Data Mining

Munawar, PhD

7. Data Mining Tools

Agenda

01 Overview

02 Tools Comparison

03 Rapid Miner

04 Weka

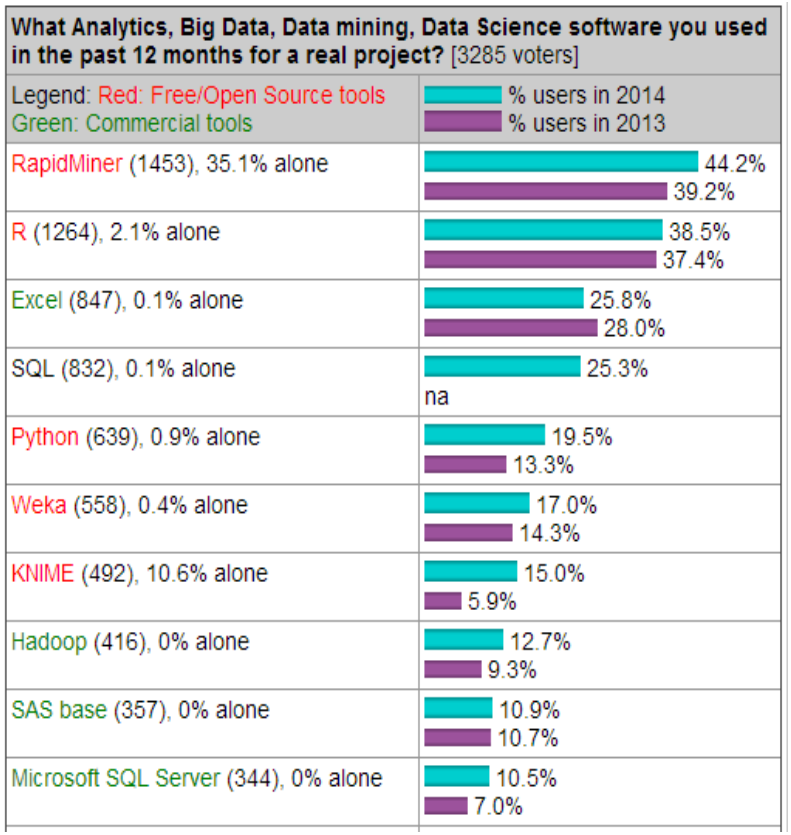
Overview

Overview

- Knowledge discovery in databases is a rapidly growing field, whose development is driven by strong research interests as well as urgent practical, social, and economical needs.
- The last few years, knowledge discovery tools have been used mainly in research environments, sophisticated software products are now rapidly emerging.
- In this course, we provide an overview of data mining tools and discuss open source tools to study basic data mining methods, including preprocessing data and association rule method:
  - ✓ RapidMiner
  - ✓ WEKA

Overview

The top 10 tools by share of users were (Kdnuggets-2014)



What Analytics, Big Data, Data mining, Data Science software you used in the past 12 months for a real project? [3285 voters]	
Legend: Red: Free/Open Source tools Green: Commercial tools	% users in 2014 % users in 2013
RapidMiner (1453), 35.1% alone	44.2% 39.2%
R (1264), 2.1% alone	38.5% 37.4%
Excel (847), 0.1% alone	25.8% 28.0%
SQL (832), 0.1% alone	25.3% na
Python (639), 0.9% alone	19.5% 13.3%
Weka (558), 0.4% alone	17.0% 14.3%
KNIME (492), 10.6% alone	15.0% 5.9%
Hadoop (416), 0% alone	12.7% 9.3%
SAS base (357), 0% alone	10.9% 10.7%
Microsoft SQL Server (344), 0% alone	10.5% 7.0%

- Most popular open source tools:
  - RapidMiner, 44.2% share (39.2% in 2013)
  - R, 38.5% (37.4% in 2013)
  - Python, 19.5% (13.3% in 2013)
  - Weka, 17.0% (14.3% in 2013)
  - KNIME, 15.0% (5.9% in 2013)
- Most popular commercial tools:
  - SAS Enterprise Miner
  - MATLAB
  - IBM SPSS Modeler

Popular Open Source DM

- **RapidMiner:** many DM algorithms (also can import Weka's methods), extendable, steady learning curve, recent problems with licensing
- **Weka:** many DM algorithms, user-friendly, extendable, not the best choice for data visualization or advanced DM tasks at this time
- **R:** strong in statistics and DM algorithms, extendable, fast implementations, complexity of extensions, not user-friendly – some improvement with Rattle GUI
- **KNIME:** user-friendly, extendable (e.g. Weka, R), covers most of the advanced DM tasks as add-ons, no significant downsides
- **Orange:** user-friendly, visually appealing GUI, moderate DM algorithms coverage, doesn't cover advanced DM tasks at this time
- **scikit-learn:** great documentation, fast implementations, moderate DM algorithms coverage, not user-friendy

Programming/ Statistic Language

Language used	% voters in 2014 (719 total) % voters in 2013 (713 total) % voters in 2012 (579 total)
R (352 voters in 2014)	49.0% 60.9% 52.5%
SAS (262)	36.4% 20.8% 19.7%
Python (252)	35.0% 38.8% 36.1%
SQL (220)	30.6% 36.6% 32.1%
Java (89)	12.4% 16.5% 21.2%
Unix shell/awk/sed (63)	8.8% 11.1% 14.7%
Pig Latin/ Hive/ other Hadoop-based languages (61)	8.5% 8.0% 6.7%
SPSS (58)	8.1% not asked not asked
MATLAB (45)	6.3% 12.5% 13.1%

- Top ten of programming/ statistics languages used for an analytics/data mining/data science work in 2014:
  - R
  - SAS
  - Python
  - Java
  - Unix
  - Pig Latin/Hive/Hadoop
  - SPSS
  - Matlab

Source: http://www.kdnuggets.com

Tools Comparison

Popular Open Source DM

Characteristic	RapidMiner	R	Weka	Orange	KNIME	scikit-learn
Developer:	RapidMiner, Germany	worldwide development	Univ. of Waikato, New Zealand	Univ. of Ljubljana, Slovenia	KNIME.com AG,Switzerland	multiple; support: INRIA, Google
Programming language:	Java	C, Fortran, R	Java	C++, Python, Qt framew.	Java	Python+NumPy+ SciPy+matplotlib
License:	open s. (v.5 or lower); closed s., free Starter ed. (v.6)	free software, GNU GPL 2+	open source, GNU GPL 3	open source, GNU GPL 3	open source, GNU GPL 3	FreeBSD
GUI/CL:	GUI	both; (GUI for DM = Rattle)	both	both	GUI	command line
Main purpose:	general data mining	sci. computation and statistics	general data mining	general data mining	general data mining	machine learning package add-on
Community support (est.):	large (~200 000 users)	very large (~ 2 M users)	large	moderate	moderate (~ 15 000 users)	moderate


Rapid Miner

Introduction

- **RapidMiner**, formerly known as YALE (Yet Another Learning Environment), was developed starting in 2001 at the Artificial Intelligence Unit of Technical University of Dortmund, Germany.
  - RapidMiner is an **open source** software platform that provides an integrated environment for machine learning, data mining, text mining, predictive analytics and business analytics.
  - Written in the **Java programming language**.
  - Follows a **modular operator concept** which allows the design of complex nedted operator chains for a huge number of learning problems.
  - Used for business and industrial applications as well as for research, education, training, rapid prototyping, and application development and supports all steps of the data mining process.
  - Available to download from https://rapidminer.com

