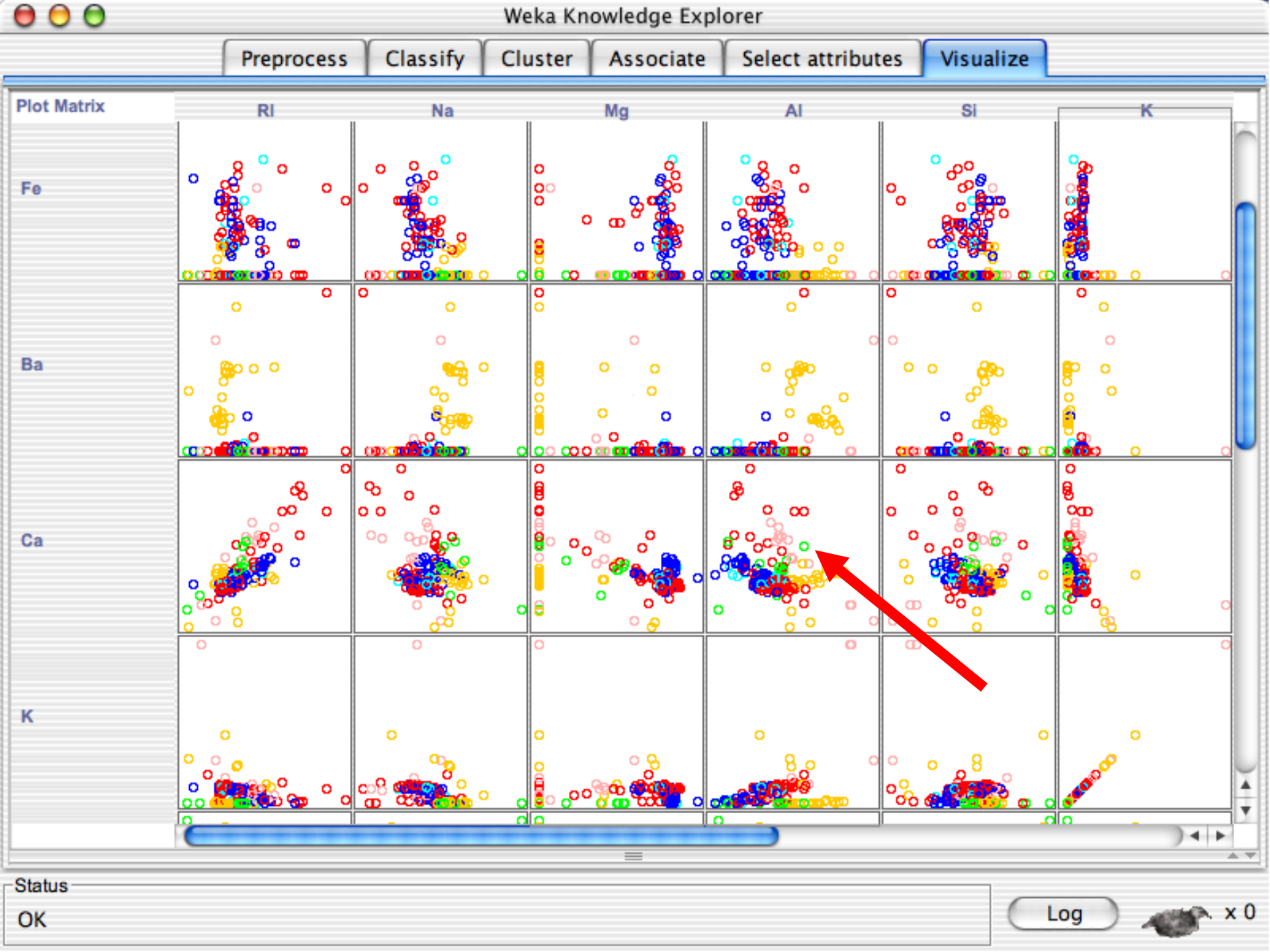



Status

OK

Log

x 0


Weka Knowledge Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Plot Matrix
RI
Na
Mg
Al
Si
K
Fe
Ba
Ca
K
Status
OK
Log
x 0

Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

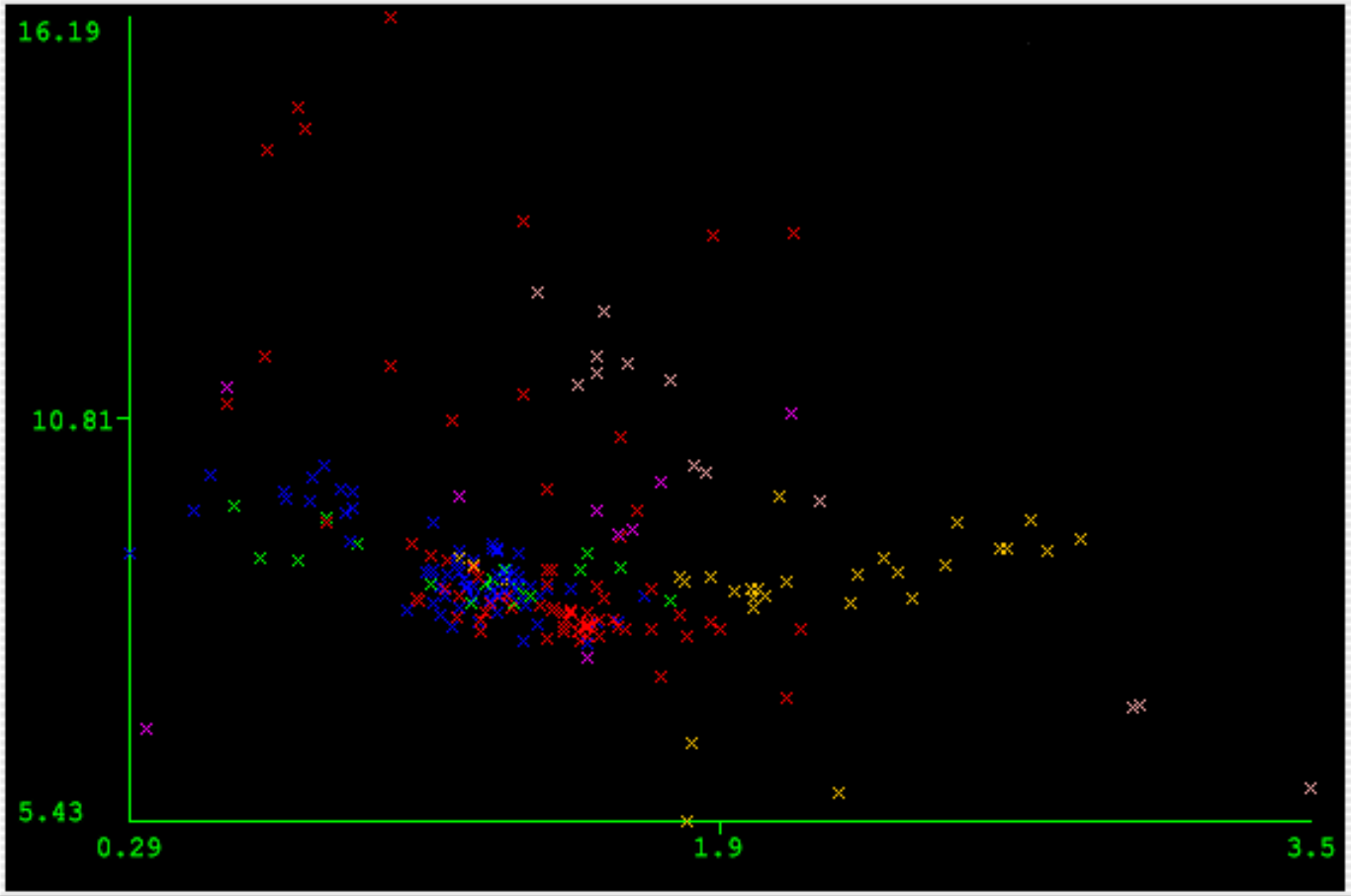
Colour: Type (Nom)

