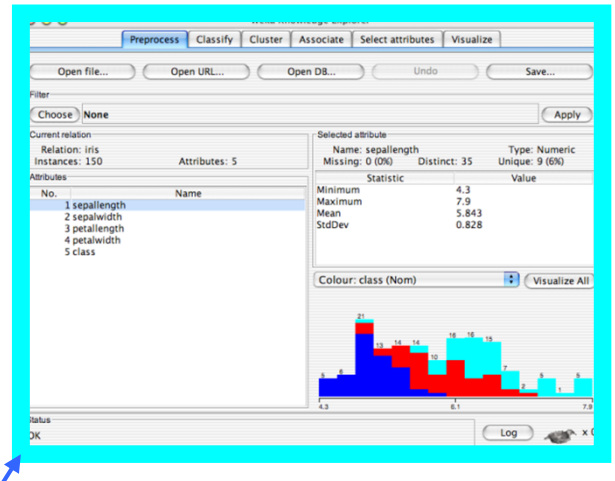
Data Mining with WEKA
(Mining Association Rules)

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 GUI Chooser
Waikato Environment for
Knowledge Analysis
Version 3.4.5
(c) 1999 - 2005
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...
Filter
Choose
None
Apply
Current relation
Relation: iris
Instances: 150
Attributes: 5
Selected attribute
Name: sepallength
Type: Numeric
Missing: 0 (0%)
Distinct: 35
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

Weka Explorer

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

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

Filter

Choose **None** Apply

Current relation

Relation: soybean
Instances: 683 Attributes: 36

Attributes

All | None | Invert

No.	Name
1	date
2	plant-stand
3	precip
4	temp
5	hail
6	crop-hist
7	area-damaged
8	severity
9	seed-tmt
10	germination
11	plant-growth
12	leaves
13	leafspots-halo
14	leafspots-marg
15	leafspot-size
16	leaf-shread
17	leaf-malf
18	leaf-mild
19	stem
20	lodging
21	stem-cankers
22	canker-lesion
23	fruiting-bodies
24	external-decay

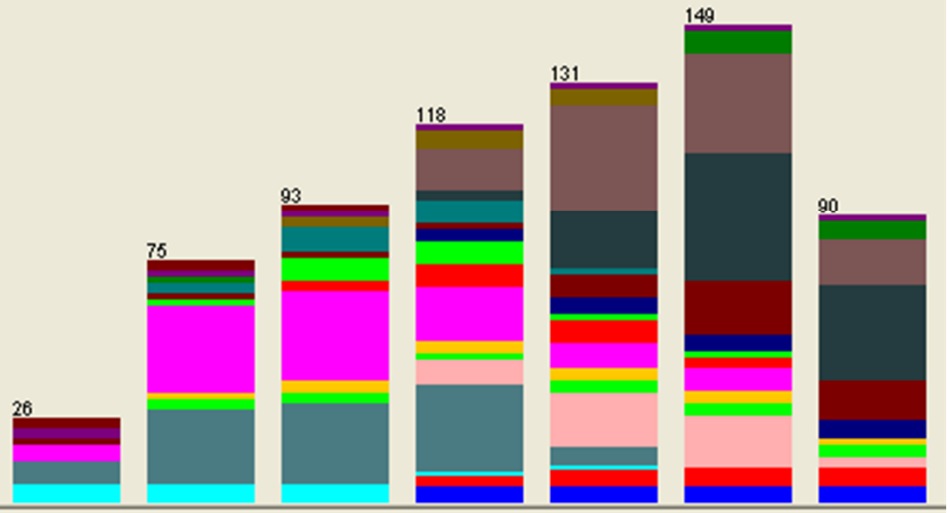
Remove

Selected attribute

Name: date Type: Nominal
Missing: 1 (0%) Distinct: 7 Unique: 0 (0%)

