고려대학교 KOREA UNIVERSITY

Weka 실습 가이드

Human Intelligence

Machine Intelligence

Index

-Data Set

-Install

-Explorer

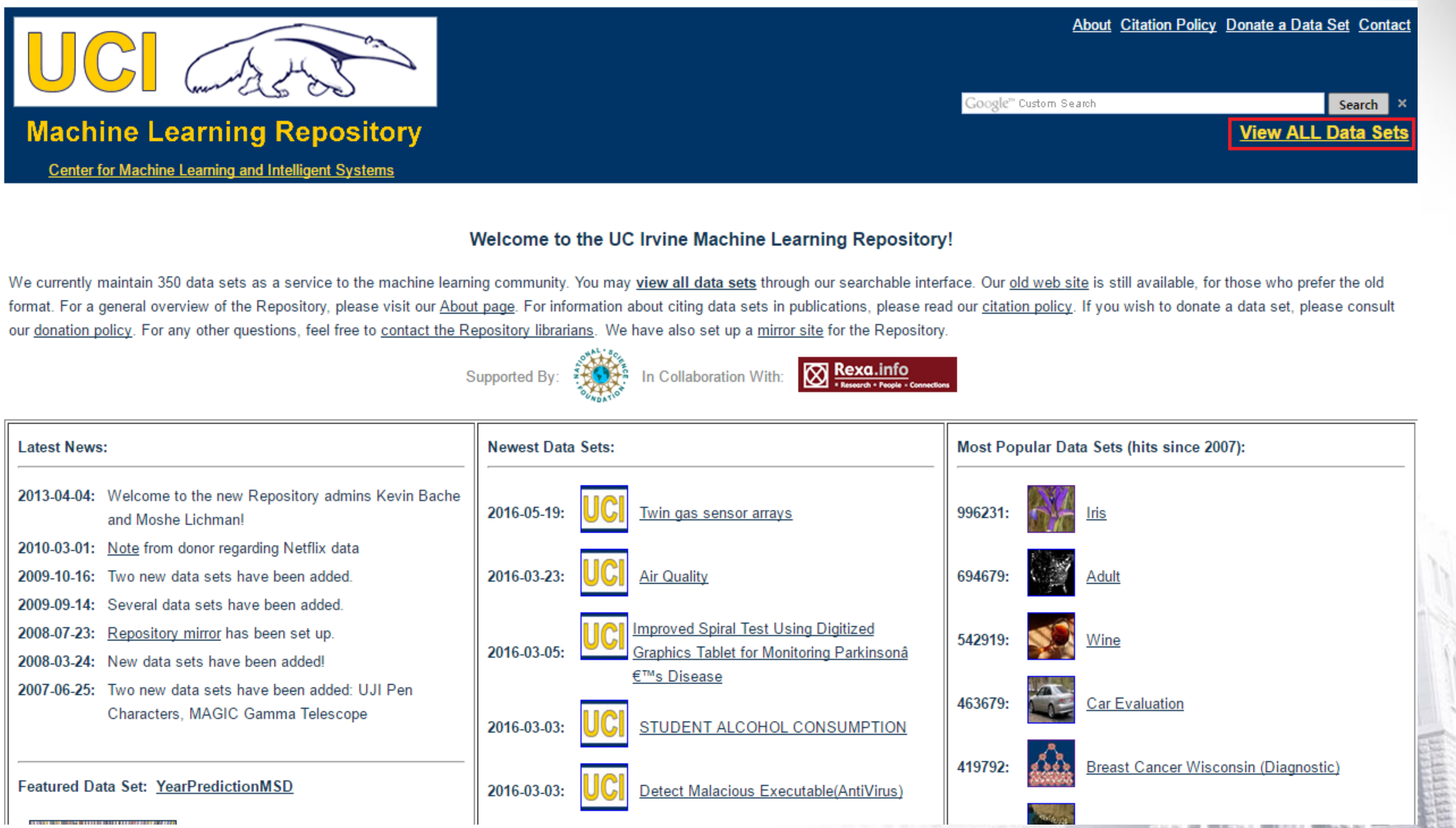
-Experimenter

-Knowledge flow

-WEKA Reference

Data Set

http://archive.ics.uci.edu/ml/



Welcome to the UC Irvine Machine Learning Repository!

We currently maintain 350 data sets as a service to the machine learning community. You may **view all data sets** through our searchable interface. Our old web site is still available, for those who prefer the old format. For a general overview of the Repository, please visit our About page. For information about citing data sets in publications, please read our citation policy. If you wish to donate a data set, please consult our donation policy. For any other questions, feel free to contact the Repository librarians. We have also set up a mirror site for the Repository.

Latest News:

2013-04-04:	Welcome to the new Repository admins Kevin Bache and Moshe Lichman!
2010-03-01:	Note from donor regarding Netflix data
2009-10-16:	Two new data sets have been added.
2009-09-14:	Several data sets have been added.
2008-07-23:	Repository mirror has been set up.
2008-03-24:	New data sets have been added!
2007-06-25:	Two new data sets have been added: UJI Pen Characters, MAGIC Gamma Telescope

Featured Data Set: YearPredictionMSD

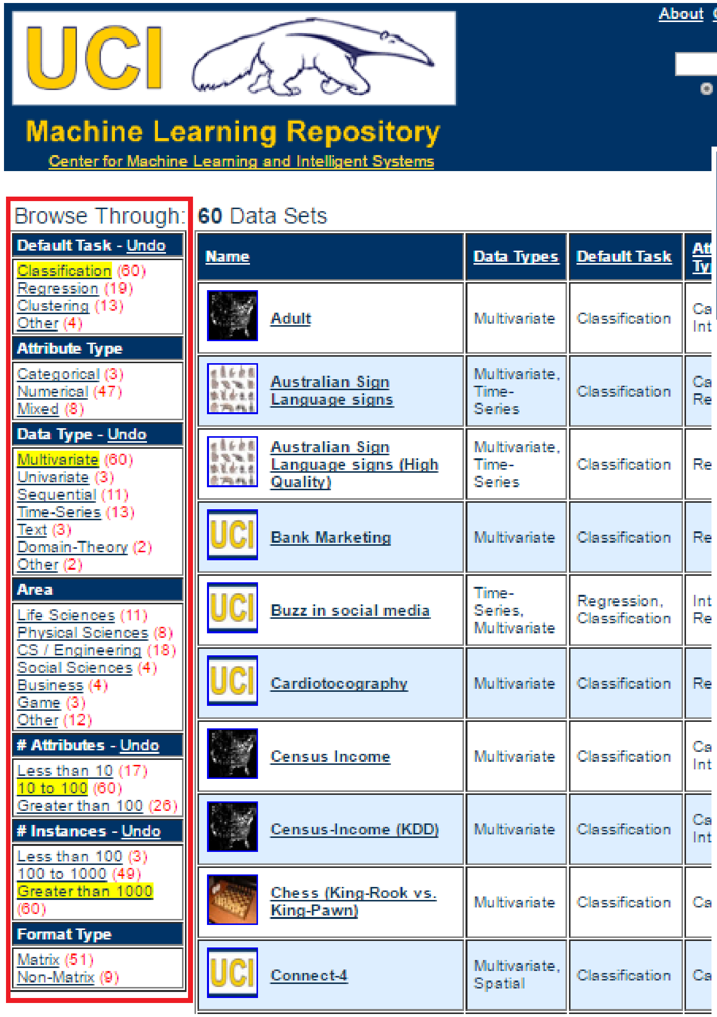
Newest Data Sets:

2016-05-19:	Twin gas sensor arrays
2016-03-23:	Air Quality
2016-03-05:	Improved Spiral Test Using Digitized Graphics Tablet for Monitoring Parkinsonâ€™s Disease
2016-03-03:	STUDENT ALCOHOL CONSUMPTION
2016-03-03:	Detect Malacious Executable(AntiVirus)

Most Popular Data Sets (hits since 2007):

996231:	Iris
694679:	Adult
542919:	Wine
463679:	Car Evaluation
419792:	Breast Cancer Wisconsin (Diagnostic)

Data Set



About Citation Policy Donate a Data Set Contact

Search

○ Repository ● Web Google

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Browse Through: **60** Data Sets

Default Task - Undo
Classification (60)
Regression (19)
Clustering (13)
Other (4)

Attribute Type
Categorical (3)
Numerical (47)
Mixed (8)

Data Type - Undo
Multivariate (60)
Univariate (3)
Sequential (11)
Time-Series (13)
Text (3)
Domain-Theory (2)
Other (2)

Area
Life Sciences (11)
Physical Sciences (8)
CS / Engineering (18)
Social Sciences (4)
Business (4)
Game (3)
Other (12)

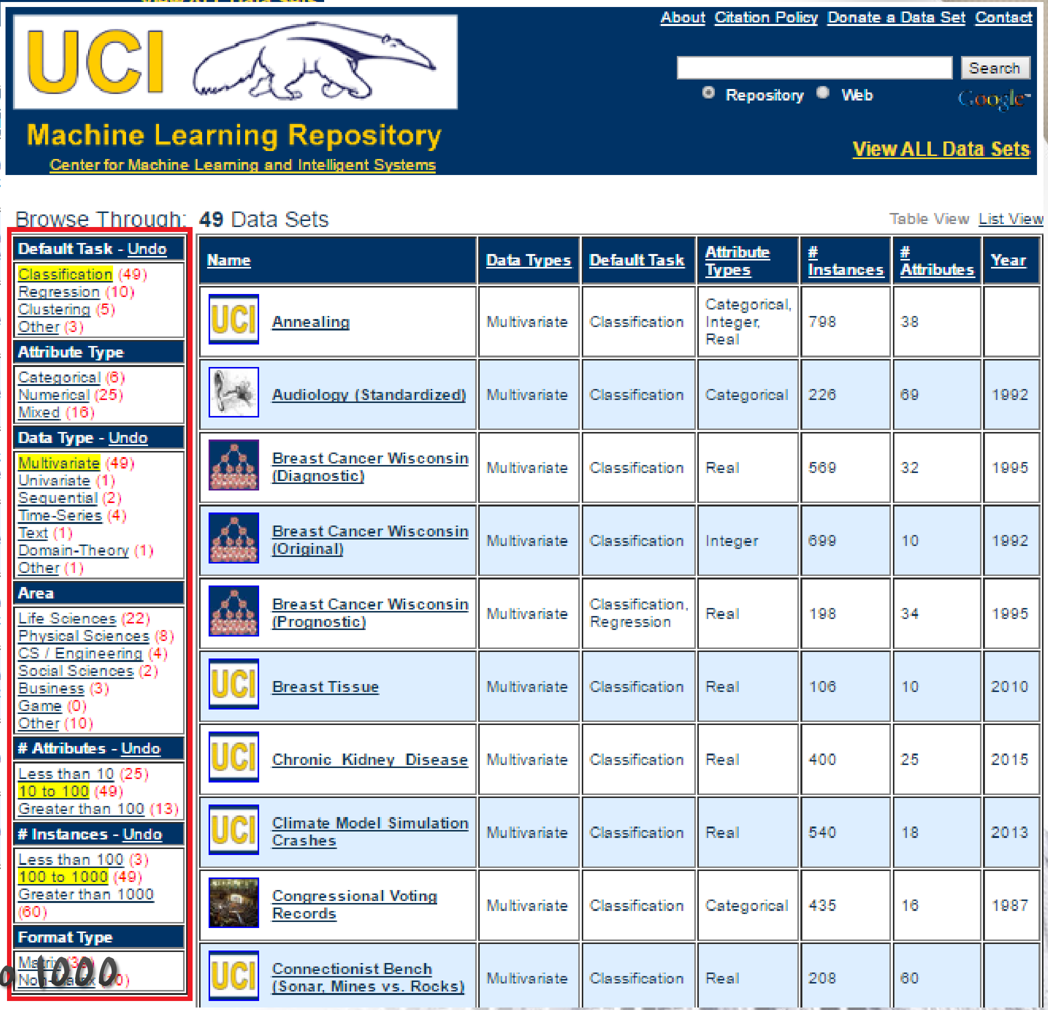
# Attributes - Undo
Less than 10 (17)
10 to 100 (60)
Greater than 100 (26)