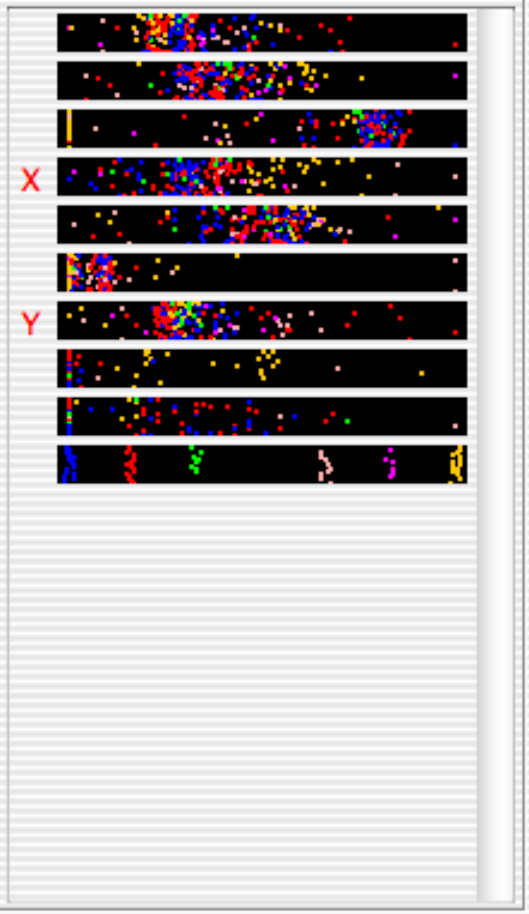
Select Instance

Reset | Clear | Save

Jitter

Plot: Glass




Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

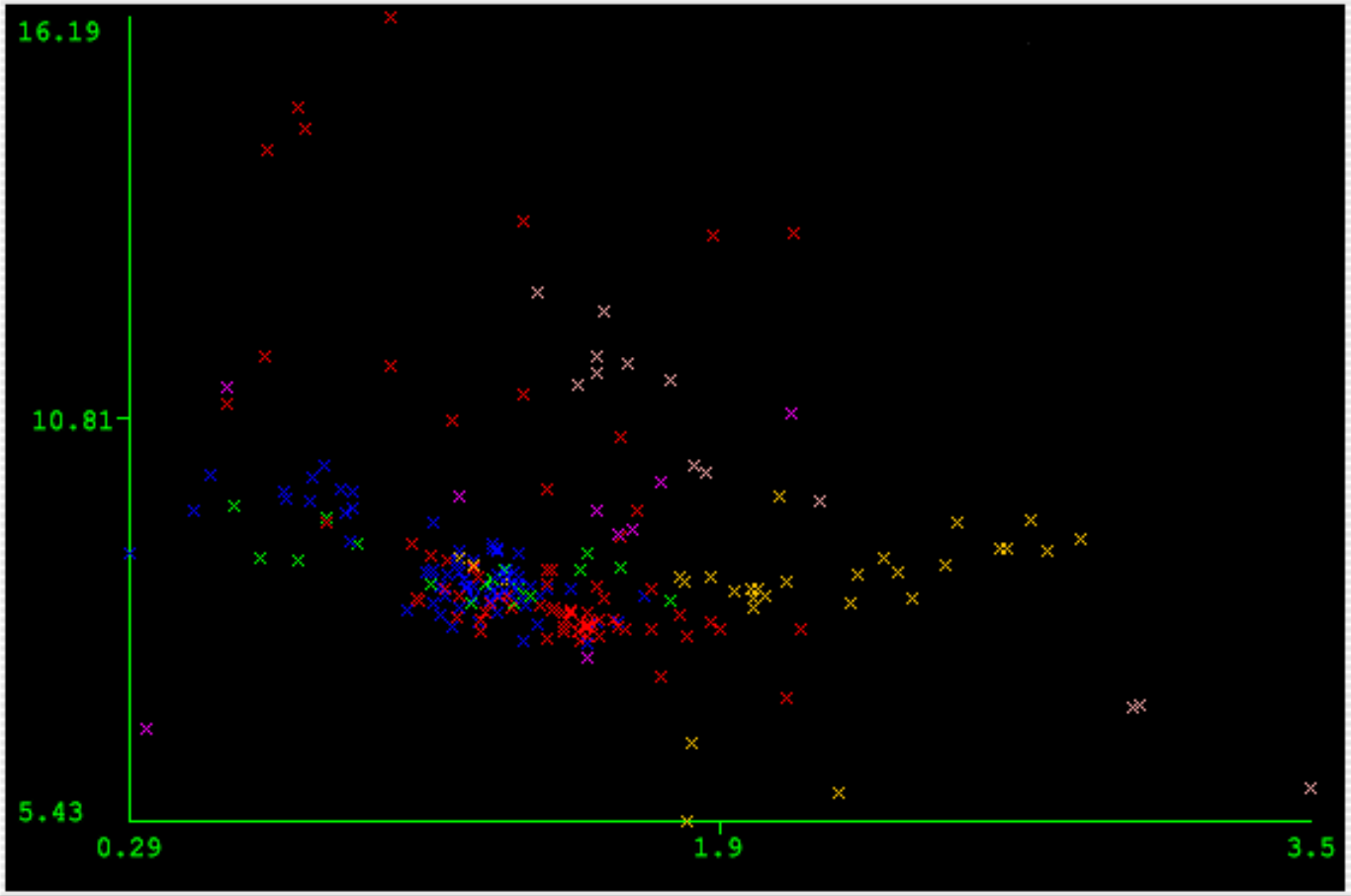
Colour: Type (Nom)