Preparation

- Download the free version of RapidMiner software from
  http://www.rapidminer.com
- Documentation
  http://docs.rapidminer.com/studio
- RapidMiner Installation Guide:
  http://docs.rapidminer.com/studio/installation/index.html
- RapidMiner manual guide
  https://rapidminer.com/wp-content/uploads/2014/10/RapidMiner-v6-user-manual.pdf
- Tutorial Video
  http://videos.rapidminerresources.com/course/index.php?categoryid=4

Installation & First Recovery

- Download the appropriate installation package for your OS and install RapidMiner Studio according to the instructions on the website. https://rapidminer.com

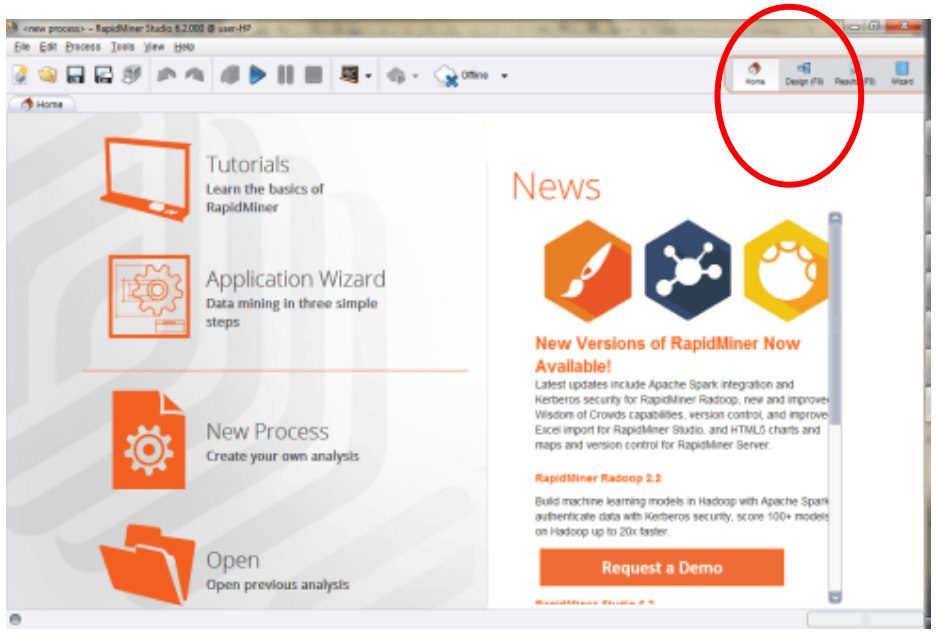
- RapidMiner is written in the Java programming language, so that an up-to-date **Java Runtime** is needed.

- Create a local repository on your computer to begin with the first use of RapidMiner Studio.

Home Perspective

- RapidMiner Studio is the GUI-based software where data mining and predictive analytics workflows can be built and deployed.

- You can select which perspective:
  - **Home** perspective
  - **Design** perspective
  - **Results** perspective
  - **Wizard** perspective

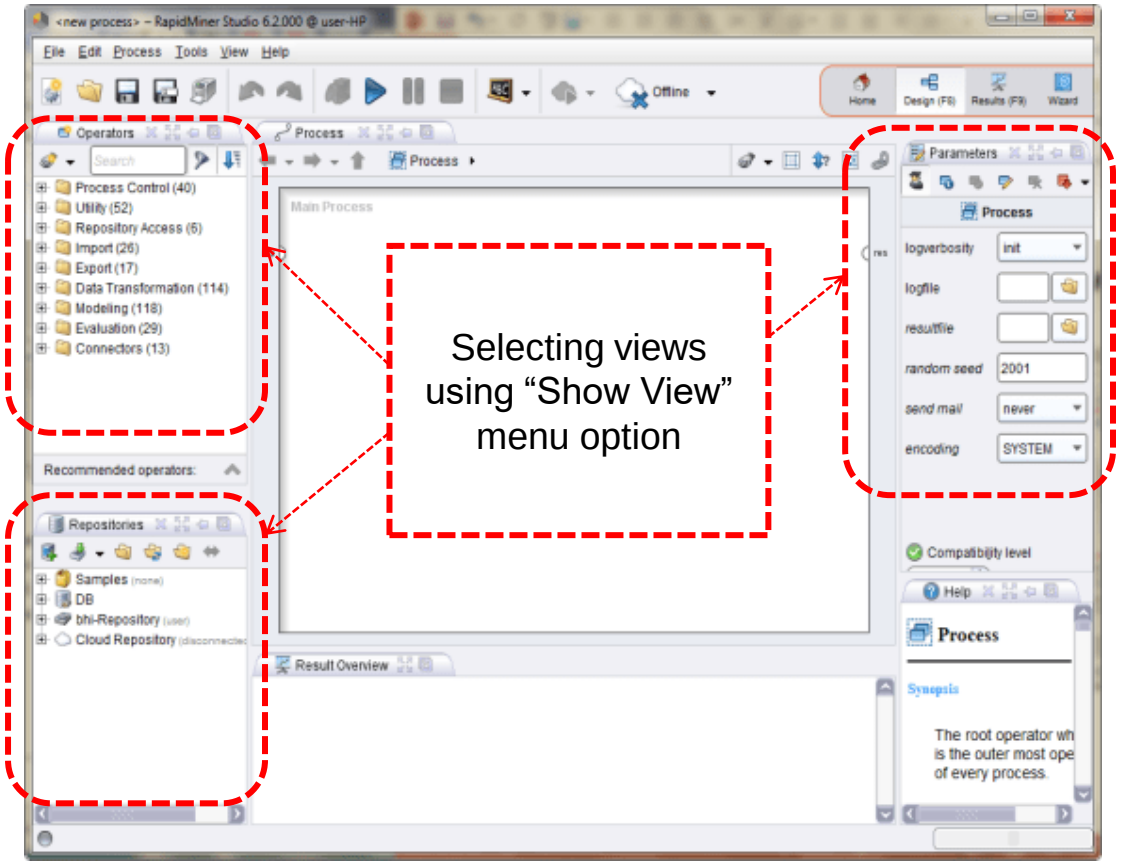


Launch view of RapidMiner 6.2 in **Home** Perspective

Design Perspective

- **Design Perspective**: This is the central RapidMiner Studio perspective where all analysis processes are created, edited and managed.

- It does not only an almost comprehensive set of **operators**, but also **structures** that express the control ow of the process.



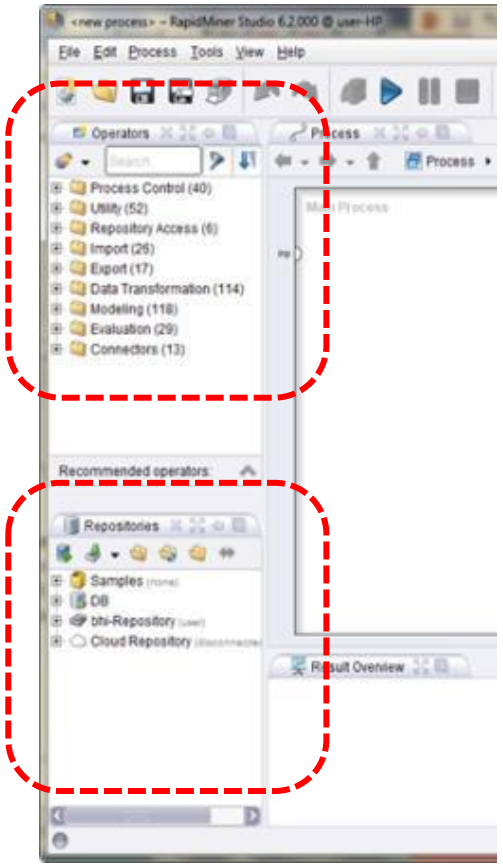
Design Perspective of RapidMiner.