# Instances - Undo
Less than 100 (3)
100 to 1000 (49)
Greater than 1000 (60)

Format Type
Matrix (51)
Non-Matrix (9)

Name	Data Types	Default Task	At Ty
Adult	Multivariate	Classification	Ca Int
Australian Sign Language signs	Multivariate, Time-Series	Classification	Ca Re
Australian Sign Language signs (High Quality)	Multivariate, Time-Series	Classification	Re
Bank Marketing	Multivariate	Classification	Re
Buzz in social media	Time-Series, Multivariate	Regression, Classification	Int Re
Cardiotocography	Multivariate	Classification	Re
Census Income	Multivariate	Classification	Ca Int
Census-Income (KDD)	Multivariate	Classification	Ca Int
Chess (King-Rook vs. King-Pawn)	Multivariate	Classification	Ca
Connect-4	Multivariate, Spatial	Classification	Ca

Data는 csv로 저장



About Citation Policy Donate a Data Set Contact

Search

○ Repository ● Web Google

UCI Machine Learning Repository

Center for Machine Learning and Intelligent Systems

View ALL Data Sets

Browse Through: **49** Data Sets

Table View List View

Default Task - Undo
Classification (49)
Regression (10)
Clustering (5)
Other (3)

Attribute Type
Categorical (6)
Numerical (25)
Mixed (16)

Data Type - Undo
Multivariate (49)
Univariate (1)
Sequential (2)
Time-Series (4)
Text (1)
Domain-Theory (1)
Other (1)

Area
Life Sciences (22)
Physical Sciences (8)
CS / Engineering (4)
Social Sciences (2)
Business (3)
Game (0)
Other (10)

# Attributes - Undo
Less than 10 (25)
10 to 100 (49)
Greater than 100 (13)

# Instances - Undo
Less than 100 (3)
100 to 1000 (49)
Greater than 1000 (60)

Format Type
Matrix ([illegible])
Non-Matrix ([illegible])

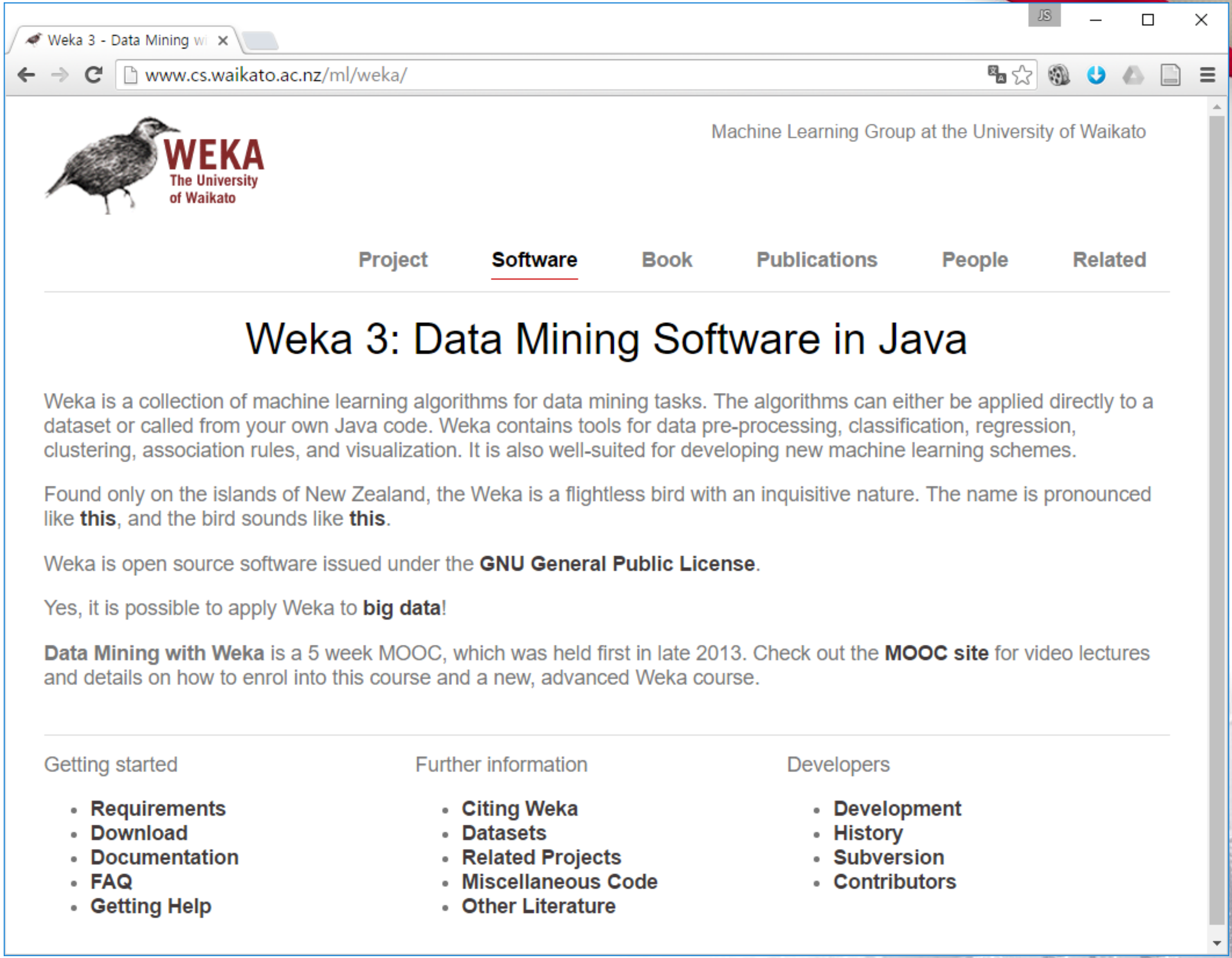
Name	Data Types	Default Task	Attribute Types	# Instances	# Attributes	Year
Annealing	Multivariate	Classification	Categorical, Integer, Real	798	38	
Audiology (Standardized)	Multivariate	Classification	Categorical	226	69	1992
Breast Cancer Wisconsin (Diagnostic)	Multivariate	Classification	Real	569	32	1995
Breast Cancer Wisconsin (Original)	Multivariate	Classification	Integer	699	10	1992
Breast Cancer Wisconsin (Prognostic)	Multivariate	Classification, Regression	Real	198	34	1995
Breast Tissue	Multivariate	Classification	Real	106	10	2010
Chronic_Kidney_Disease	Multivariate	Classification	Real	400	25	2015
Climate Model Simulation Crashes	Multivariate	Classification	Real	540	18	2013
Congressional Voting Records	Multivariate	Classification	Categorical	435	16	1987
Connectionist Bench (Sonar, Mines vs. Rocks)	Multivariate	Classification	Real	208	60	

Data 선택조건

1. Classification 3. # of Attribute 10 to 1000

2. Multivariate 4. # of Instance (100 to 1000, Greater than 1,000)

Install



Machine Learning Group at the University of Waikato

Project Software Book Publications People Related

Weka 3: Data Mining Software in Java

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

Found only on the islands of New Zealand, the Weka is a flightless bird with an inquisitive nature. The name is pronounced like **this**, and the bird sounds like **this**.

Weka is open source software issued under the **GNU General Public License**.

Yes, it is possible to apply Weka to **big data**!

Data Mining with Weka is a 5 week MOOC, which was held first in late 2013. Check out the **MOOC site** for video lectures and details on how to enrol into this course and a new, advanced Weka course.

Getting started

- **Requirements**
- **Download**
- **Documentation**
- **FAQ**
- **Getting Help**

Further information

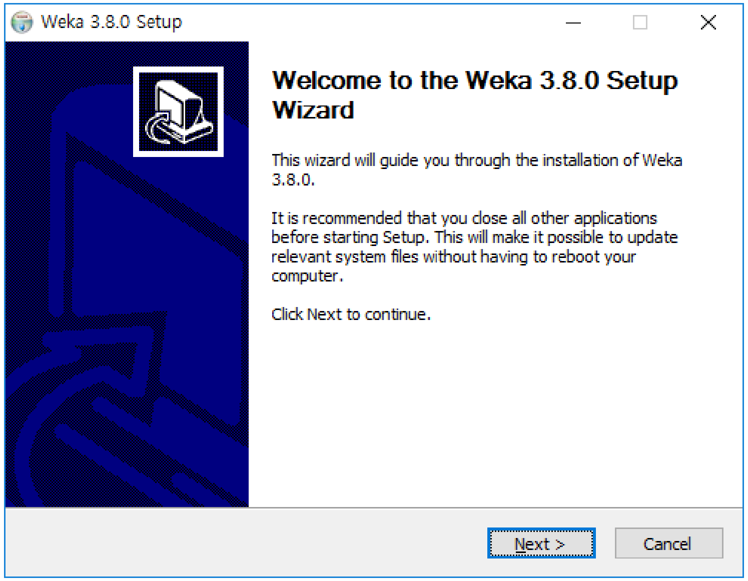
- **Citing Weka**
- **Datasets**
- **Related Projects**
- **Miscellaneous Code**
- **Other Literature**

Developers

- **Development**
- **History**
- **Subversion**
- **Contributors**

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

Install

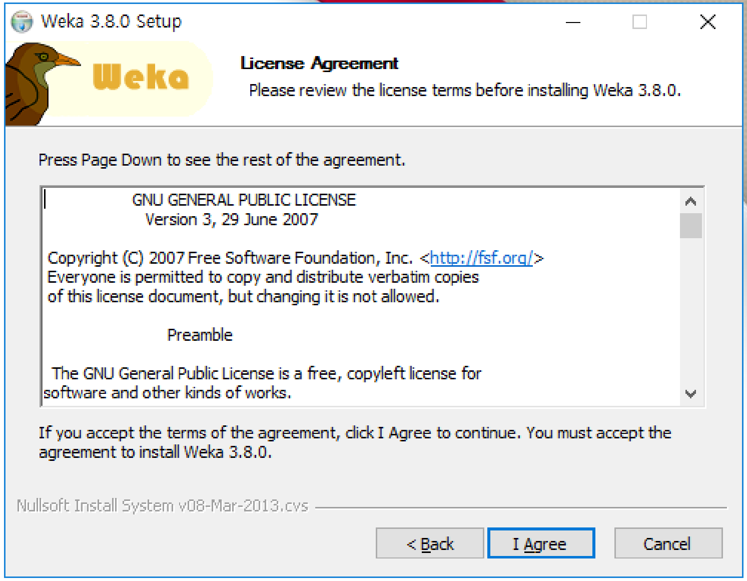

Weka 3.8.0 Setup

Welcome to the Weka 3.8.0 Setup Wizard

This wizard will guide you through the installation of Weka 3.8.0.

It is recommended that you close all other applications before starting Setup. This will make it possible to update relevant system files without having to reboot your computer.

Click Next to continue.