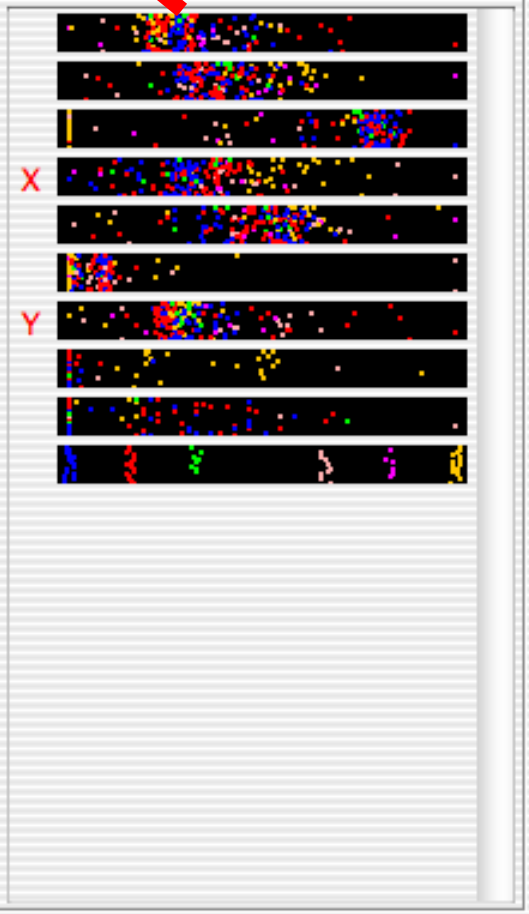
Select Instance

Reset | Clear | Save

Jitter

Plot: Glass




Class colour

build wind float

build wind non-float

vehic wind float

vehic wind non-float

containers

tableware

headlamps

Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

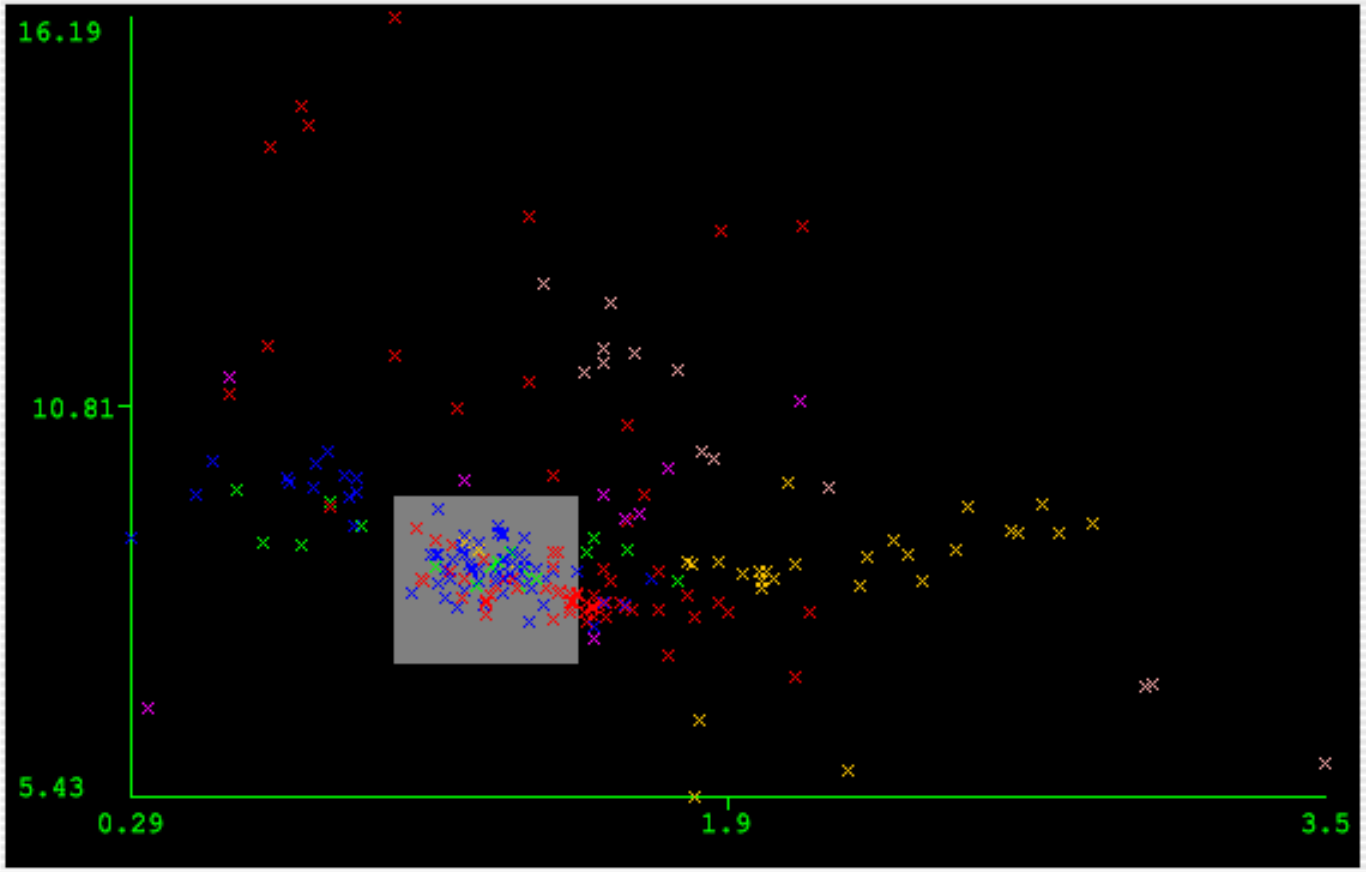
Colour: Type (Nom)