Operators & Repositories View

- There are two very meaningful views in this Design area:
- **Operators View**

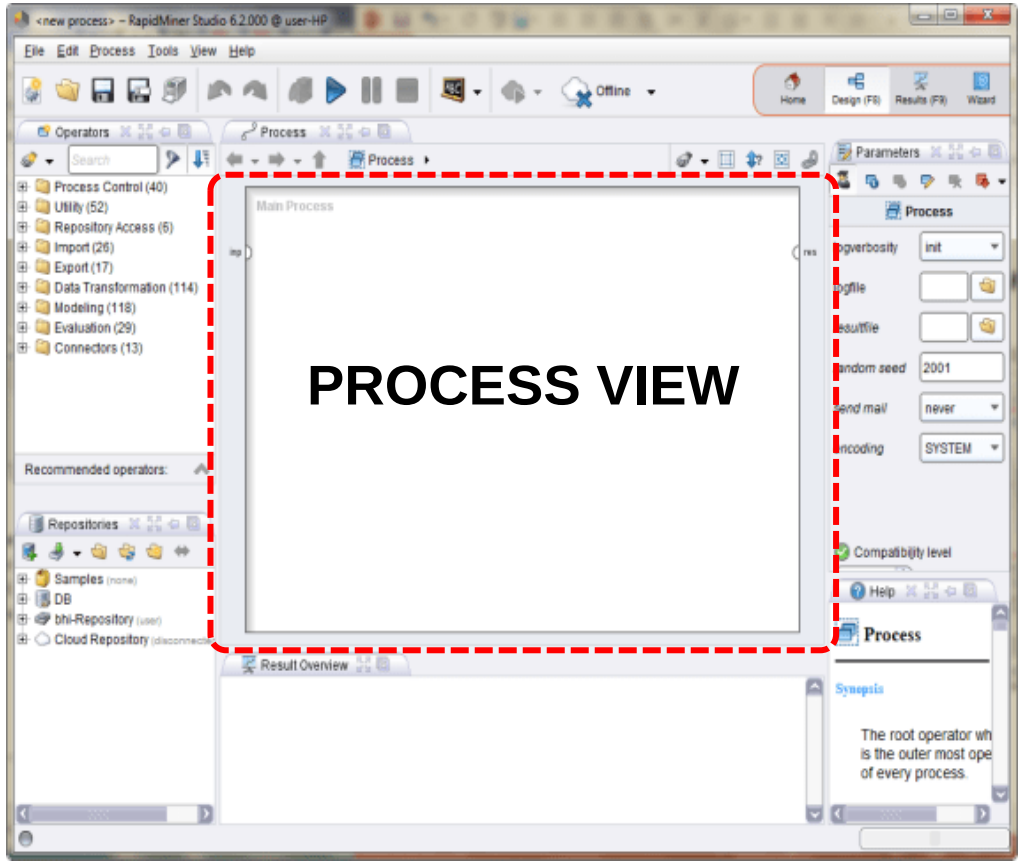
  All work steps (operators) available in RapidMiner Studio are presented in groups here and can therefore be included in the current process.
- **Repositories View**

  The repository is a central component of RapidMiner Studio which was introduced in Version 5. It is used for the management and structuring of your analysis processes into projects and at the same time as both a source of data as well as of the associated meta data.



Process View

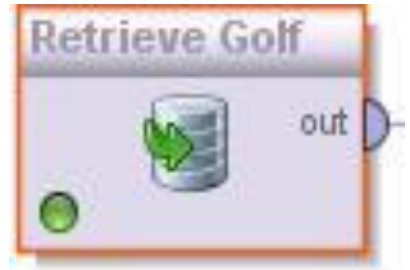
- In RapidMiner, process components are called **operators**.
- An operator is dened by several things:
  - The description of the expected inputs,
  - The description of the supplied outputs,
  - The action performed by the operator on the inputs, which ultimately leads to the supply of the outputs,
  - A number of parameters which can control the action performed.



Operators

- An **operator** is an atomic piece of functionality (which in fact is a chunk of encapsulated code) performing a certain task.

- The data mining tasks:
  - importing a data set into the RapidMiner repository,
  - cleaning it by getting rid of spurious examples,
  - reducing the number of attributes by using feature selection techniques,
  - building predictive models, or scoring new data sets using models built earlier.



An operator can be connected via its input ports (left) and output ports (right). Below: status indicator of operators

Group of Operators

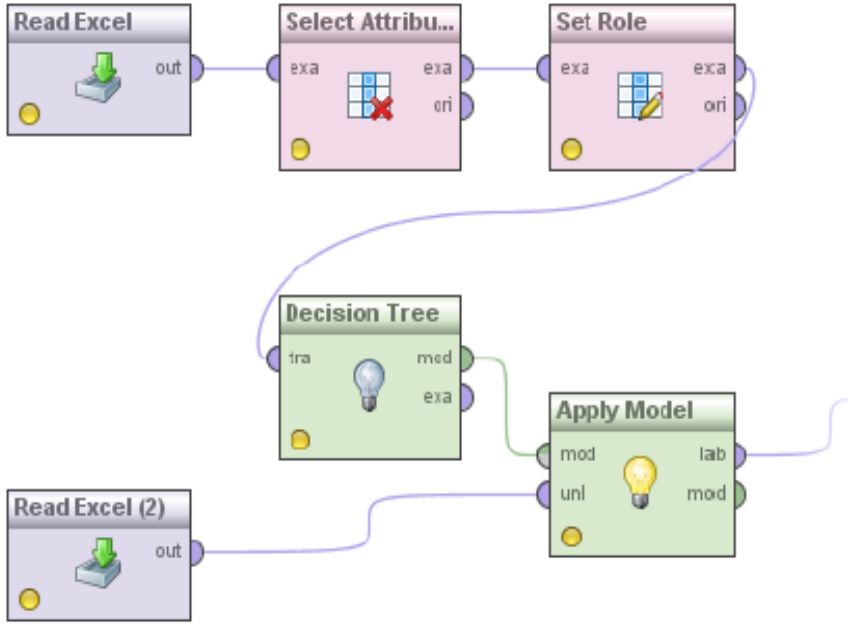
- Groups of operators in the tree structure.
- **Process Control**: Operators such as loops or conditional branches which can control the process ow.
- **Utility**: Auxiliary operators which, alongside the operator"Subprocess" for grouping subprocesses, also contain the important macro-operators as well as the operators for logging.
- **Repository Access**: Contains operators for read and write access in repositories.
- **Import**: Contains a large number of operators in order to read data and objects from external formats such as les, databases etc.
- **Export**: Contains a large number of operators for writing data and objects into external formats such as les, databases etc.

Process

- All data mining and predictive analytics problem solving require a series of calculations and logical operations, therefore a single operator by itself cannot perform data mining.
- All of these steps can be accomplished by connecting a number of different operators, each uniquely customized for a specific task as we saw earlier.
- There is typically a certain flow to these problems:
  - import data,
  - clean and prepare data,
  - train a model to learn the data,
  - validate the model and rank its performance, then finally apply the model to score new and unseen data.

Process (2)

- An analysis process consisting of several operators.
- You can insert new operators into the process in dierent ways, e.g.: Via *drag & drop* from the Operators View as described above.
- After you have inserted new operators, you can *interconnect* the operators inserted.