Weka 3.8.0 Setup

License Agreement

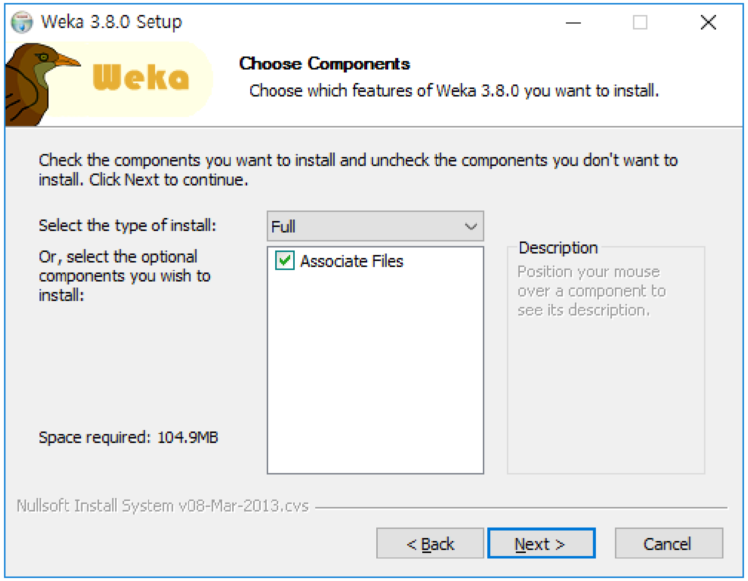
Please review the license terms before installing Weka 3.8.0.

Press Page Down to see the rest of the agreement.

```
              GNU GENERAL PUBLIC LICENSE
                 Version 3, 29 June 2007

 Copyright (C) 2007 Free Software Foundation, Inc. <http://fsf.org/>
 Everyone is permitted to copy and distribute verbatim copies
 of this license document, but changing it is not allowed.

                    Preamble

  The GNU General Public License is a free, copyleft license for
software and other kinds of works.
```

If you accept the terms of the agreement, click I Agree to continue. You must accept the agreement to install Weka 3.8.0.

Nullsoft Install System v08-Mar-2013.cvs


Weka 3.8.0 Setup

Choose Components

Choose which features of Weka 3.8.0 you want to install.

Check the components you want to install and uncheck the components you don't want to install. Click Next to continue.

Select the type of install: Full

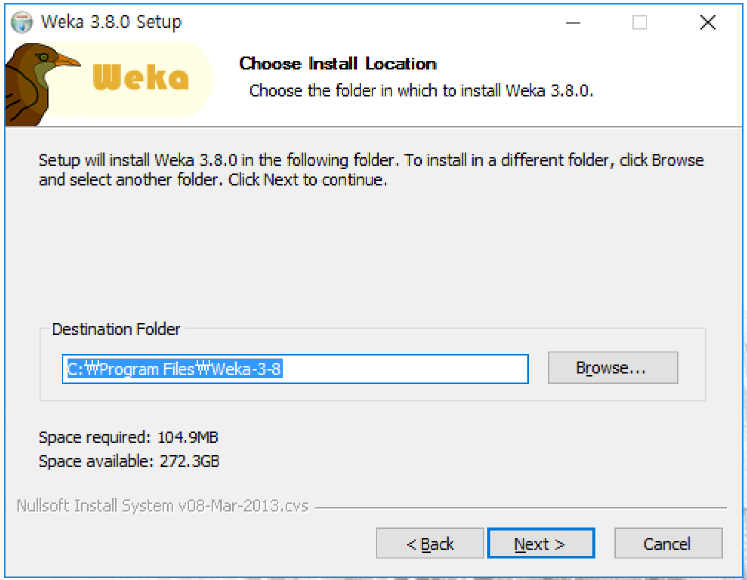
Or, select the optional components you wish to install: ☑ Associate Files

Description

Position your mouse over a component to see its description.

Space required: 104.9MB

Nullsoft Install System v08-Mar-2013.cvs


Weka 3.8.0 Setup

Choose Install Location

Choose the folder in which to install Weka 3.8.0.

Setup will install Weka 3.8.0 in the following folder. To install in a different folder, click Browse and select another folder. Click Next to continue.

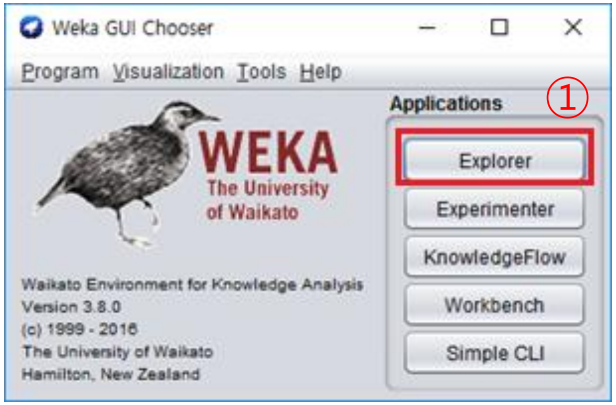
Destination Folder

C:₩Program Files₩Weka-3-8

Space required: 104.9MB
Space available: 272.3GB

Nullsoft Install System v08-Mar-2013.cvs

Explorer



Weka GUI Chooser

Program Visualization Tools Help

Applications

① Explorer

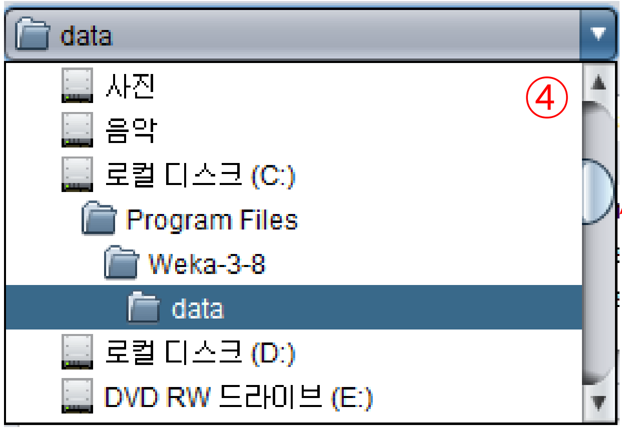
Experimenter

KnowledgeFlow

Workbench

Simple CLI

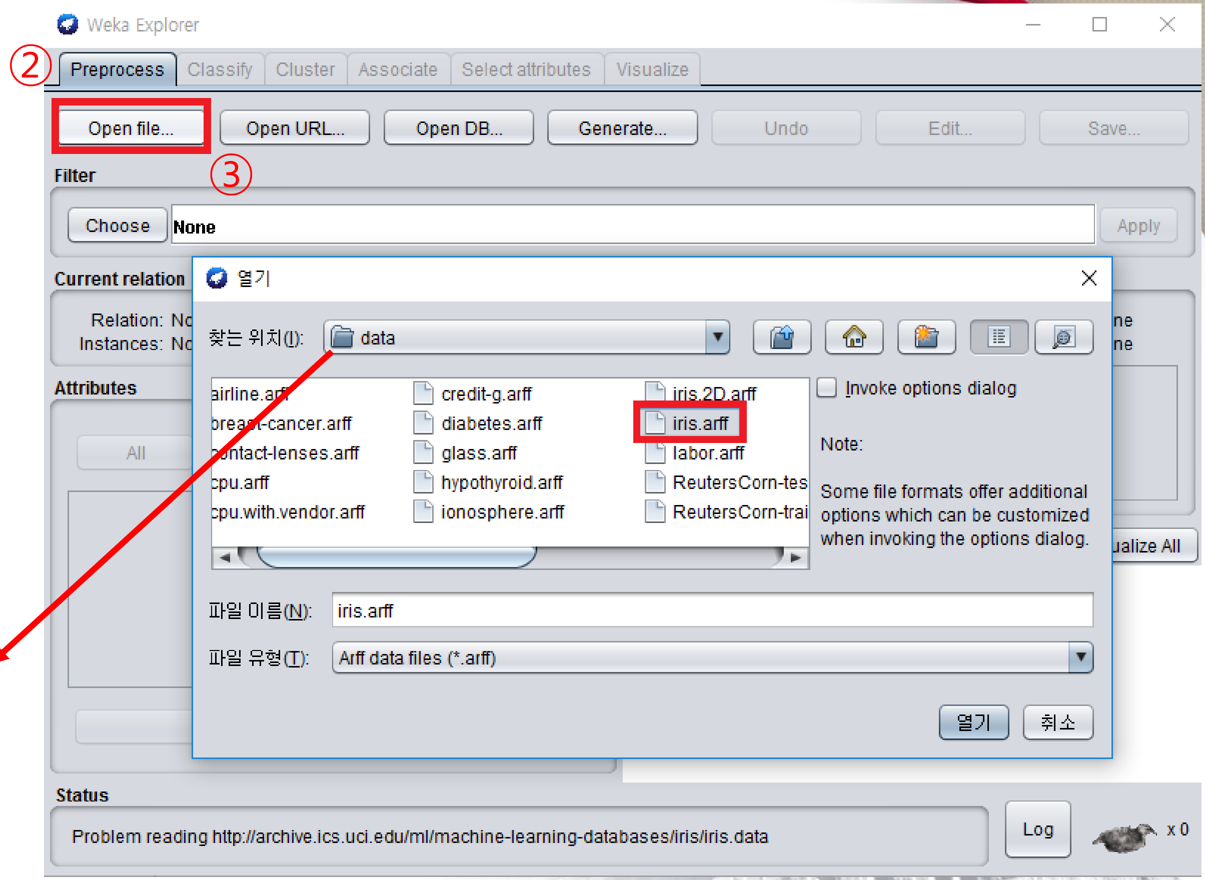
Waikato Environment for Knowledge Analysis
Version 3.8.0
(c) 1999 - 2016
The University of Waikato
Hamilton, New Zealand



data

④

- 사진
- 음악
- 로컬 디스크 (C:)
  - Program Files
    - Weka-3-8
      - data
- 로컬 디스크 (D:)
- DVD RW 드라이브 (E:)



Weka Explorer

② Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

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

③

Filter

Choose None Apply

Current relation

Relation: No
Instances: No

Attributes

All

열기

찾는 위치(I): data

airline.arff	credit-g.arff	iris.2D.arff
breast-cancer.arff	diabetes.arff	iris.arff
contact-lenses.arff	glass.arff	labor.arff
cpu.arff	hypothyroid.arff	ReutersCorn-tes
cpu.with.vendor.arff	ionosphere.arff	ReutersCorn-trai

□ Invoke options dialog

Note:

Some file formats offer additional options which can be customized when invoking the options dialog.