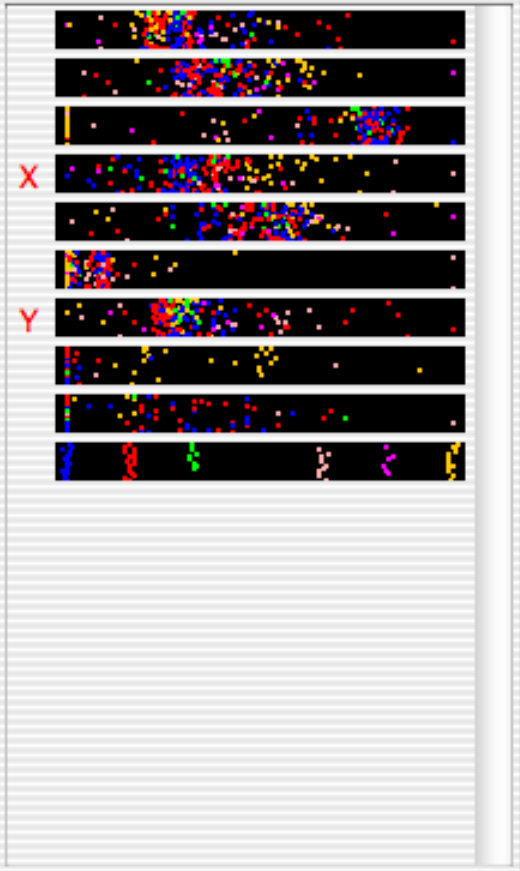
Rectangle

Submit | Clear | Save

Jitter

Plot: Glass




Class colour

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

Weka Knowledge Explorer: Visualizing Glass

X: Al (Num)

Y: Ca (Num)

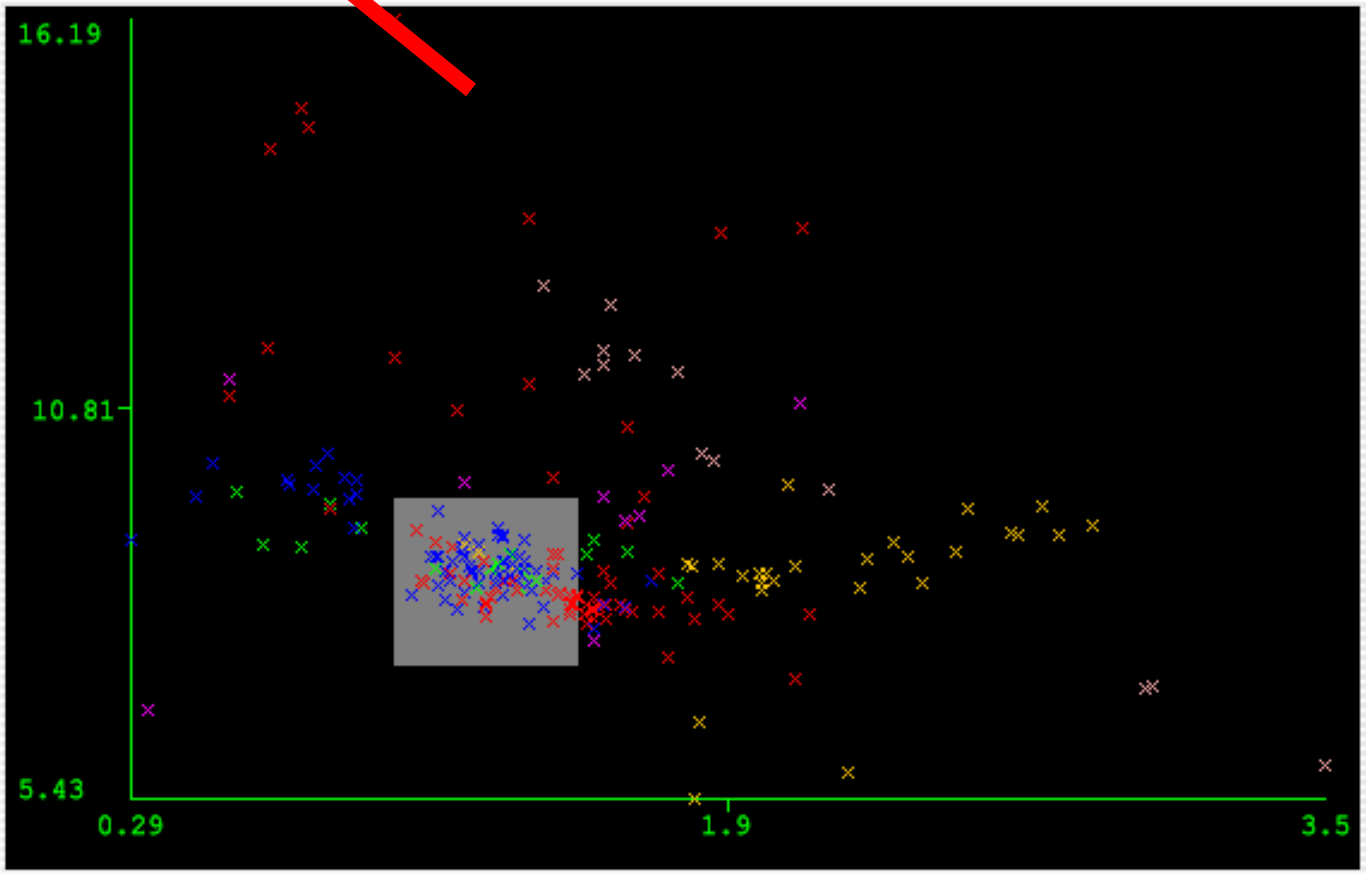
Colour: Type (Nom)