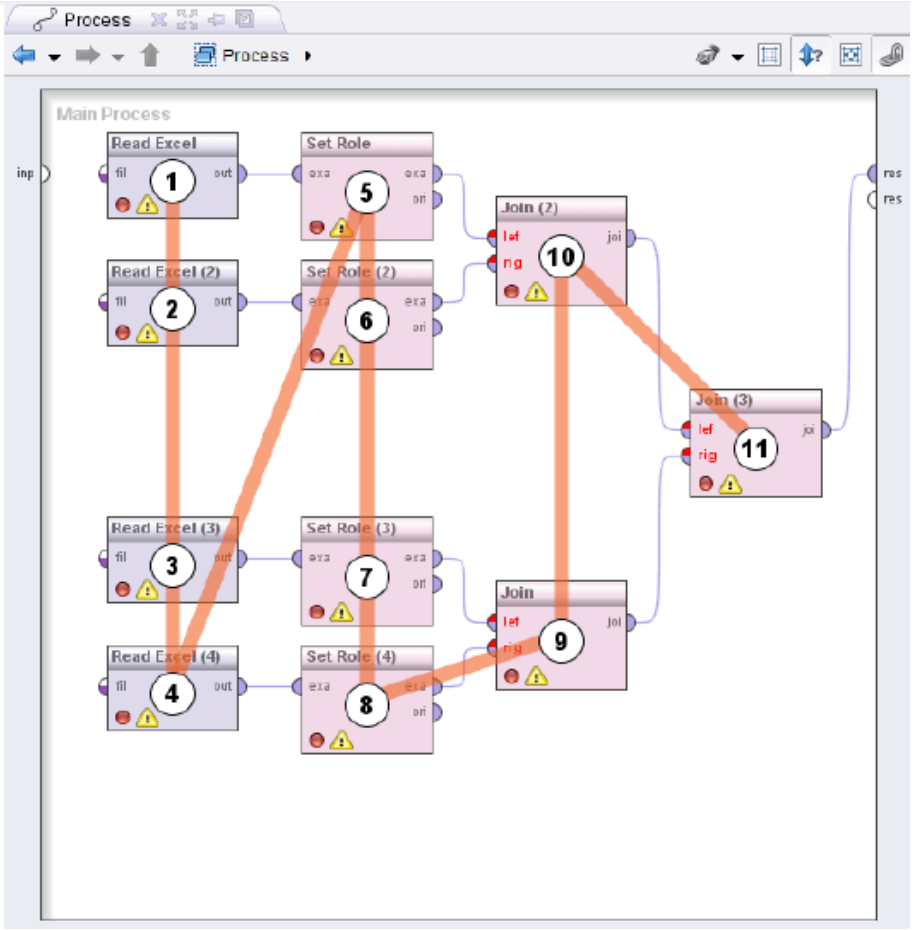


Further Options of the Process View

- The ve icons on the right-hand side of the Process View toolbar perform the following actions:
  - Auto-wire and Re-wire connections The plug symbol allows to auto-wire and re-wire the connections between operators.
  - Automatic arrangement: Rearranges all operators of the current process according to the connections and the current execution order.
  - Show and alter execution order This action allows you to see the execution order of the operators and to change it.
  - Automatic size: Changes the size of the white working area in such a manner that all operators currently positioned have just enough space.

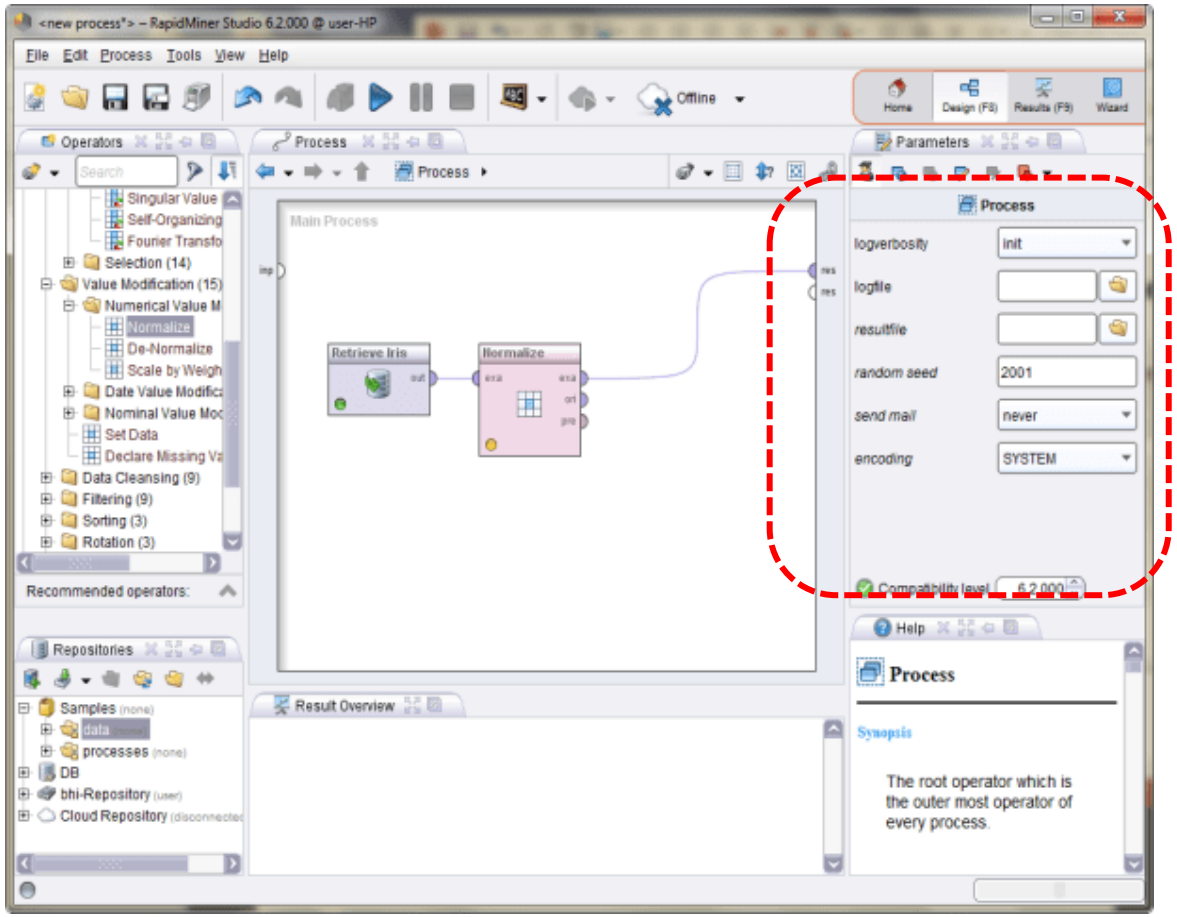
Further Options of the Process View

- In further options of the Process View, representation of the execution order is unfavourable however, since more data sets have to be handled at the same time.



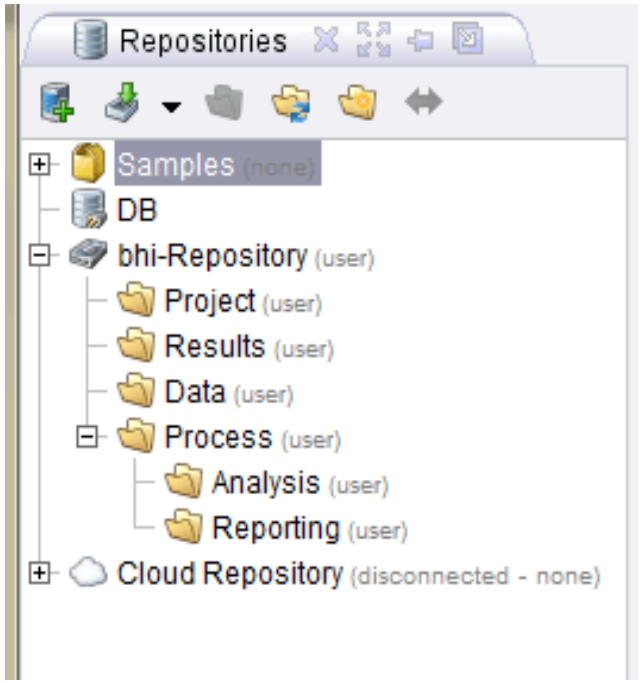
Parameter View

- Parameters of the currently selected operator are set in the parameter view.
- Numerous operators require one or several parameters to be indicated for a correct functionality.