파일 이름(N): iris.arff

파일 유형(T): Arff data files (*.arff)

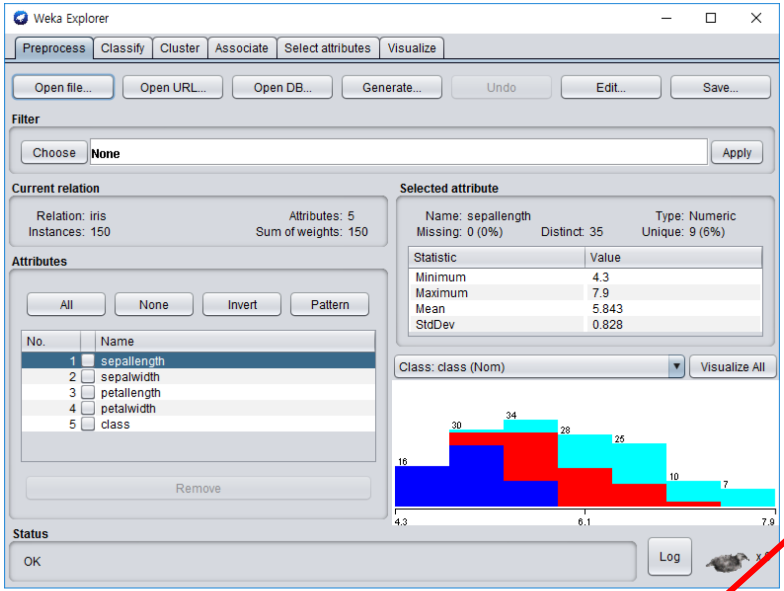
열기 | 취소

Status

Problem reading http://archive.ics.uci.edu/ml/machine-learning-databases/iris/iris.data

Log x 0

Explorer



Weka Explorer

Filter: None

Current relation

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

Selected attribute

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

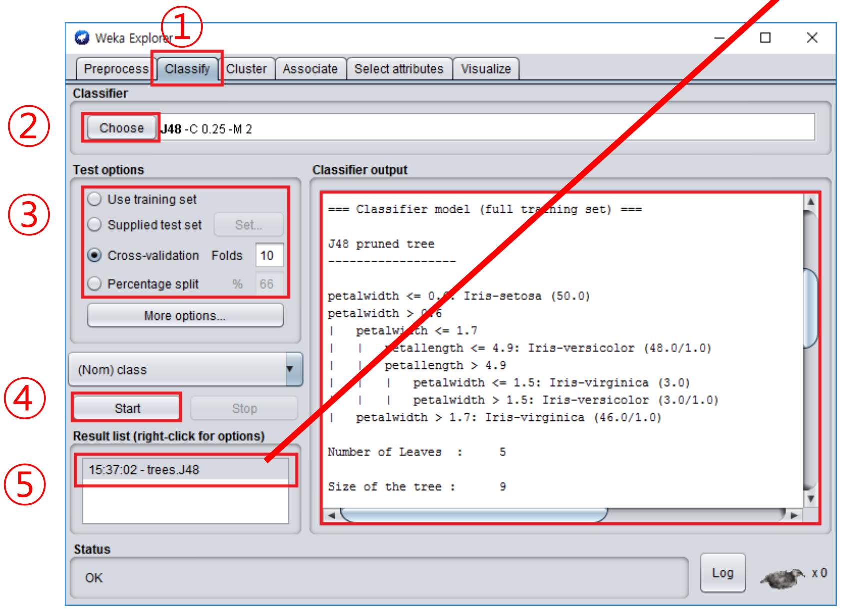
Statistic	Value
Minimum	4.3
Maximum	7.9
Mean	5.843
StdDev	0.828

Attributes

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

Class: class (Nom)

Status
OK



Weka Explorer

Classifier: J48 -C 0.25 -M 2

Test options

○ Use training set
○ Supplied test set
● Cross-validation Folds 10
○ Percentage split % 66

(Nom) class

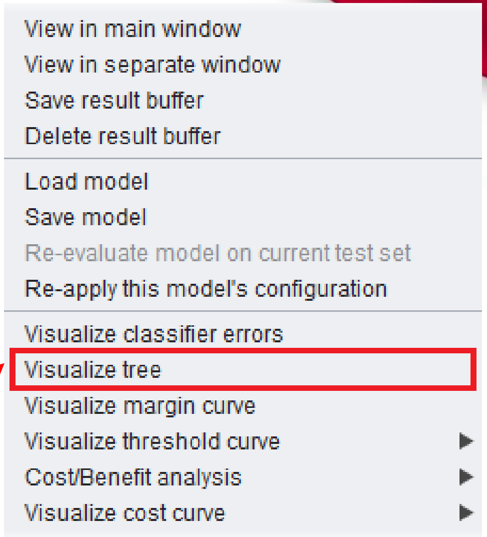
Result list (right-click for options)

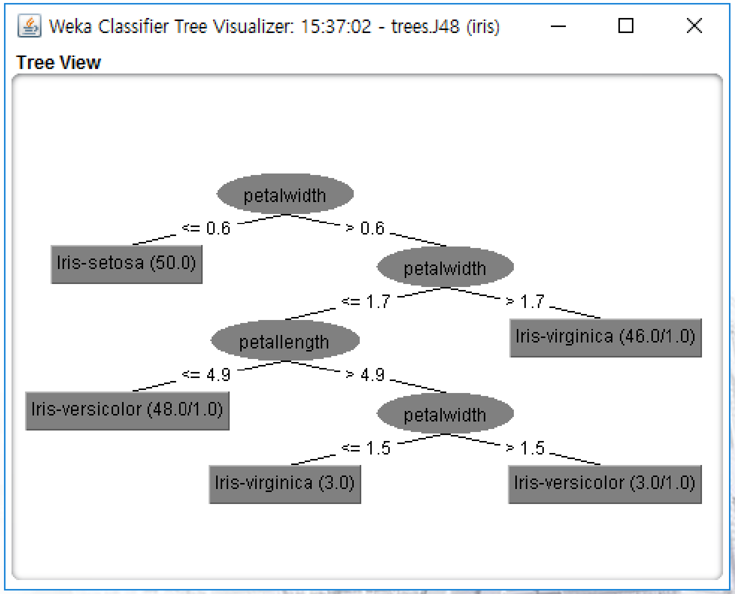
15:37:02 - trees.J48

Classifier output

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

J48 pruned tree
------------------

petalwidth <= 0.6: Iris-setosa (50.0)
petalwidth > 0.6
|   petalwidth <= 1.7
|   |   petallength <= 4.9: Iris-versicolor (48.0/1.0)
|   |   petallength > 4.9
|   |   |   petalwidth <= 1.5: Iris-virginica (3.0)
|   |   |   petalwidth > 1.5: Iris-versicolor (3.0/1.0)
|   petalwidth > 1.7: Iris-virginica (46.0/1.0)