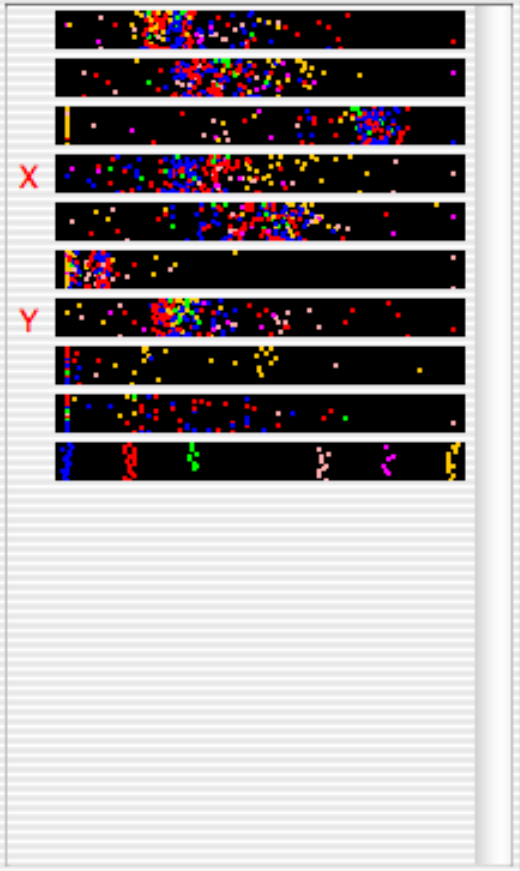
Rectangle

Submit | Clear | Save

Jitter

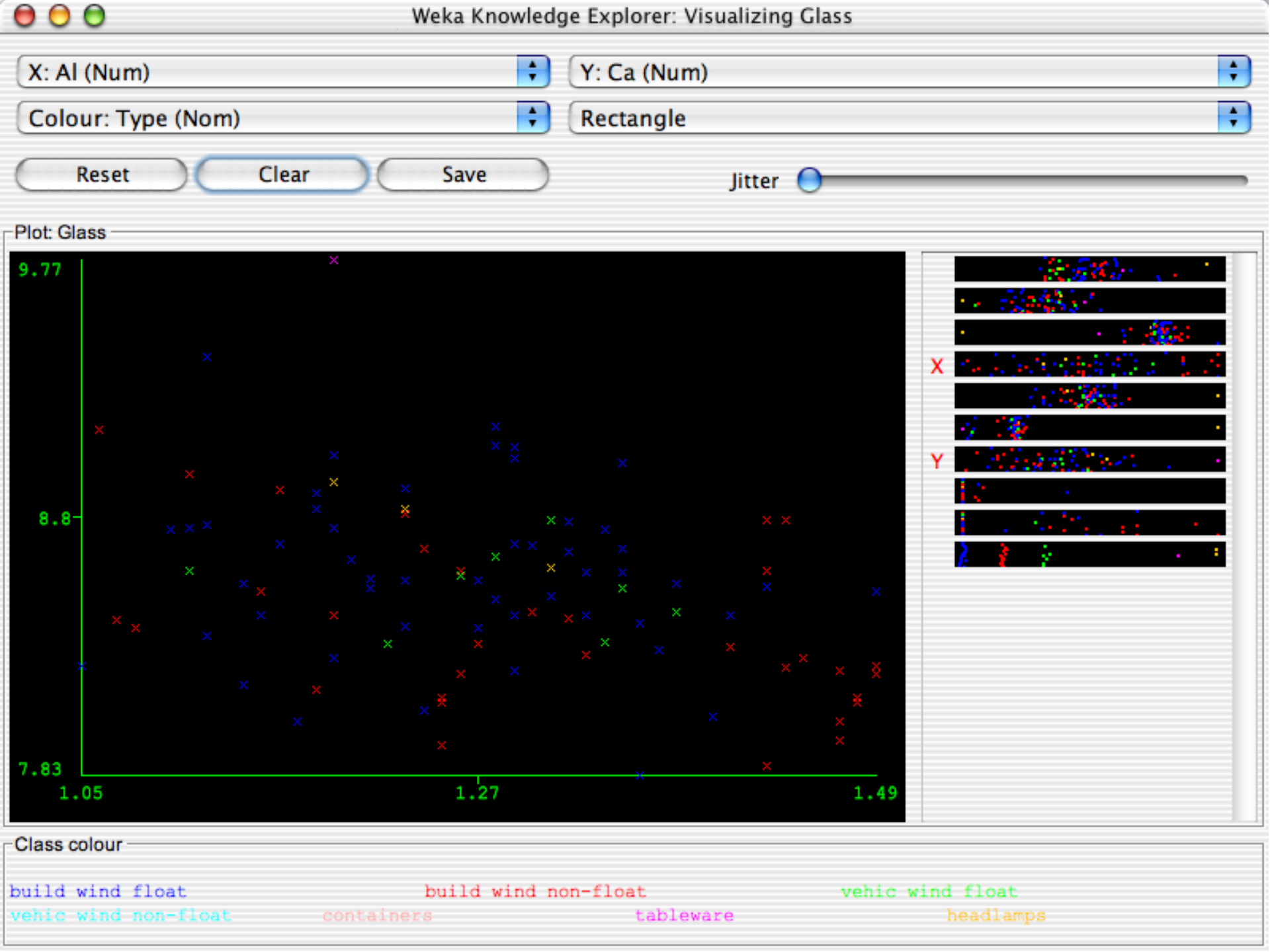
Plot: Glass



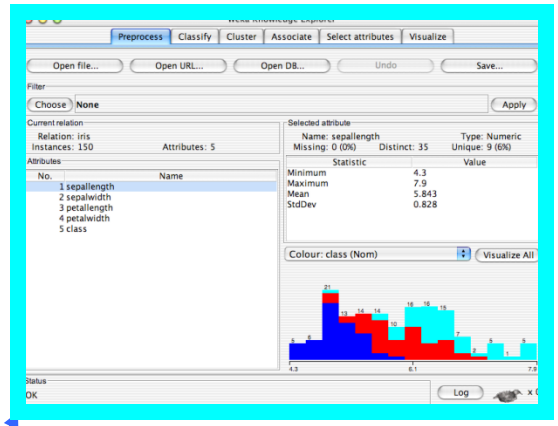

Class colour

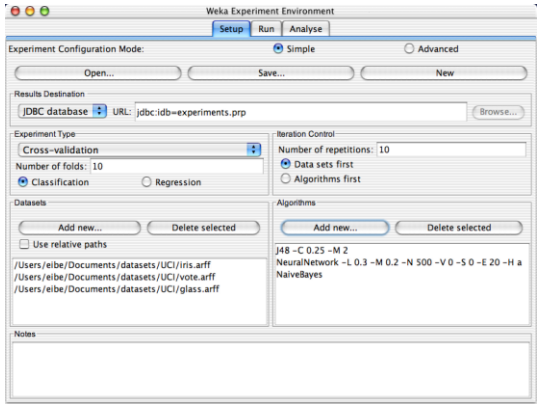
build wind float build wind non-float vehic wind float vehic wind non-float containers tableware headlamps

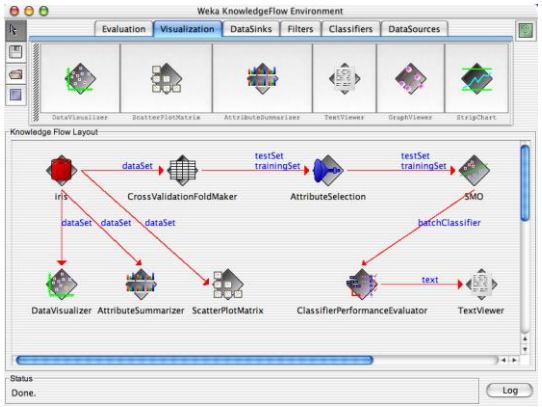

Weka Knowledge Explorer: Visualizing Glass
X: Al (Num)
Y: Ca (Num)
Colour: Type (Nom)
Rectangle
Reset
Clear
Save
Jitter
Plot: Glass
9.77
8.8
7.83
1.05
1.27
1.49
X
Y
Class colour
build wind float
build wind non-float
vehic wind float
vehic wind non-float
containers
tableware
headlamps