Label	Count
april	26
may	75
june	93
july	118
august	131
september	149
october	90

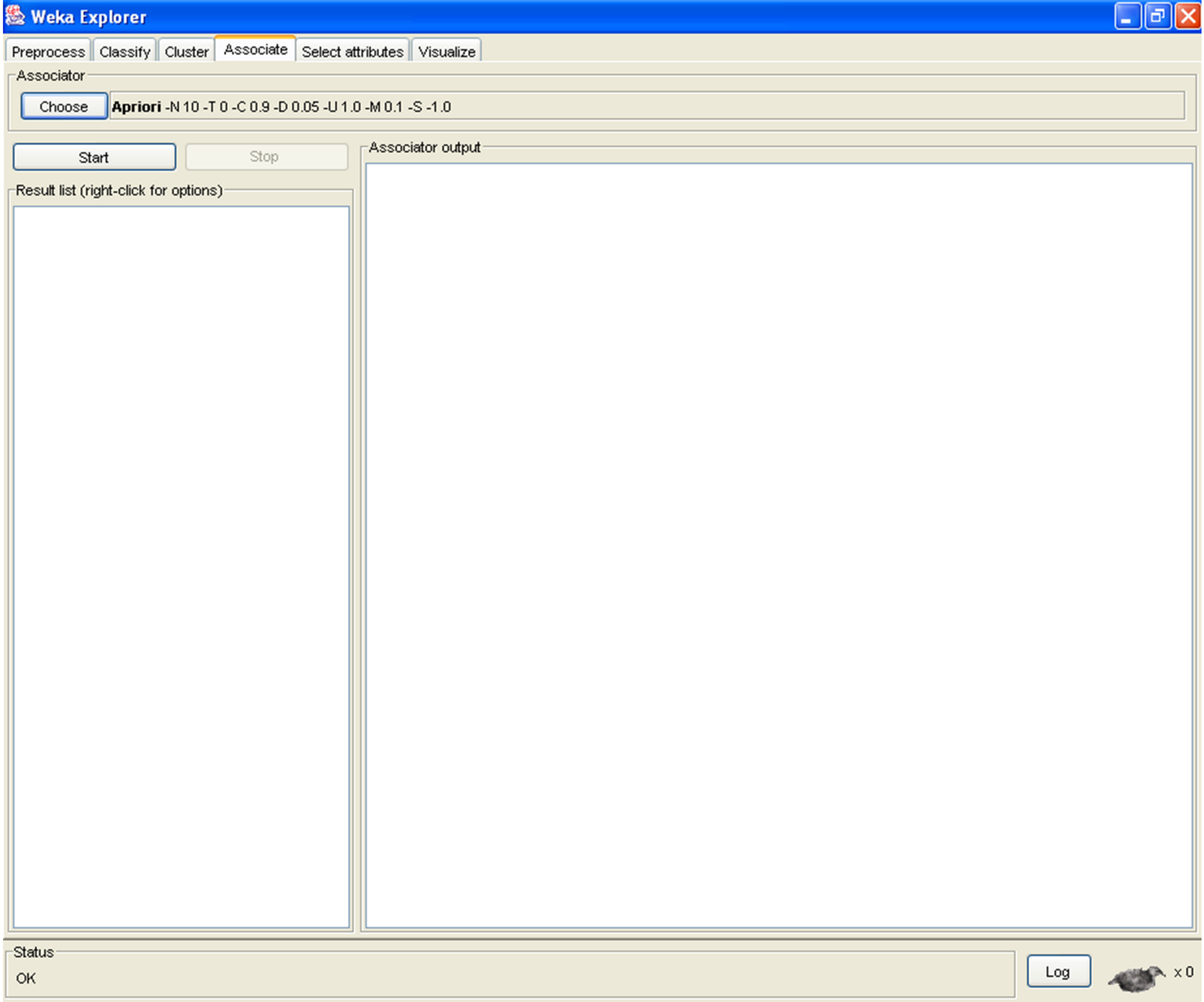
Class: class (Nom) Visualize All

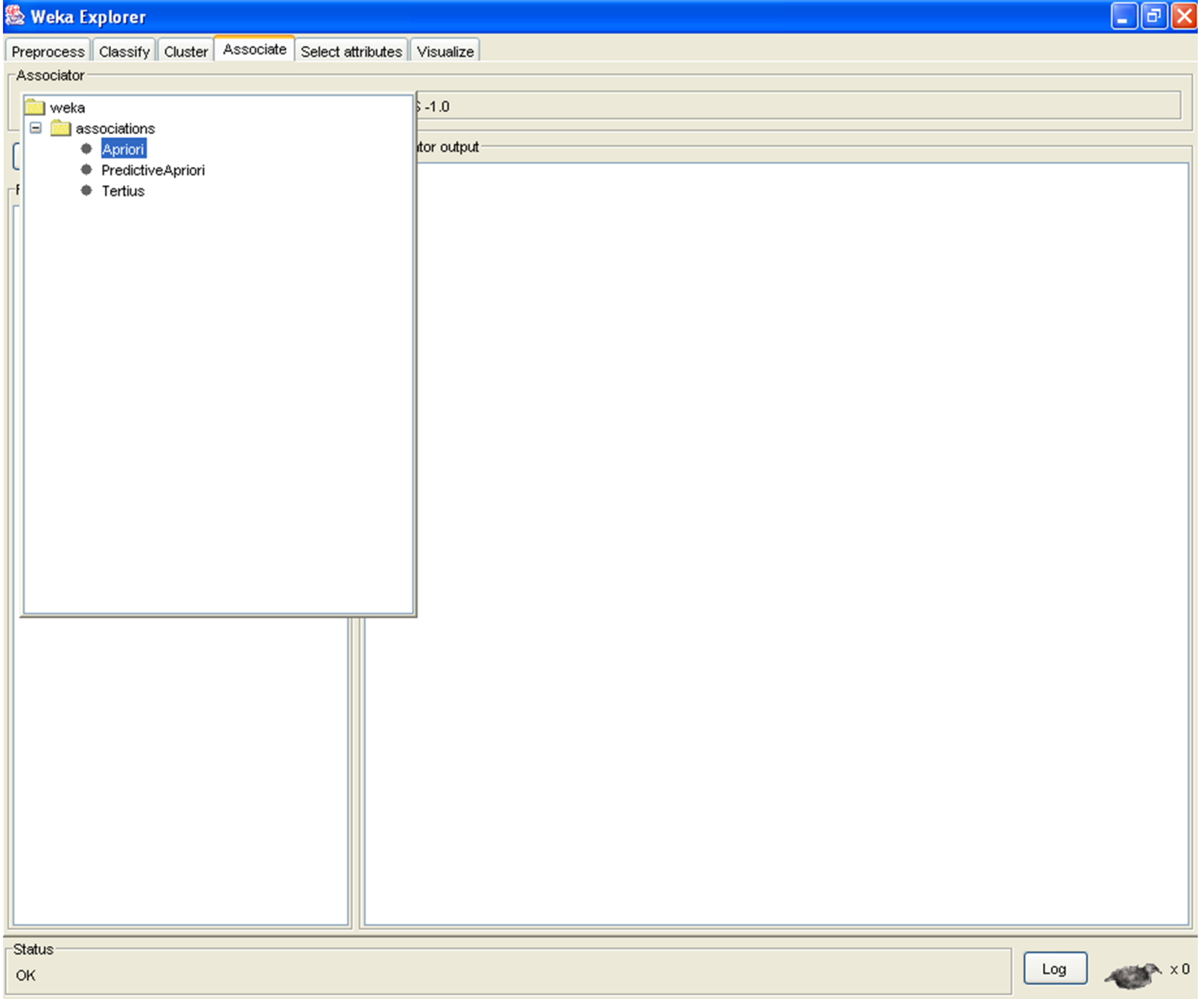


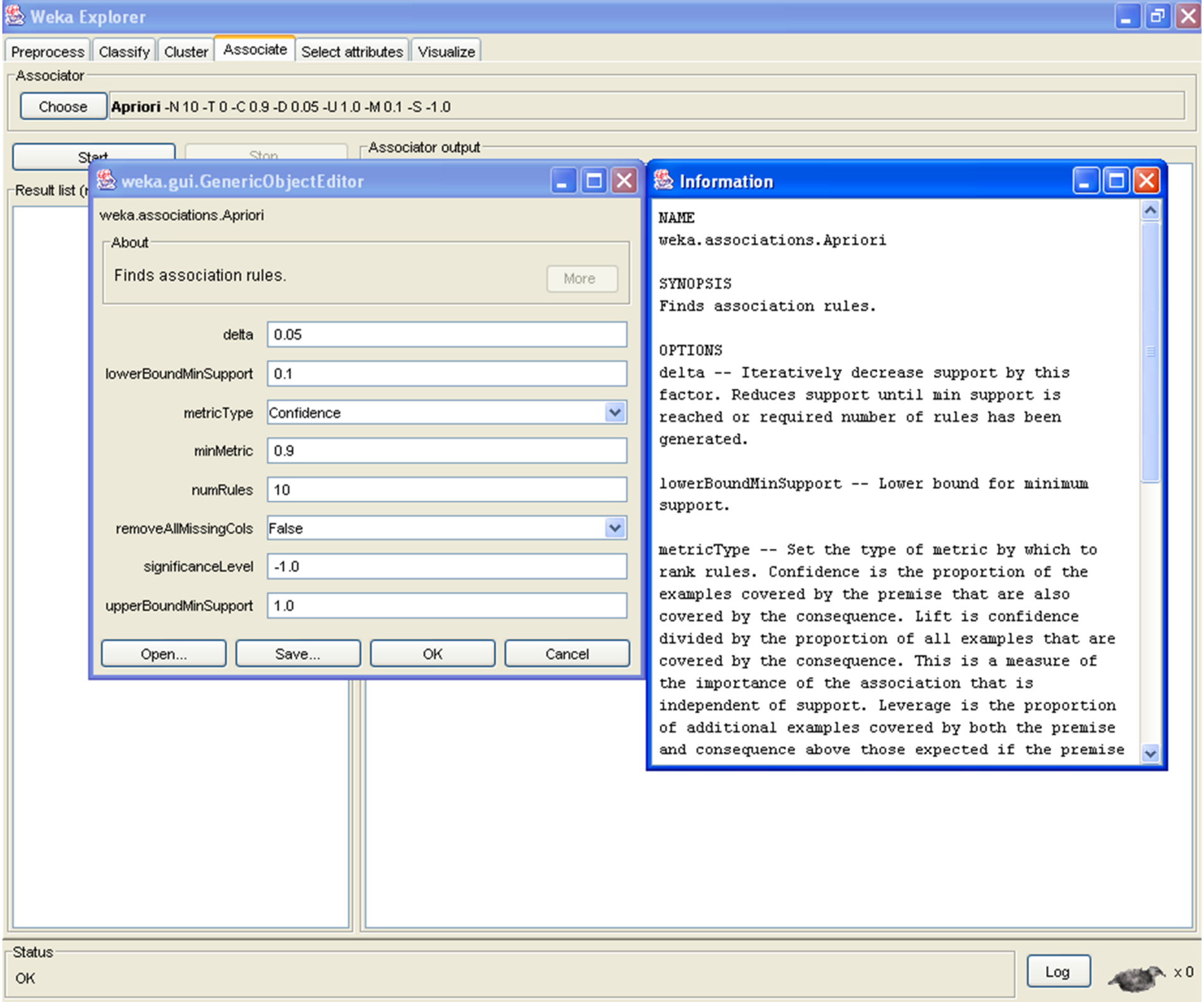
Status

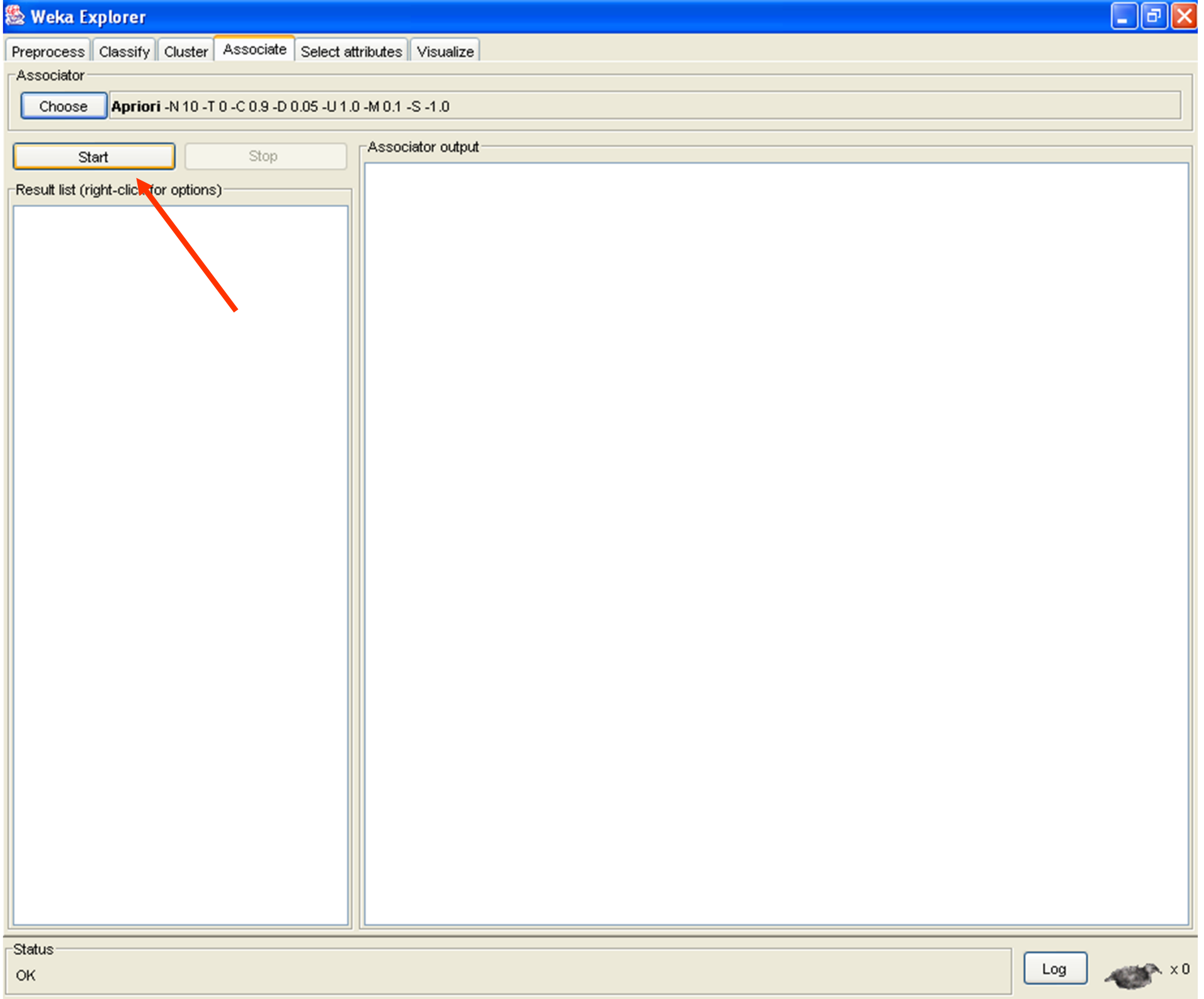
OK

Log x 0


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


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
weka
associations
Apriori
PredictiveApriori
Tertius
-1.0
tor output
Status
OK
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Associator
Choose
Apriori -N 10 -T 0 -C 0.9 -D 0.05 -U 1.0 -M 0.1 -S -1.0
Start
Stop
Associator output
Result list (
weka.gui.GenericObjectEditor
weka.associations.Apriori
About
Finds association rules.
More
delta
0.05
lowerBoundMinSupport
0.1
metricType
Confidence
minMetric
0.9
numRules
10
removeAllMissingCols
False
significanceLevel
-1.0
upperBoundMinSupport
1.0
Open...
Save...
OK
Cancel
Information
NAME
weka.associations.Apriori
SYNOPSIS
Finds association rules.
OPTIONS