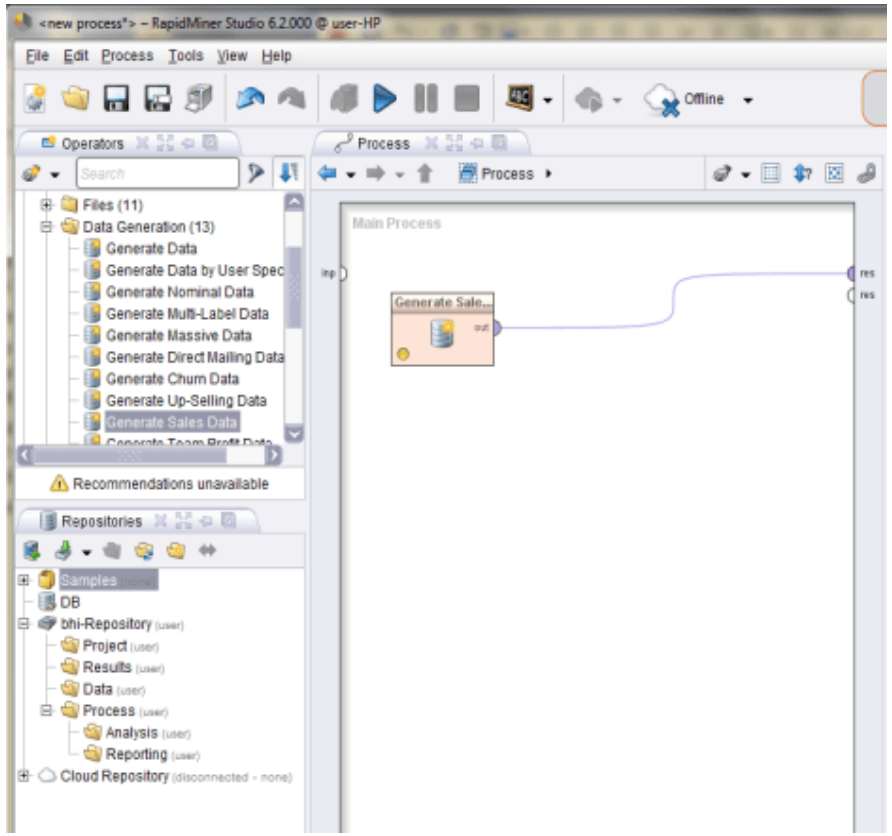
Creating a New Process

- You can start a new process by selecting the “New” button under the “File" menu from the Home Perspective.
- In principle, you are completely free in how you structure your repository.
- A repository structured into projects and each of those structured according to data, **processes** and **results**.



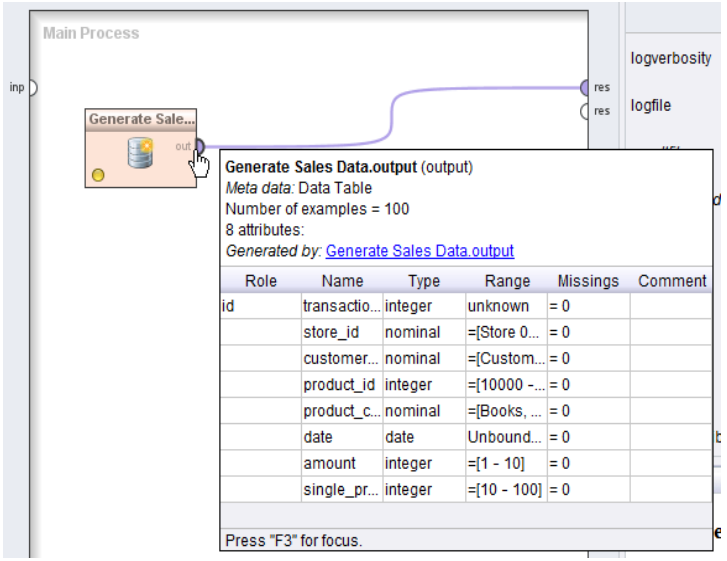
First Analysis Process

- After the creation of the process, RapidMiner Studio automatically switches to the **Design Perspective** and you can start with the process design.
- Now we will begin our new process starting with the **generating of data** which we can work on.
- Expand the group “**Utility**" in the **Operators View** and then the group Data Generation“, e.g. the operator “Generate Sales Data"



Transforming Meta Data

- The most fascinating aspects of RapidMiner Studio, namely the ability to compute the output of an operator or process beforehand and to even do this during the design time, so without having to load the actual data or even perform the process.
- This is made possible by the so-called **meta data transformation** of RapidMiner Studio.



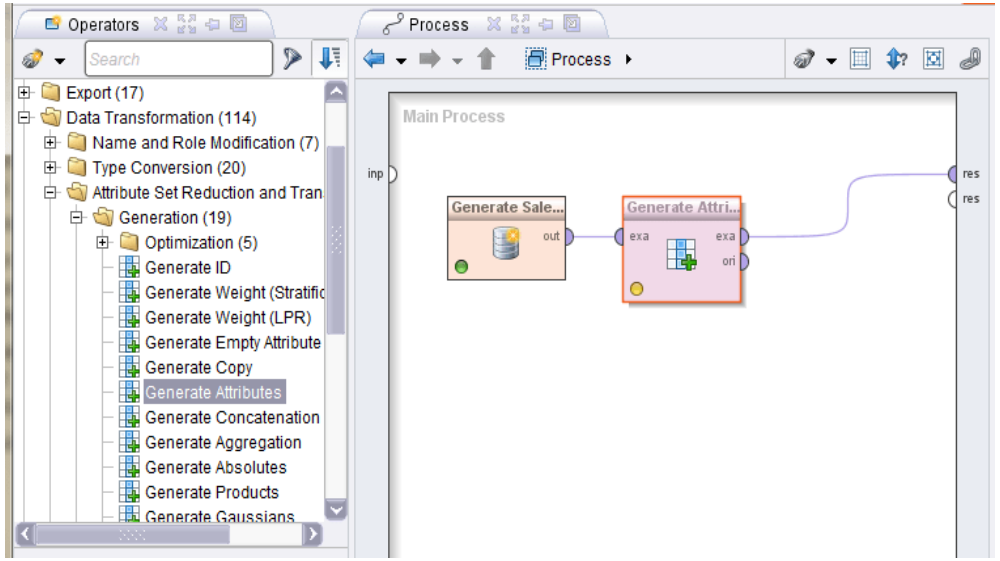
The meta data of the output port of the operator “*Generate Sales Data*".

Transforming Meta Data (2)

- The most important part of the meta data is the table which describes the meta data of the individual attributes. The individual columns are:
  - **Role**: The role of the attribute. If nothing is indicated then it is a regular attribute
  - **Name**: The name of the attribute
  - **Type**: The value type of the attribute
  - **Range**: The value range of the attribute, so the minimum and maximum in the case of numerical attributes and an excerpt of possible values in the case of nominal attributes
  - **Missings**: The number of examples where the value of this attribute is unknown
  - **Comment**: A comment depending on the attribute

View Meta Data

- When we want to generate a **new attribute** with the name "*total price*", we will use a further operator named "Generate Attributes", which is located in the group "Data Transformation" - Attribute Set Reduction and Transformation -"Generation".