Number of Leaves  :     5

Size of the tree :      9
```

Status
OK

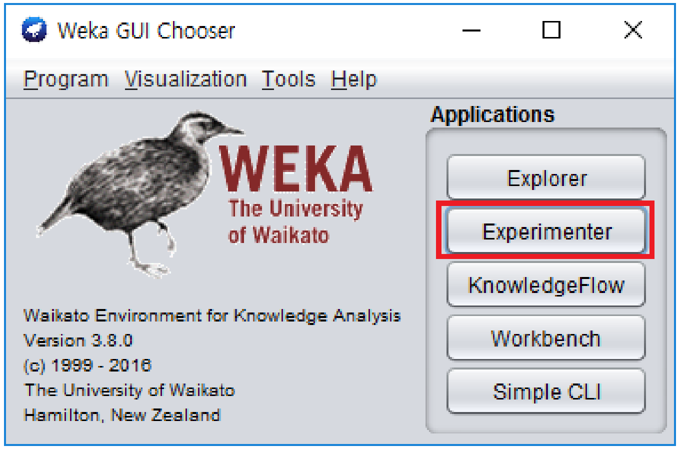




Weka Classifier Tree Visualizer: 15:37:02 - trees.J48 (iris)

Tree View

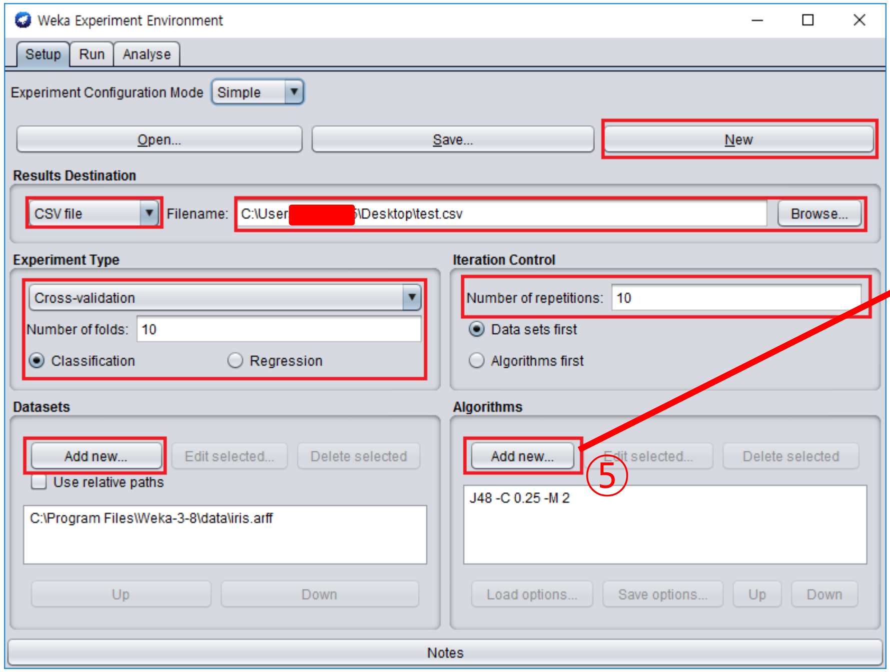
Experimenter



Weka GUI Chooser

Waikato Environment for Knowledge Analysis
Version 3.8.0
(c) 1999 - 2016
The University of Waikato
Hamilton, New Zealand

①



Weka Experiment Environment

Setup Run Analyse
Open... Save... New
Browse...
Add new... Edit selected... Delete selected
Up Down
Load options... Save options... Up Down
Notes

Experiment Configuration Mode: Simple

②

Results Destination

③ CSV file ▼ Filename: C:\User[redacted]\Desktop\test.csv ④

Experiment Type

④ Cross-validation

Number of folds: 10

● Classification ○ Regression

Iteration Control

Number of repetitions: 10

● Data sets first

○ Algorithms first

Datasets

④ Add new...

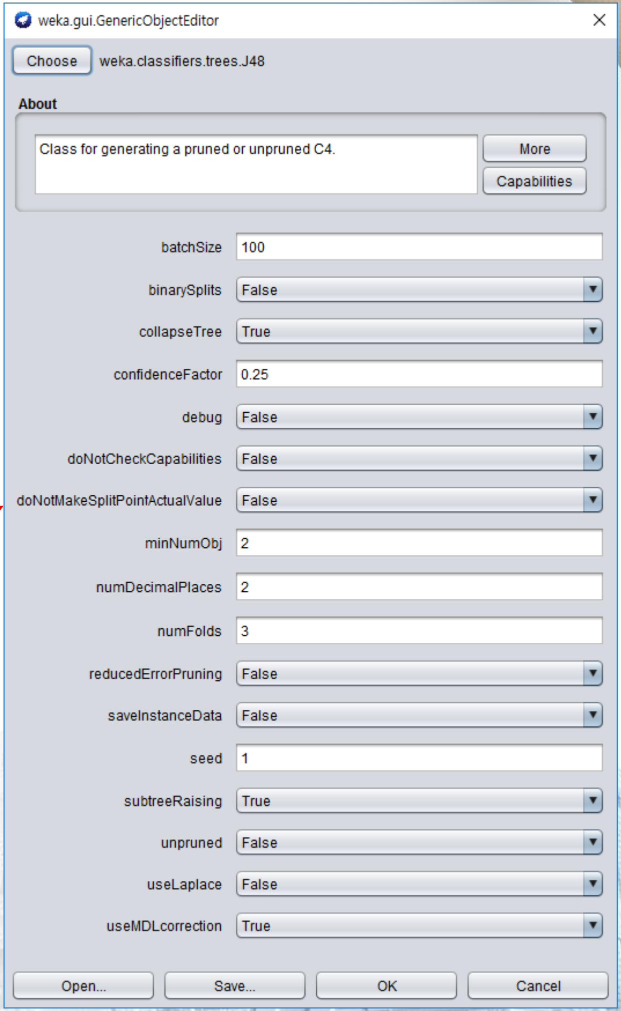
□ Use relative paths

C:\Program Files\Weka-3-8\data\iris.arff

Algorithms

Add new... ⑤

J48 -C 0.25 -M 2



weka.gui.GenericObjectEditor

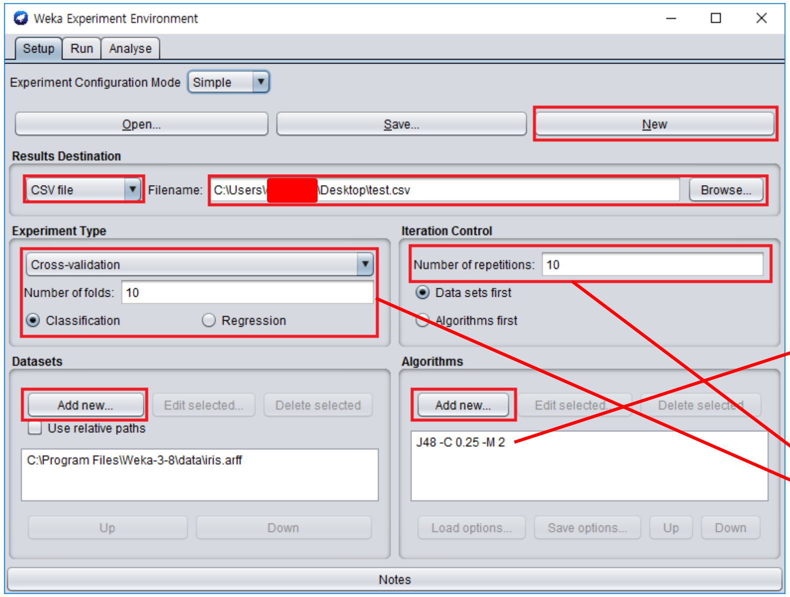
weka.classifiers.trees.J48

About

Class for generating a pruned or unpruned C4.

batchSize	100
binarySplits	False
collapseTree	True
confidenceFactor	0.25
debug	False
doNotCheckCapabilities	False
doNotMakeSplitPointActualValue	False
minNumObj	2
numDecimalPlaces	2
numFolds	3
reducedErrorPruning	False
saveInstanceData	False
seed	1
subtreeRaising	True
unpruned	False
useLaplace	False
useMDLcorrection	True

Experimenter



Weka Experiment Environment

Setup | Run | Analyse

Experiment Configuration Mode: Simple

Open... | Save... | New

Results Destination

CSV file — Filename: C:\Users\[redacted]\Desktop\test.csv — Browse...

Experiment Type

Cross-validation

Number of folds: 10

● Classification ○ Regression

Iteration Control

Number of repetitions: 10

● Data sets first

○ Algorithms first

Datasets

Add new... | Edit selected... | Delete selected

□ Use relative paths

C:\Program Files\Weka-3-8\data\iris.arff

Up | Down

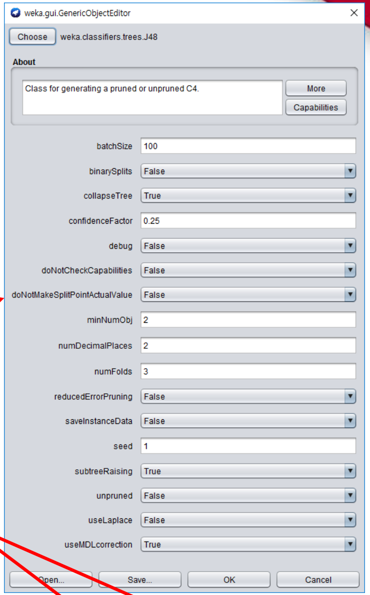
Algorithms

Add new... | Edit selected... | Delete selected

J48 -C 0.25 -M 2

Load options... | Save options... | Up | Down

Notes



weka.gui.GenericObjectEditor

Choose — weka.classifiers.trees.J48

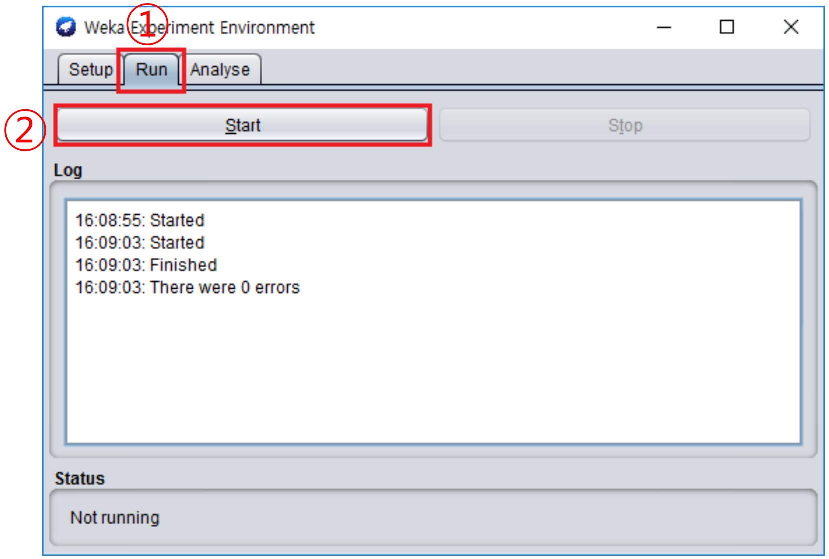
About

Class for generating a pruned or unpruned C4.

More | Capabilities

batchSize	100
binarySplits	False
collapseTree	True
confidenceFactor	0.25
debug	False
doNotCheckCapabilities	False
doNotMakeSplitPointActualValue	False
minNumObj	2
numDecimalPlaces	2
numFolds	3
reducedErrorPruning	False
saveInstanceData	False
seed	1
subtreeRaising	True
unpruned	False
useLaplace	False
useMDLcorrection	True

Open... | Save... | OK | Cancel



①

②

Weka Experiment Environment

Setup | Run | Analyse

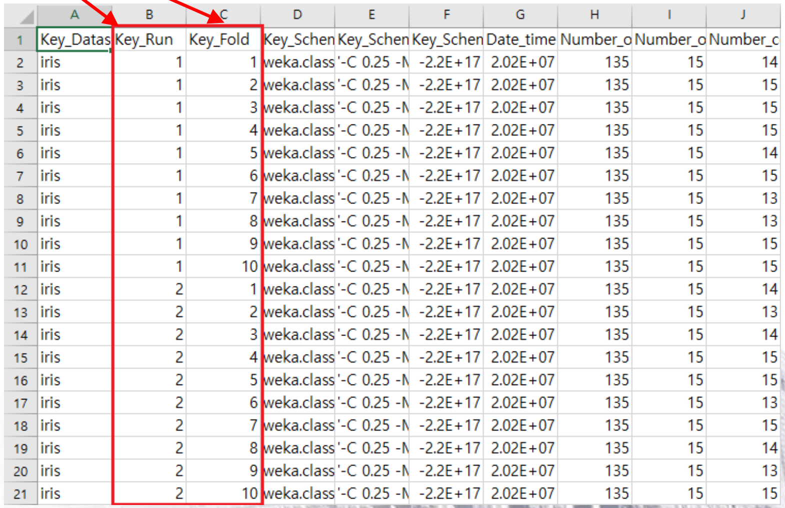
Start | Stop

Log

```
16:08:55: Started
16:09:03: Started
16:09:03: Finished
16:09:03: There were 0 errors
```

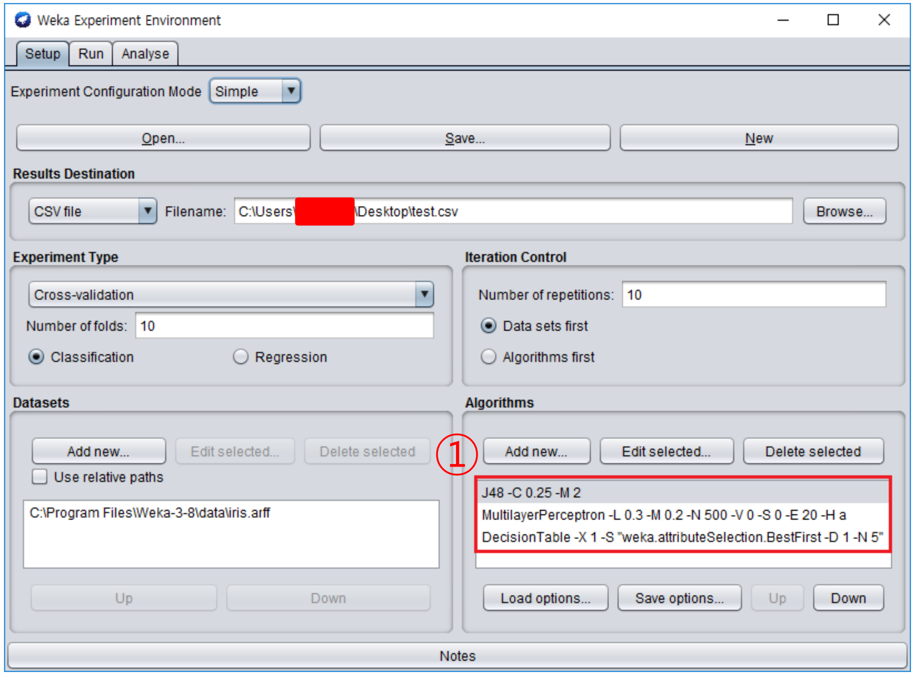
Status

Not running



	A	B	C	D	E	F	G	H	I	J
1	Key_Datas	Key_Run	Key_Fold	Key_Schen	Key_Schen	Key_Schen	Date_time	Number_o	Number_o	Number_c
2	iris	1	1	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	14
3	iris	1	2	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
4	iris	1	3	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
5	iris	1	4	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
6	iris	1	5	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	14
7	iris	1	6	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
8	iris	1	7	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	13
9	iris	1	8	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	13
10	iris	1	9	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
11	iris	1	10	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
12	iris	2	1	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	14
13	iris	2	2	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	13
14	iris	2	3	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	14
15	iris	2	4	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
16	iris	2	5	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
17	iris	2	6	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	13
18	iris	2	7	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15
19	iris	2	8	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	14
20	iris	2	9	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	13
21	iris	2	10	weka.class	'-C 0.25 -N	-2.2E+17	2.02E+07	135	15	15

Experimenter – 알고리즘끼리 비교



Weka Experiment Environment

Experiment Configuration Mode: Simple

Results Destination

CSV file — Filename: C:\Users\[redacted]\Desktop\test.csv

Experiment Type

Cross-validation

Number of folds: 10

● Classification ○ Regression

Iteration Control

Number of repetitions: 10

● Data sets first

○ Algorithms first

Datasets

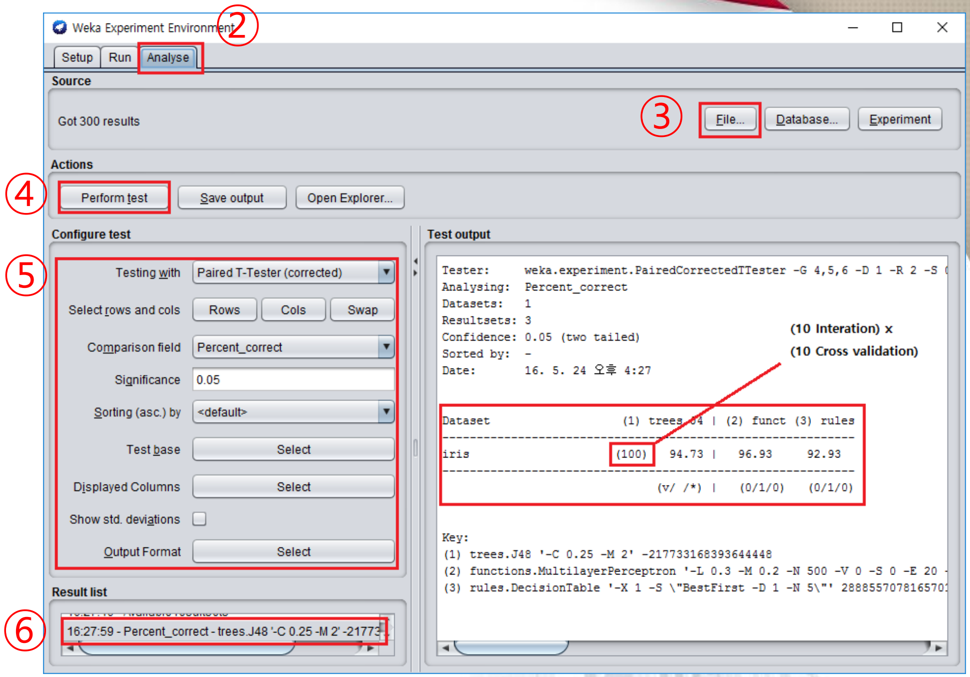
□ Use relative paths

C:\Program Files\Weka-3-8\data\iris.arff

Algorithms

①