delta -- Iteratively decrease support by this factor. Reduces support until min support is reached or required number of rules has been generated.
lowerBoundMinSupport -- Lower bound for minimum support.
metricType -- Set the type of metric by which to rank rules. Confidence is the proportion of the examples covered by the premise that are also covered by the consequence. Lift is confidence divided by the proportion of all examples that are covered by the consequence. This is a measure of the importance of the association that is independent of support. Leverage is the proportion of additional examples covered by both the premise and consequence above those expected if the premise
Status
OK
Log
x 0


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

Weka Explorer

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

Associator

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

Start | Stop

Result list (right-click for

21:33:51 - Apriori

Associator output

```
               roots
               class
=== Associator model (full training set) ===


Apriori
=======

Minimum support: 0.8
Minimum metric <confidence>: 0.9
Number of cycles performed: 4

Generated sets of large itemsets:

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

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

Size of set of large itemsets L(3): 2

Best rules found:

 1. int-discolor=none 581 ==> sclerotia=absent 581    conf:(1)
 2. mycelium=absent int-discolor=none 575 ==> sclerotia=absent 575    conf:(1)
 3. leaves=abnorm sclerotia=absent 548 ==> mycelium=absent 547    conf:(1)
 4. sclerotia=absent 625 ==> mycelium=absent 619    conf:(0.99)
 5. int-discolor=none 581 ==> mycelium=absent sclerotia=absent 575    conf:(0.99)
 6. int-discolor=none sclerotia=absent 581 ==> mycelium=absent 575    conf:(0.99)
 7. int-discolor=none 581 ==> mycelium=absent 575    conf:(0.99)
 8. leaf-malf=absent 554 ==> mycelium=absent 548    conf:(0.99)
 9. mycelium=absent 639 ==> sclerotia=absent 619    conf:(0.97)
10. leaves=abnorm mycelium=absent 567 ==> sclerotia=absent 547    conf:(0.96)
```