Welcome to the WEKA SimpleCLI
> help
Command must be one of:
java <classname> <args>
break
kill
cls
exit
help <command>


Weka GUI Chooser
Waikato Environment for Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow

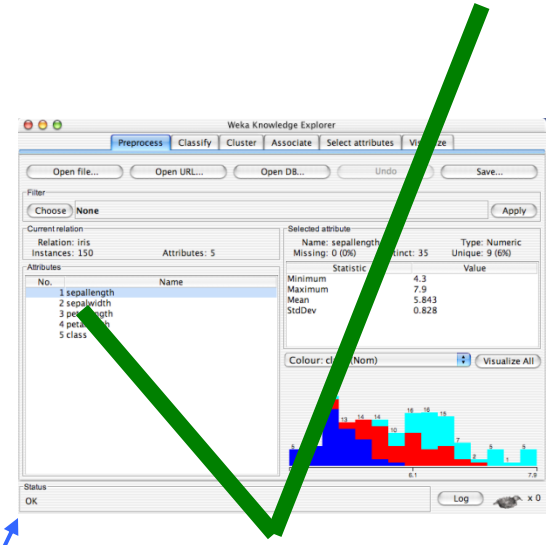

Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Choose
None
Apply
Relation: iris
Instances: 150
Attributes: 5
1 sepallength
2 sepalwidth
3 petallength
4 petalwidth
5 class
Minimum 4.3
Maximum 7.9
Mean 5.843
StdDev 0.828
Visualize All
Status
OK
Log

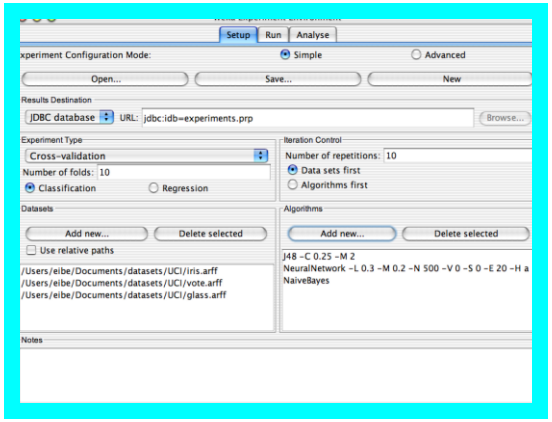

Weka Experiment Environment
Setup
Run
Analyse
Experiment Configuration Mode:
Simple
Advanced
Open...
Save...
New
Results Destination
JDBC database
URL:
jdbc:idb=experiments.prp
Browse...
Experiment Type
Cross-validation
Number of folds: 10
Classification
Regression
Iteration Control
Number of repetitions: 10
Data sets first
Algorithms first
Datasets
Add new...
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
Algorithms
Add new...
Delete selected
J48 -C 0.25 -M 2
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
NaiveBayes
Notes

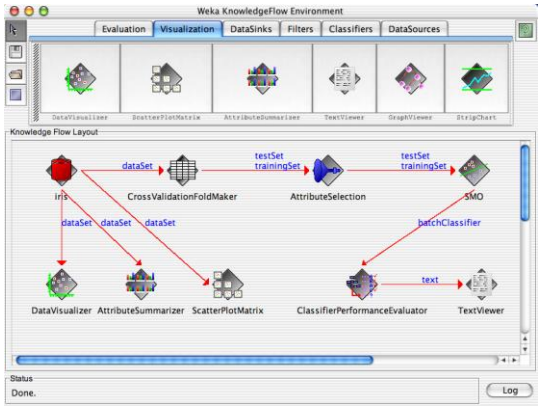

Weka KnowledgeFlow Environment
Evaluation
Visualization
DataSinks
Filters
Classifiers
DataSources
Knowledge Flow Layout
iris
CrossValidationFoldMaker
AttributeSelection
SMO
DataVisualizer
AttributeSummarizer
ScatterPlotMatrix
ClassifierPerformanceEvaluator
TextViewer
dataSet
testSet
trainingSet
batchClassifier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


Welcome to the WEKA SimpleCLI
> help
Command must be one of:
java <classname> <args>
break
kill
cls
exit
help <command>


Weka GUI Chooser
Waikato Environment for Knowledge Analysis
(c) 1999 - 2003
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow


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
J48 -C 0.25 -M 2
NeuralNetwork -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
NaiveBayes


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

# 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
- 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