```
J48 -C 0.25 -M 2
MultilayerPerceptron -L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 -H a
DecisionTable -X 1 -S "weka.attributeSelection.BestFirst -D 1 -N 5"
```



Weka Experiment Environment ②

Source

Got 300 results

③

Actions

④

Configure test

⑤

Testing with: Paired T-Tester (corrected)

Select rows and cols: Rows | Cols | Swap

Comparison field: Percent_correct

Significance: 0.05

Sorting (asc.) by: <default>

Test base: Select

Displayed Columns: Select

Show std. deviations: □

Output Format: Select

Result list

⑥

```
16:27:59 - Percent_correct - trees.J48 '-C 0.25 -M 2' -21773
```

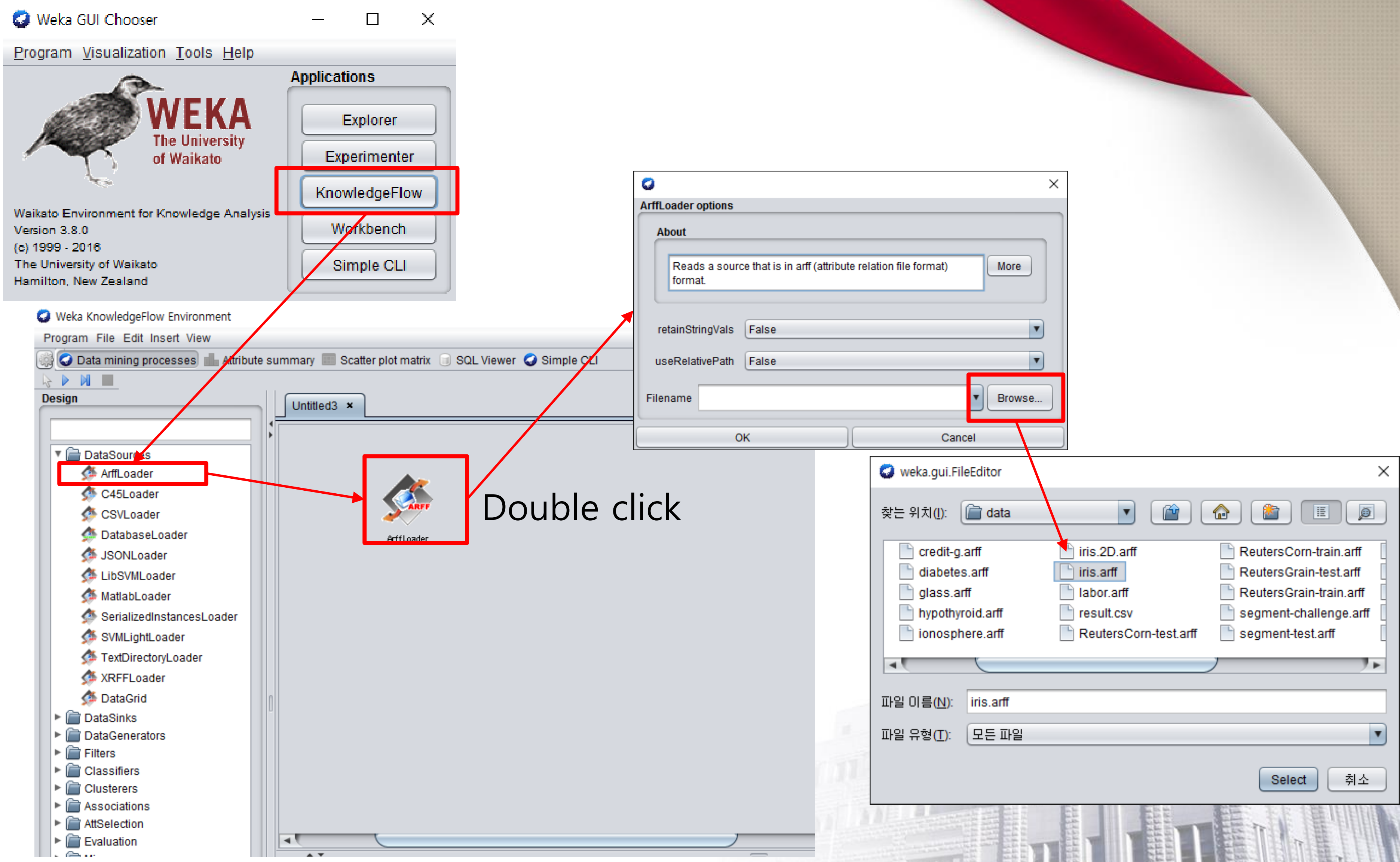
Test output

```
Tester:     weka.experiment.PairedCorrectedTTester -G 4,5,6 -D 1 -R 2 -S 
Analysing:  Percent_correct
Datasets:   1
Resultsets: 3
Confidence: 0.05 (two tailed)
Sorted by:  -
Date:       16. 5. 24 오후 4:27


Dataset                   (1) trees.J4 | (2) funct (3) rules
------------------------------------------------------------
iris                      (100)  94.73 |   96.93     92.93
------------------------------------------------------------
                                (v/ /*) |   (0/1/0)   (0/1/0)