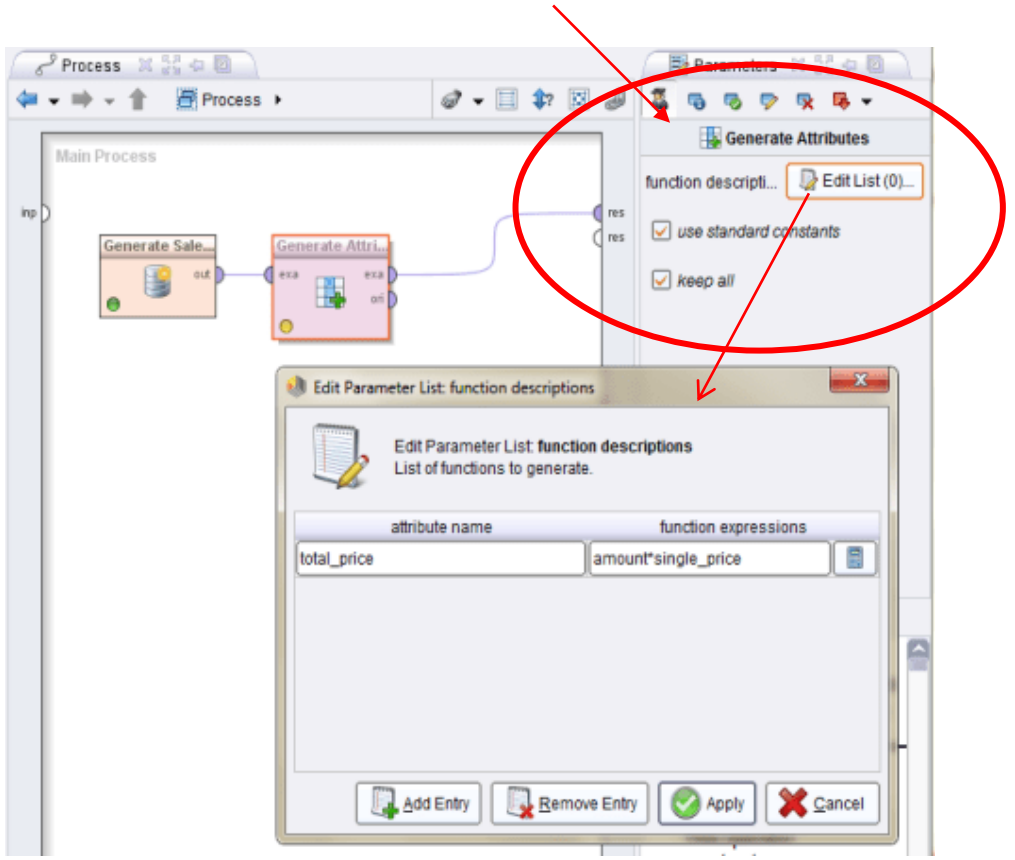


Generate New Attribute

- Computation of the new attribute “total price" as a product of “amount" and “single price“ can be done by select “Edit List” in “Generate Attribute”

- A new attribute “total_price” will be created by selecting “Apply” button.

The parameters of the operator “Generate Attributes".



Generate New Attribute

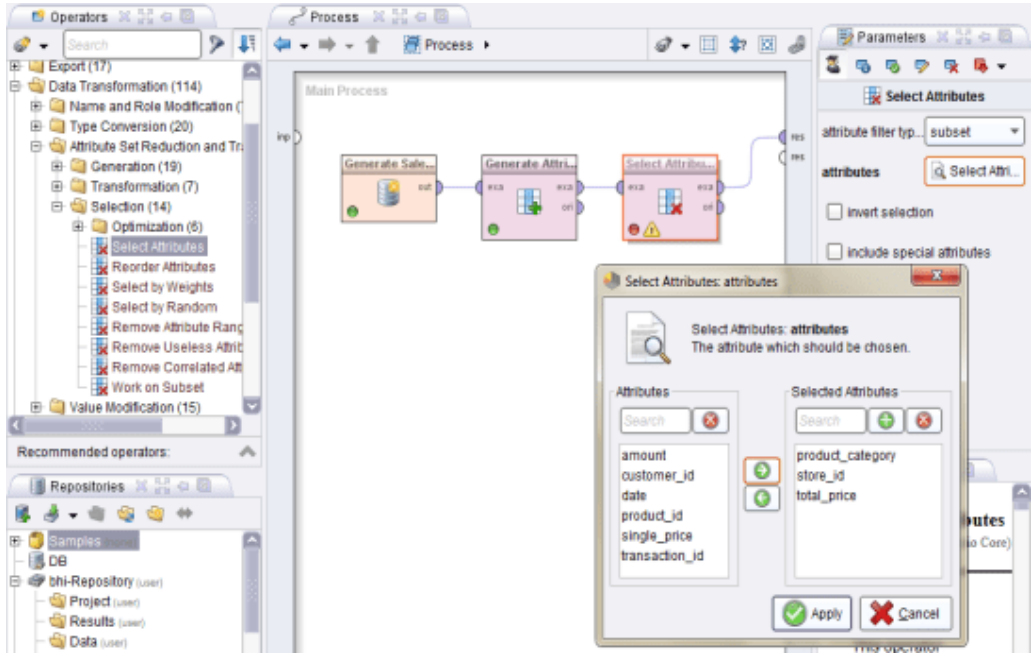
- A new attribute “total_price” has been created

ExampleSet (100 examples, 1 special attribute, 8 regular attributes) Filter (100 / 100 examples): all

Row No.	transaction...	store_id	customer_id	product_id	product_cat...	date	amount	single_price	total_price
1	1	Store 01	Customer 1	53642	Toys	Apr 1, 2007	3	90.246	270.739
2	2	Store 15	Customer 1	90945	Movies	Feb 15, 200	2	60.586	121.173
3	3	Store 12	Customer 1	18548	Movies	Sep 27, 200	5	96.613	483.063
4	4	Store 05	Customer 1	85359	Books	May 7, 2005	5	16.963	84.813
5	5	Store 01	Customer 4	80069	Clothing	Jan 6, 2008	5	65.215	326.077
6	6	Store 11	Customer 7	55848	Sports	Jun 3, 2006	3	56.475	169.424
7	7	Store 10	Customer 7	11762	Health	Sep 19, 200	3	26.873	80.619
8	8	Store 10	Customer 1	75667	Health	Nov 29, 200	7	67.075	469.522
9	9	Store 11	Customer 1	97291	Health	Mar 21, 200	3	47.246	141.737
10	10	Store 14	Customer 5	39580	Toys	Sep 5, 2005	8	42.669	341.352
11	11	Store 14	Customer 1	58636	Health	Jun 5, 2008	5	81.849	409.243
12	12	Store 12	Customer 1	64853	Sports	Dec 8, 2007	8	29.309	234.470

Selection of Attribute

- RapidMiner allows generation of data, generation of a new attribute, and also **selection** of a subset of attributes.
- We can try : Open the group “Data Transformation" – “Attribute Set Reduction and Transformation" “Selection" and drag the operator named “Select Attributes" into the process



Individual attributes or subsets can be selected or even deleted with the operator “Select Attributes".