Status

OK

Log

x 0

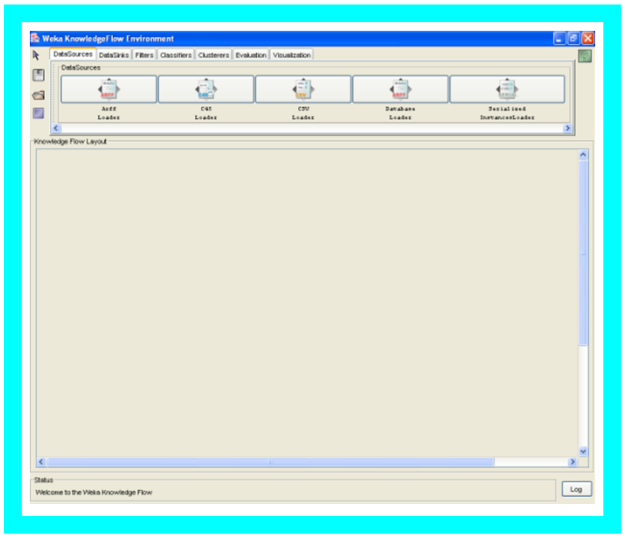
Data Mining with WEKA (Knowledge Flow GUI)

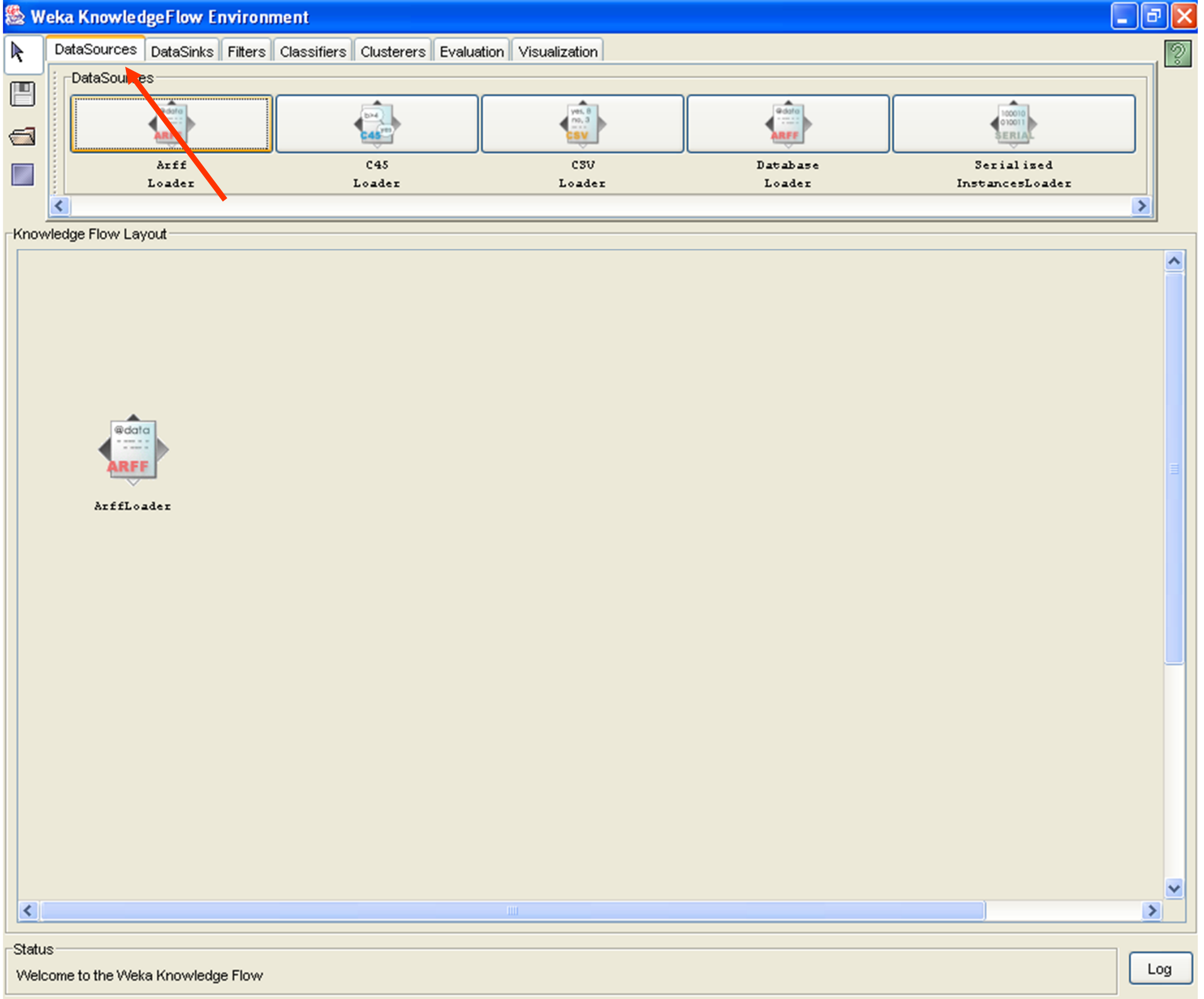
The Knowledge Flow GUI

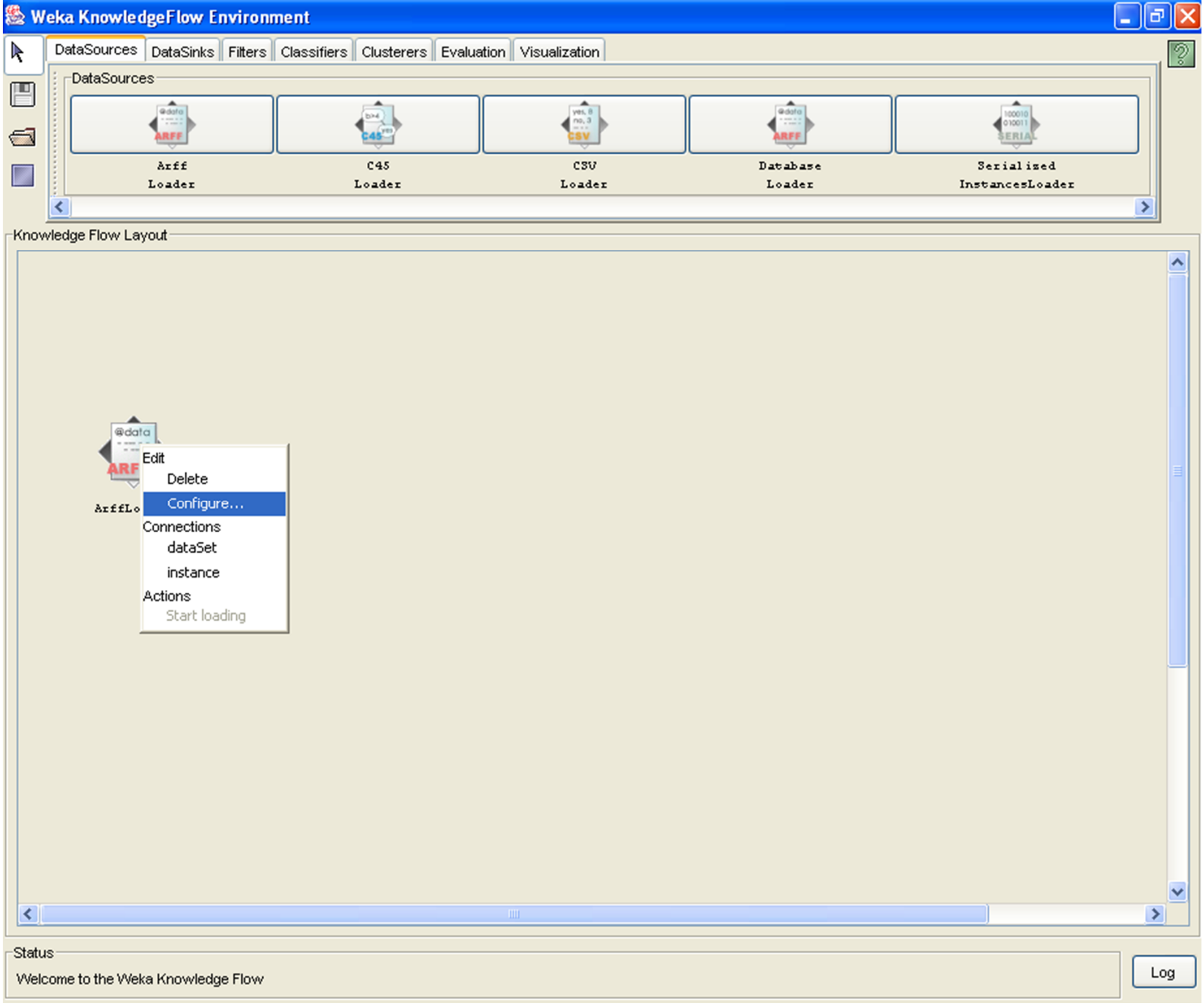
- Graphical user interface for WEKA
- Java-Beans-based interface for setting up and running machine learning experiments
- Data sources, classifiers, etc. are beans and can be connected graphically
- Data "flows" through components: e.g.,