Key:
(1) trees.J48 '-C 0.25 -M 2' -217733168393644448
(2) functions.MultilayerPerceptron '-L 0.3 -M 0.2 -N 500 -V 0 -S 0 -E 20 
(3) rules.DecisionTable '-X 1 -S \"BestFirst -D 1 -N 5\"' 288855707816570
```

(10 Interation) x

(10 Cross validation)

Knowledge flow



Weka GUI Chooser

Waikato Environment for Knowledge Analysis
Version 3.8.0
(c) 1999 - 2016
The University of Waikato
Hamilton, New Zealand

ArffLoader options

About: Reads a source that is in arff (attribute relation file format) format.

retainStringVals: False

useRelativePath: False

Filename:

weka.gui.FileEditor

찾는 위치(I): data

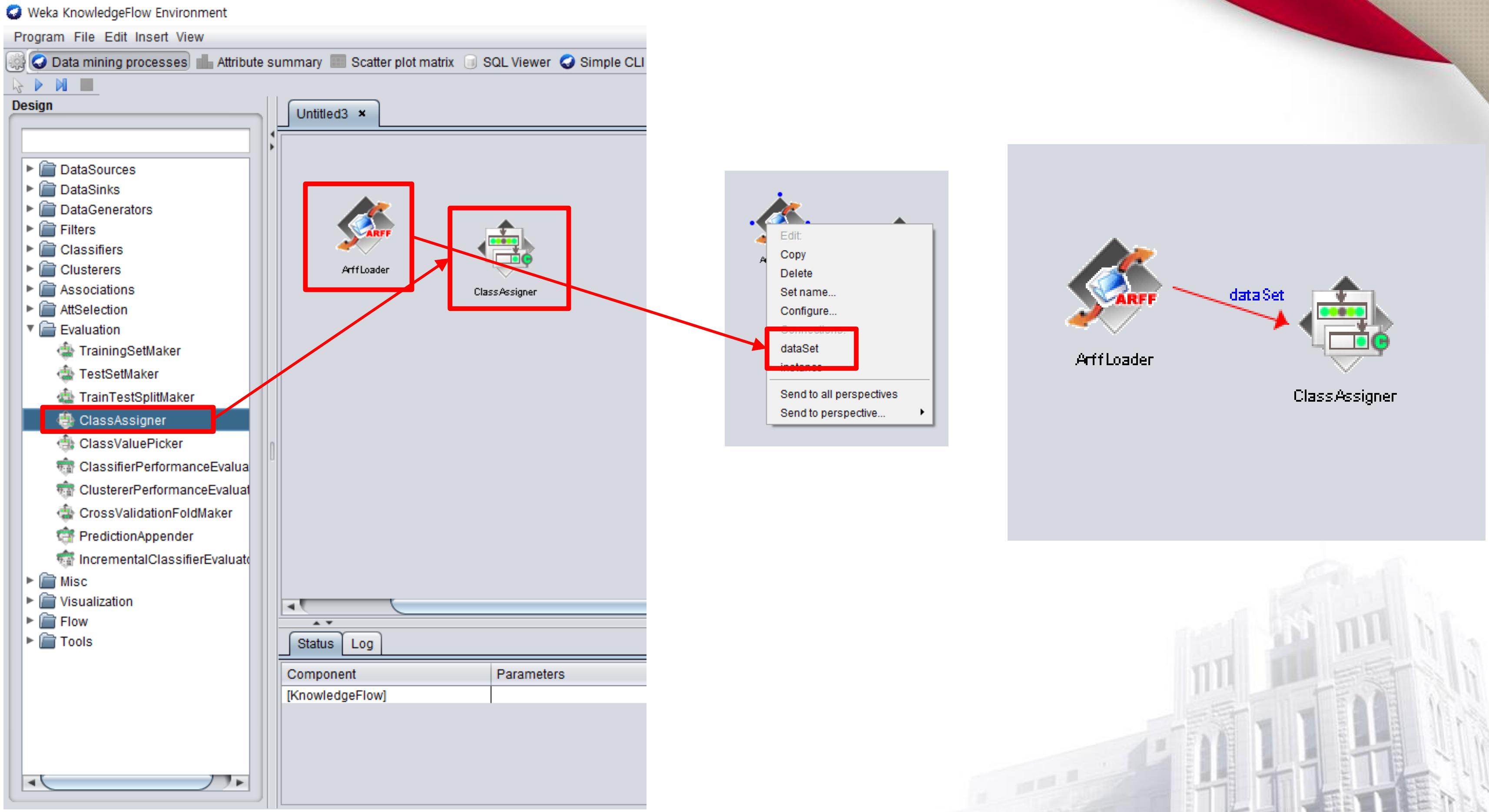
credit-g.arff, diabetes.arff, glass.arff, hypothyroid.arff, ionosphere.arff, iris.2D.arff, iris.arff, labor.arff, result.csv, ReutersCorn-test.arff, ReutersCorn-train.arff, ReutersGrain-test.arff, ReutersGrain-train.arff, segment-challenge.arff, segment-test.arff

파일 이름(N): iris.arff