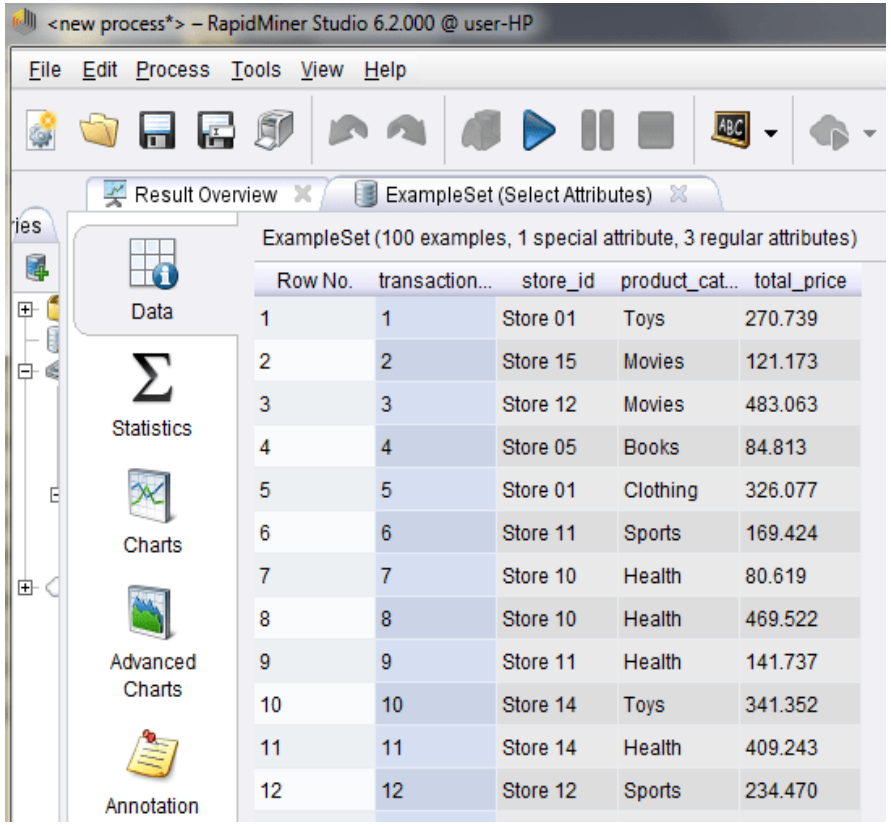
Executing Process & Results

Executing Processes:

You have the following options for starting the process:

1. Press the large play button in the toolbar of RapidMiner,
2. Select the menu entry “Process" - “Run",
3. Press F11.

Result Overview

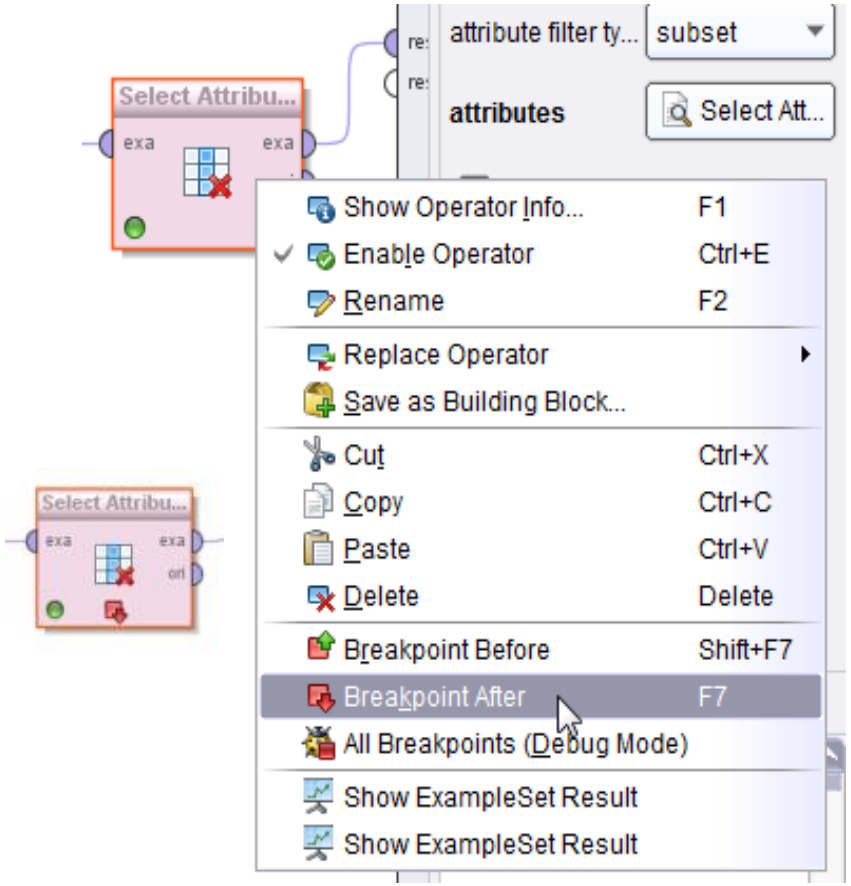


Row No.	transaction...	store_id	product_cat...	total_price
1	1	Store 01	Toys	270.739
2	2	Store 15	Movies	121.173
3	3	Store 12	Movies	483.063
4	4	Store 05	Books	84.813
5	5	Store 01	Clothing	326.077
6	6	Store 11	Sports	169.424
7	7	Store 10	Health	80.619
8	8	Store 10	Health	469.522
9	9	Store 11	Health	141.737
10	10	Store 14	Toys	341.352
11	11	Store 14	Health	409.243
12	12	Store 12	Sports	234.470

Breakpoints

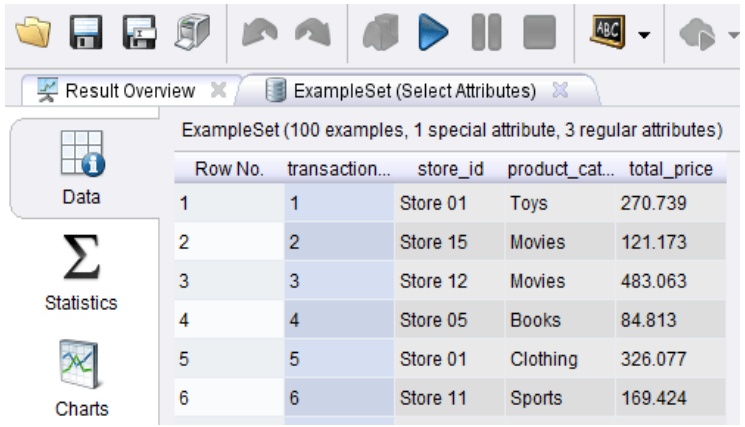
- RapidMiner allows us to restrict inter-operator process by providing process termination through “Breakpoint Before” and “Breakpoint After”

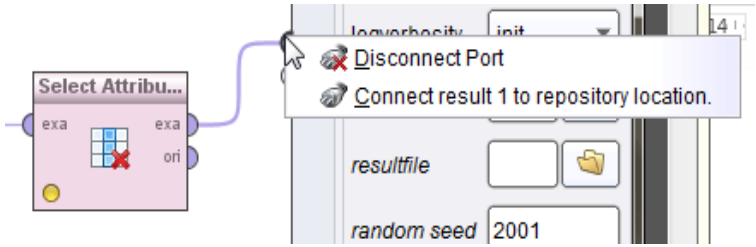
- If a breakpoint was inserted after an operator for example, then the execution of the process will be interrupted here and the results of all connected output ports will be indicated in the Results Perspective.