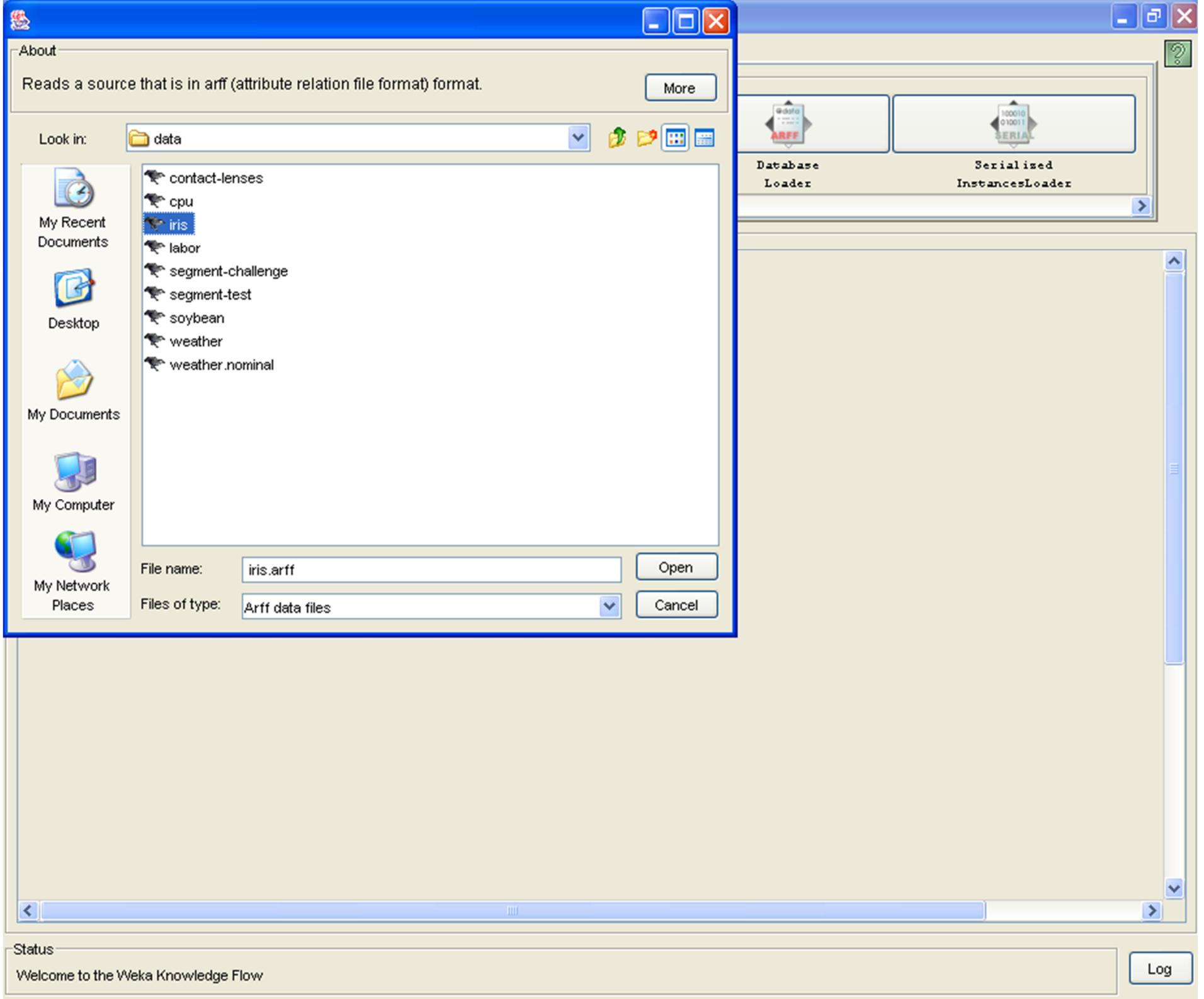
  "data source" -> "filter" -> "classifier" -> "evaluator"
- Layouts can be saved and loaded again later

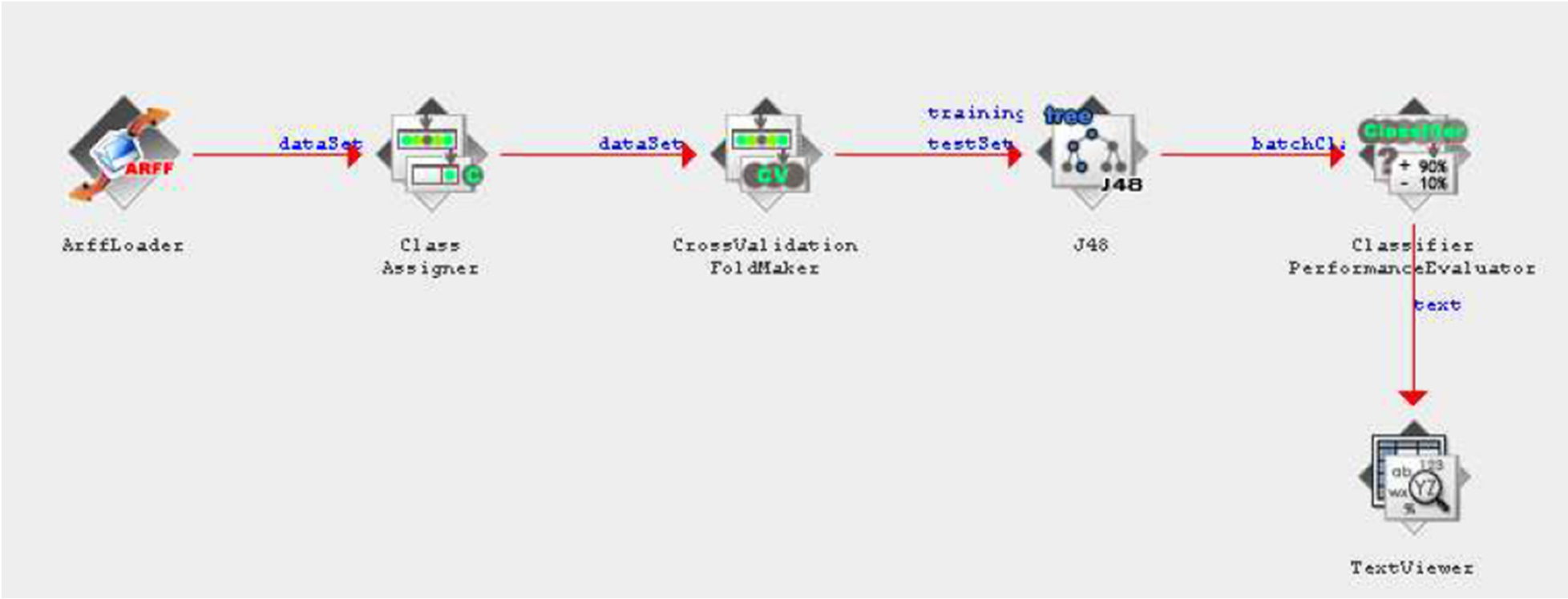

Weka GUI Chooser
Waikato Environment for
Knowledge Analysis
Version 3.4.5
(c) 1999 - 2005
University of Waikato
New Zealand
GUI
Simple CLI
Explorer
Experimenter
KnowledgeFlow