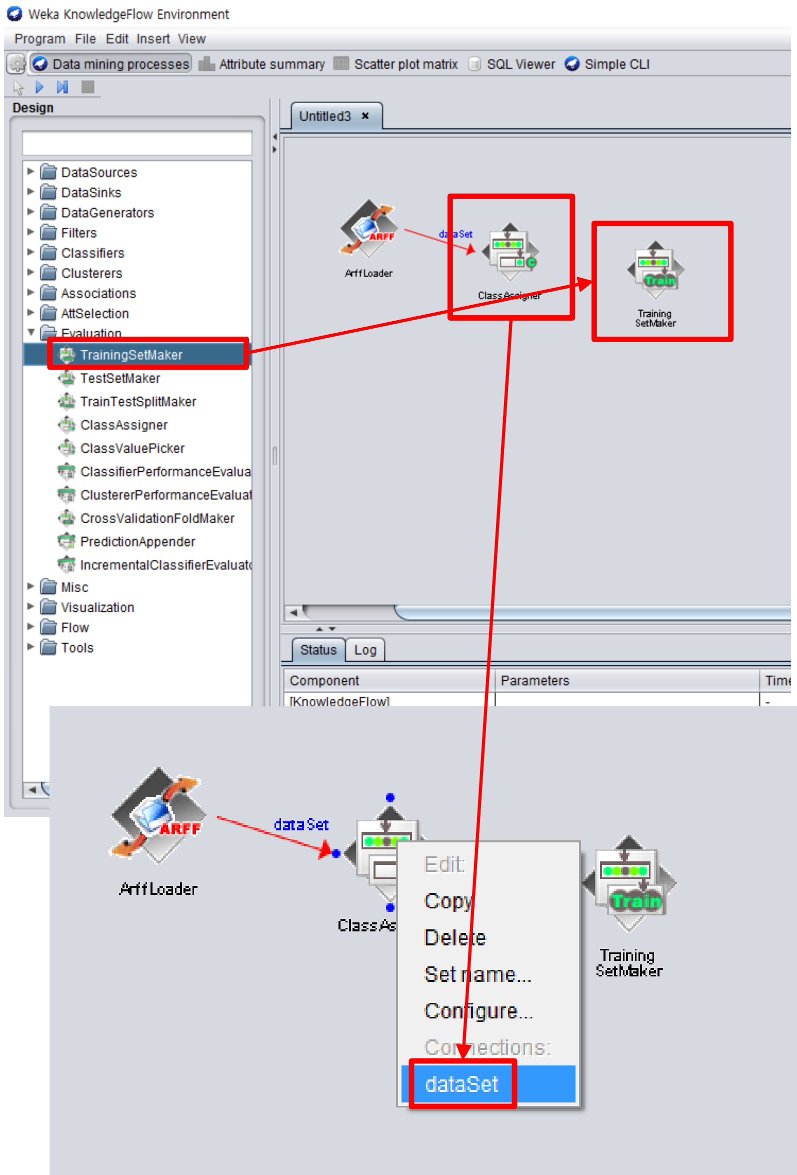
파일 유형(T): 모든 파일

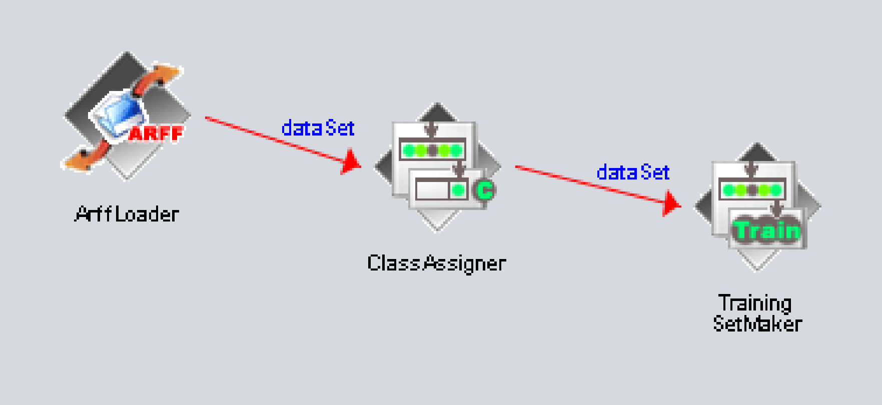
Double click

Knowledge flow

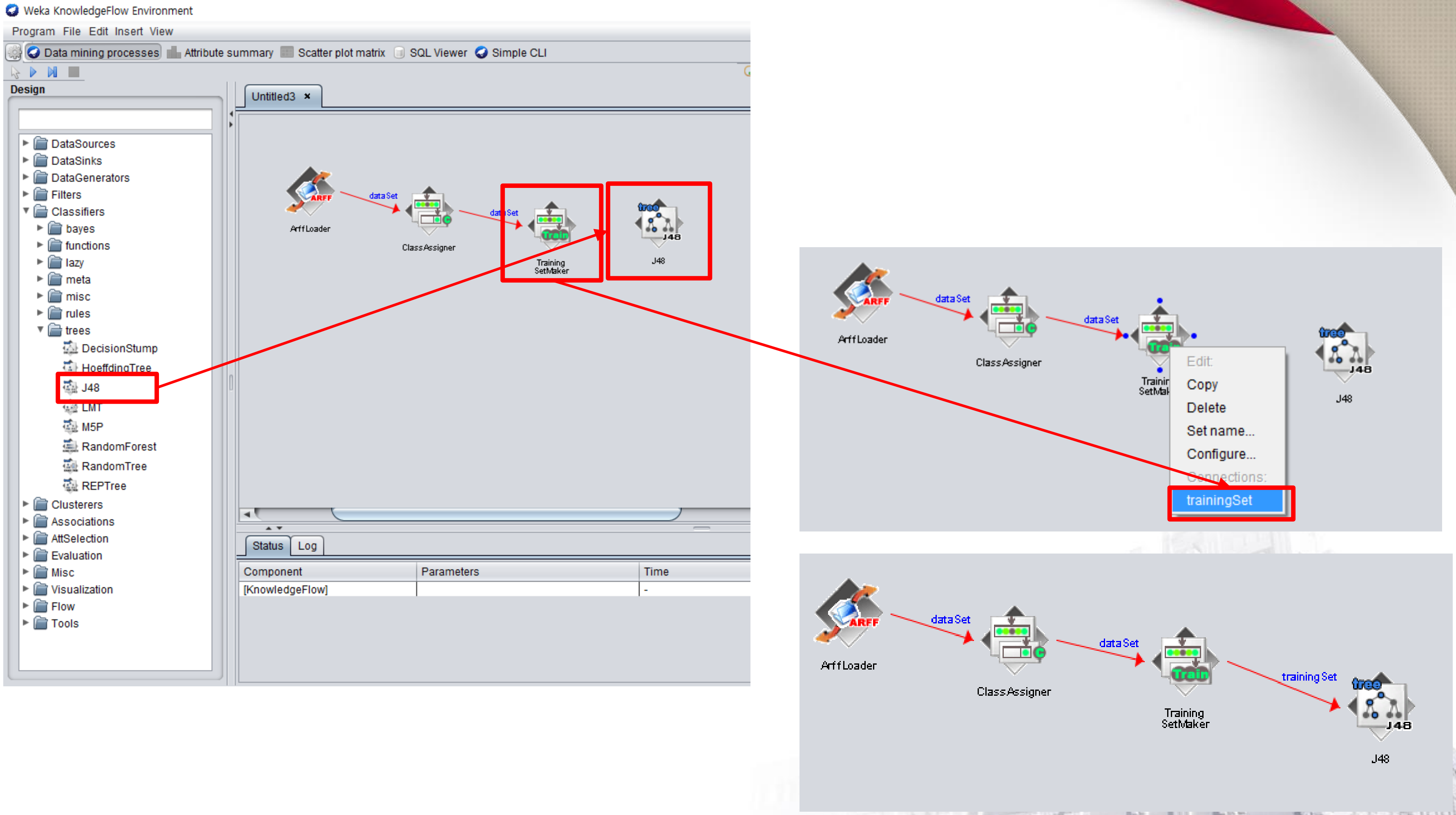


Knowledge flow



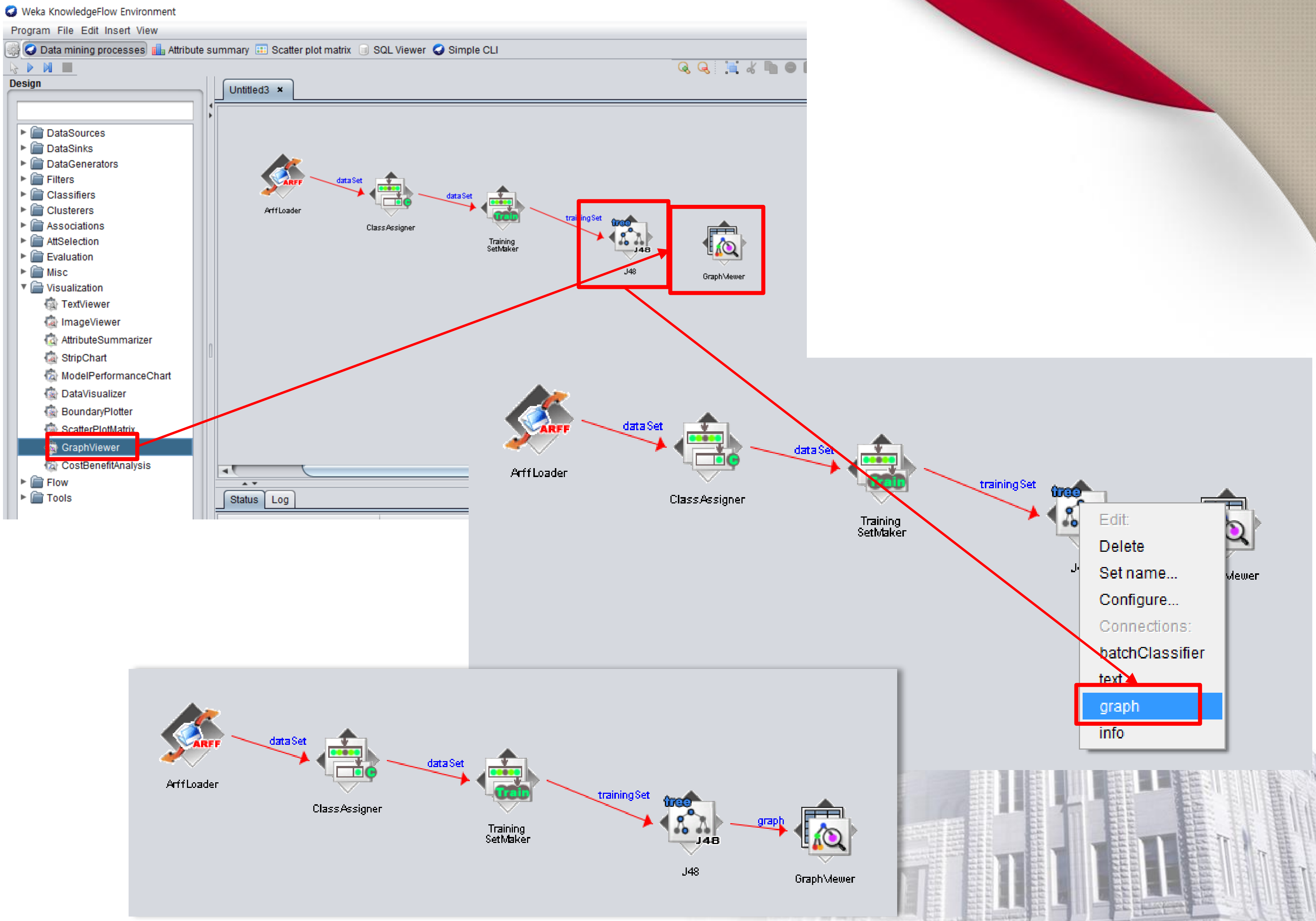


Knowledge flow

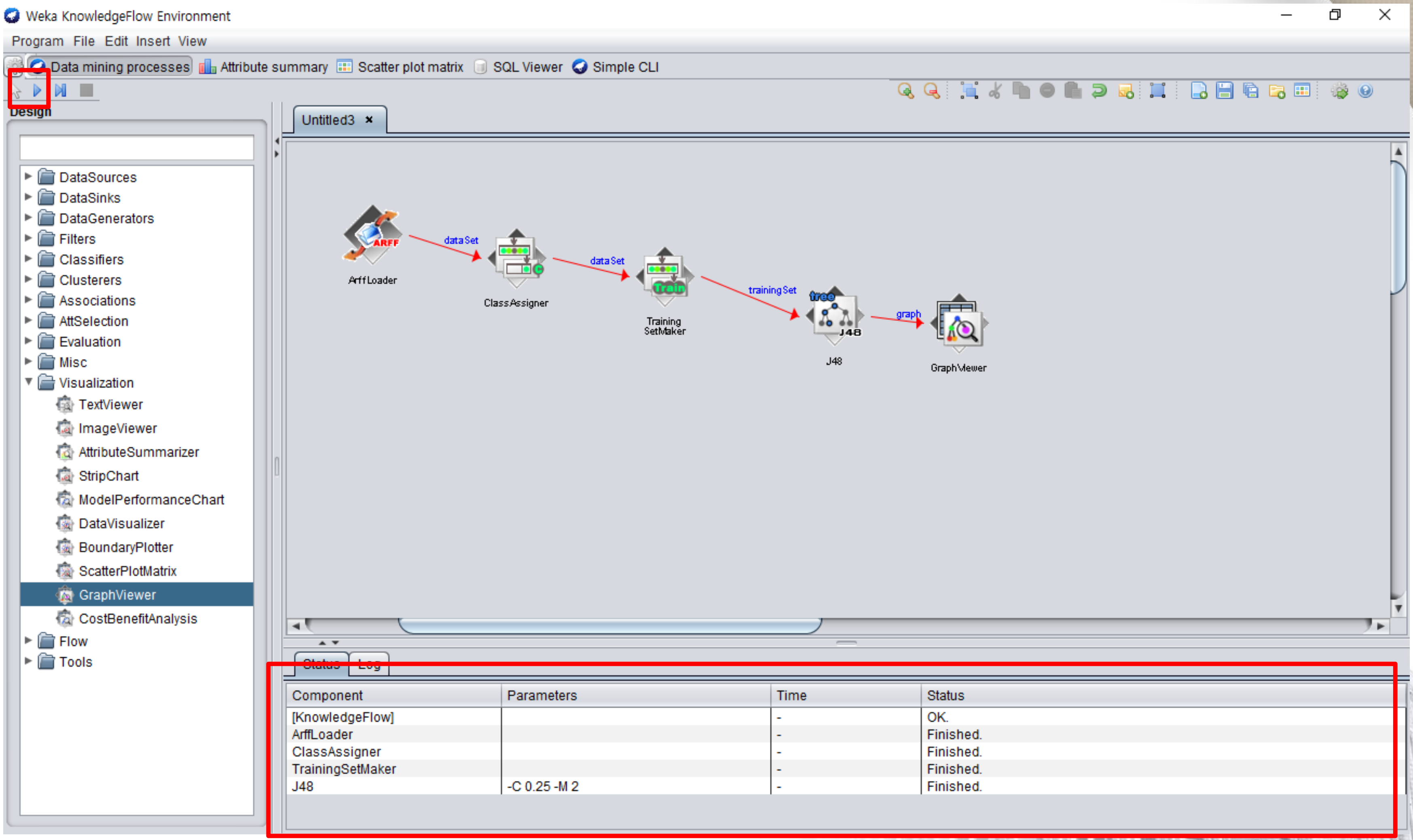


Component	Parameters	Time
[KnowledgeFlow]		-

Knowledge flow

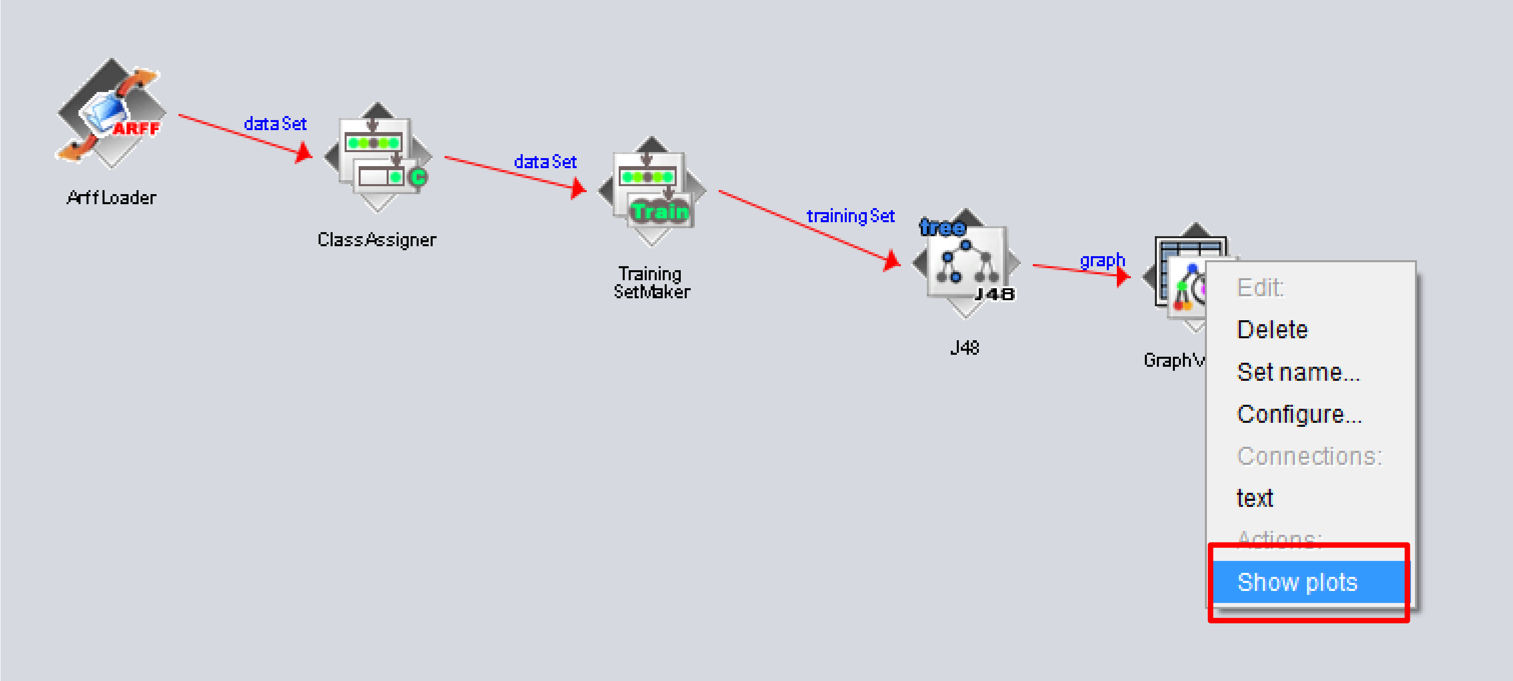


Knowledge flow

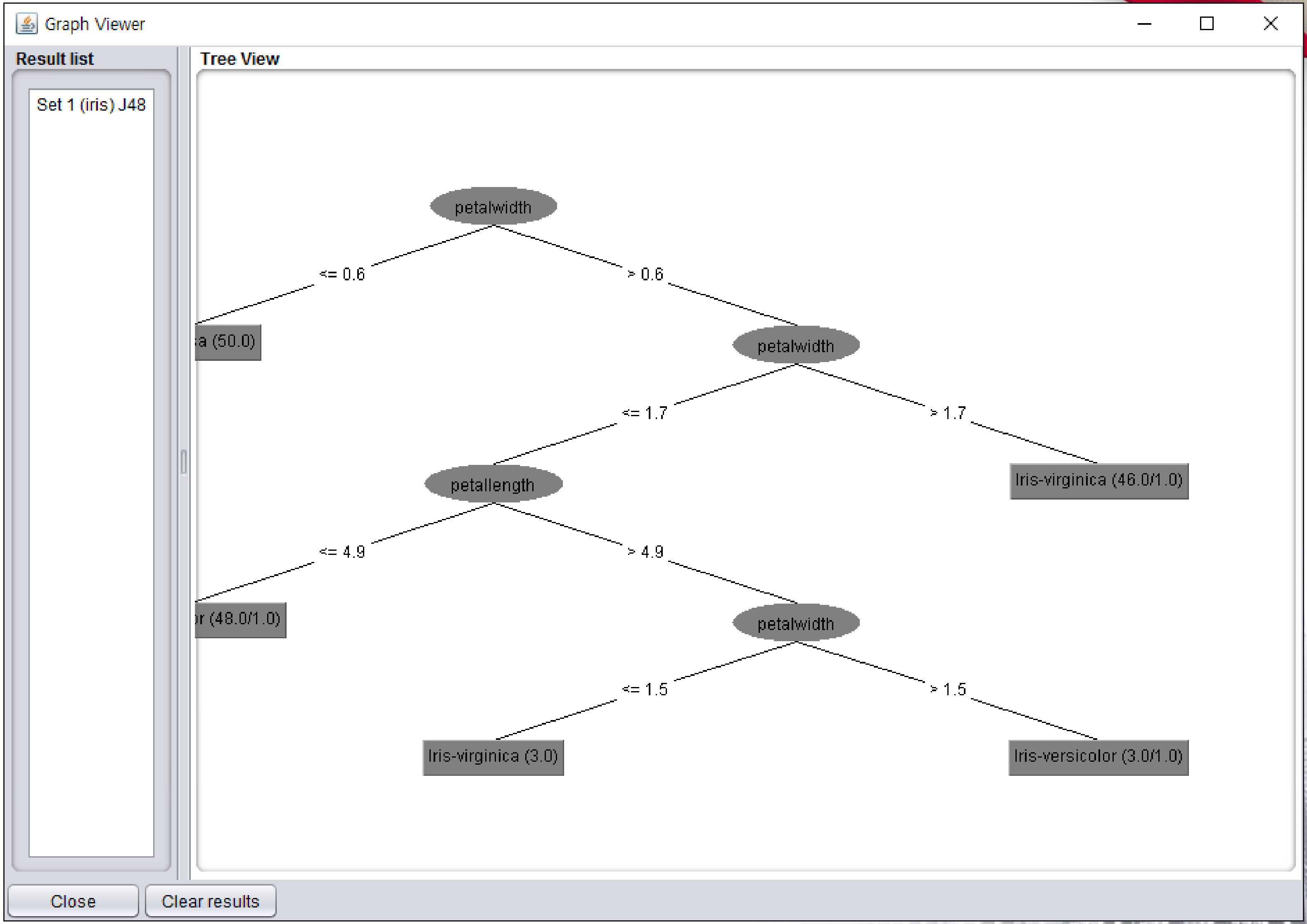


Component	Parameters	Time	Status
[KnowledgeFlow]		-	OK.
ArffLoader		-	Finished.
ClassAssigner		-	Finished.
TrainingSetMaker		-	Finished.
J48	-C 0.25 -M 2	-	Finished.

Knowledge flow



Knowledge flow



Weka Reference

http://statweb.stanford.edu/~lpekelis/13_datafest_cart/WekaManual-3-7-8.pdf

http://software.ucv.ro/~eganea/AIR/KnowledgeFlowTutorial-3-5-8.pdf

https://www.youtube.com/watch?v=bPrTeUAS6_I&list=PLJbE6j2EG1pZnBhOg3_Rb63WLCprtyJag&index=26