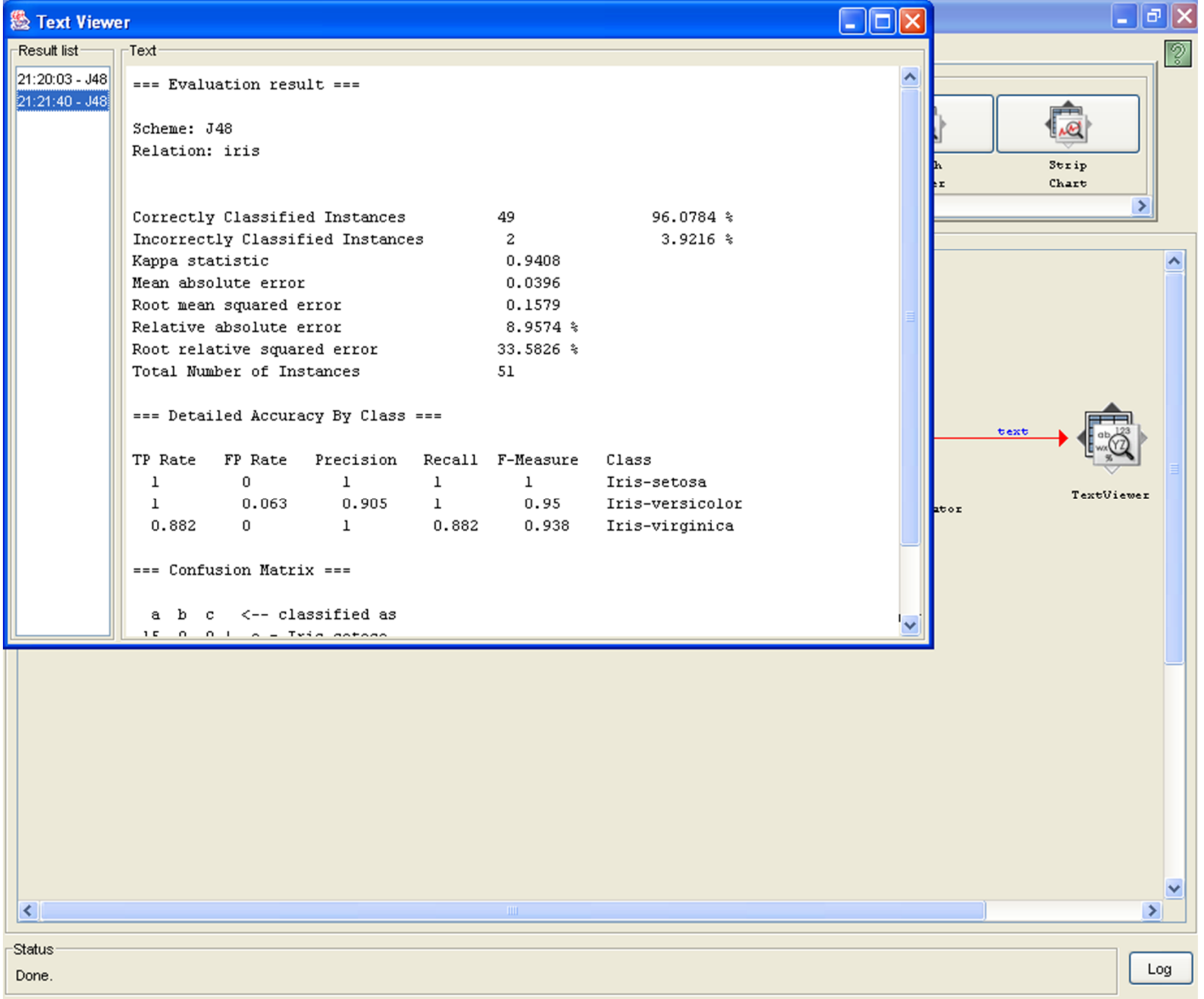



Weka KnowledgeFlow Environment
DataSources
DataSinks
Filters
Classifiers
Clusterers
Evaluation
Visualization
DataSources
Arff Loader
C45 Loader
CSV Loader
Database Loader
Serialized InstancesLoader
Knowledge Flow Layout
ArffLoader
Status
Welcome to the Weka Knowledge Flow
Log


Weka KnowledgeFlow Environment
DataSources
DataSinks
Filters
Classifiers
Clusterers
Evaluation
Visualization
DataSources
Arff Loader
C45 Loader
CSV Loader
Database Loader
Serialized InstancesLoader
Knowledge Flow Layout
ArffLo
Edit
Delete
Configure...
Connections
dataSet
instance
Actions
Start loading
Status
Welcome to the Weka Knowledge Flow
Log


About
Reads a source that is in arff (attribute relation file format) format.
More
Look in:
data
contact-lenses
cpu
iris
labor
segment-challenge
segment-test
soybean
weather
weather.nominal
My Recent Documents
Desktop
My Documents
My Computer
My Network Places
File name:
iris.arff
Open
Files of type:
Arff data files
Cancel
Database Loader
Serialized InstancesLoader
Status
Welcome to the Weka Knowledge Flow
Log


ARFF
dataSet
dataSet
training
testSet
J48
batchCl
Classifier
+ 90%
- 10%
text
ArffLoader
Class
Assigner
CrossValidation
FoldMaker
J48
Classifier
PerformanceEvaluator
TextViewer


Text Viewer
Result list
21:20:03 - J48
21:21:40 - J48
Text
=== Evaluation result ===
Scheme: J48
Relation: iris
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 8.9574 %
Root relative squared error 33.5826 %
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
Strip Chart
text
TextViewer
Status
Done.
Log