Visualization Data & Result

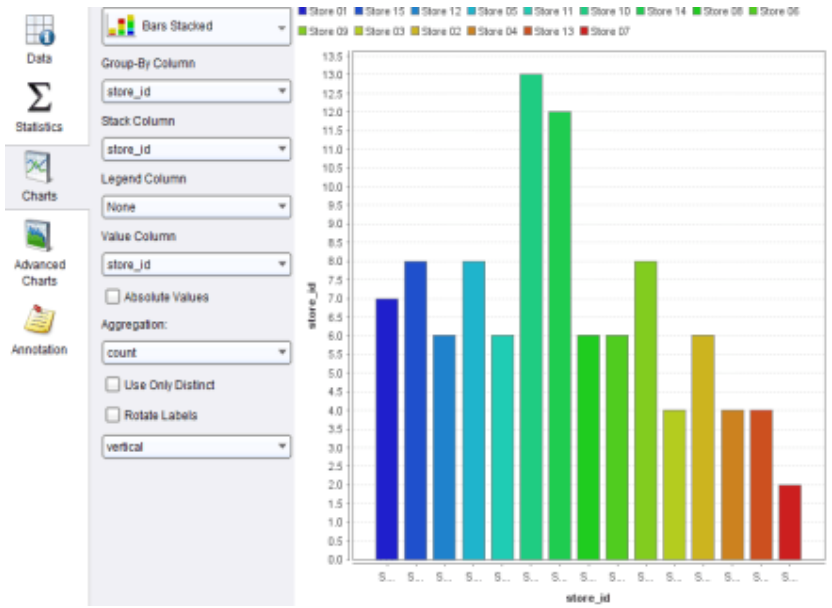
- Results of a process will be display on Result Perspective of RapidMiner. However, there are some different mode:
  - Each open result is displayed as an **additional tab** in the large area on the left-hand side as Automatic Opening**.**
  - The second option for displaying results is loading results from one of your **repositories**.
  - A third possibility for looking at results and even intermediate results is displaying results which are still at **ports**.

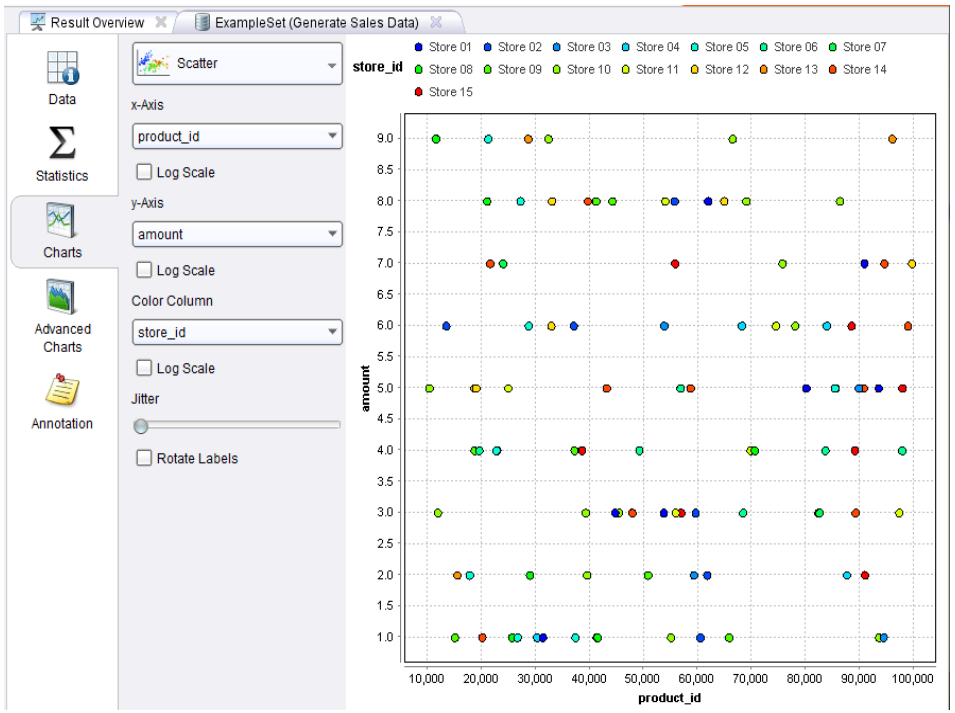




Chart

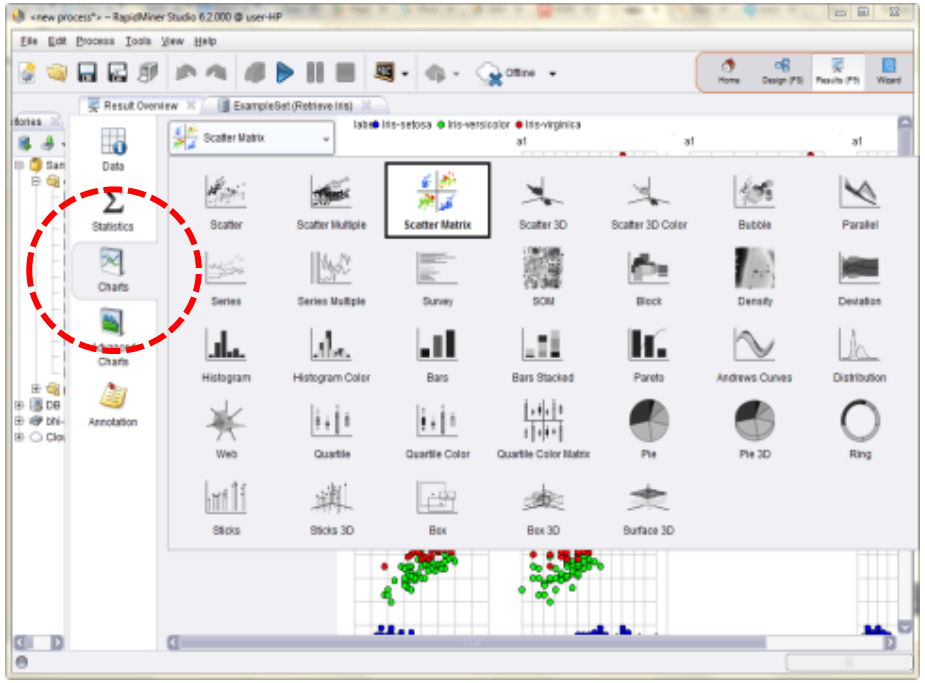
- One of the strongest features of RapidMiner Studio are the numerous visualisation methods for data, other tables, models and results oered in the “Charts View“ and “Advanced Charts View".

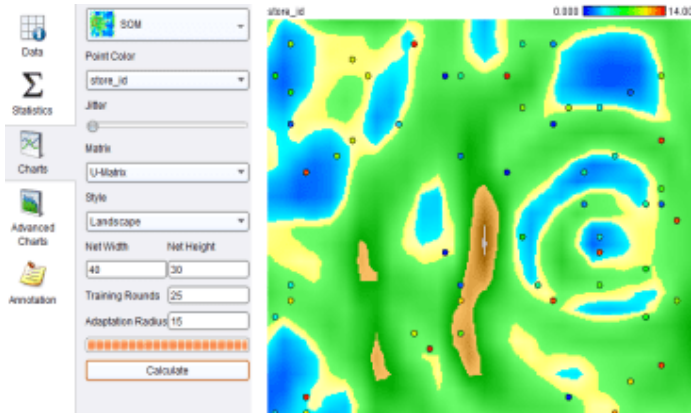




More Complex Visualization

- **Univariate Plots:** RapidMiner provides a lot of options to visualize the data by click Scatter Matrix button in Charts view of ExampleSet
- Complex visualisations such as SOMs over a “Calculate" button for starting the computation. The progress is indicated by a bar.






Weka

Introduction

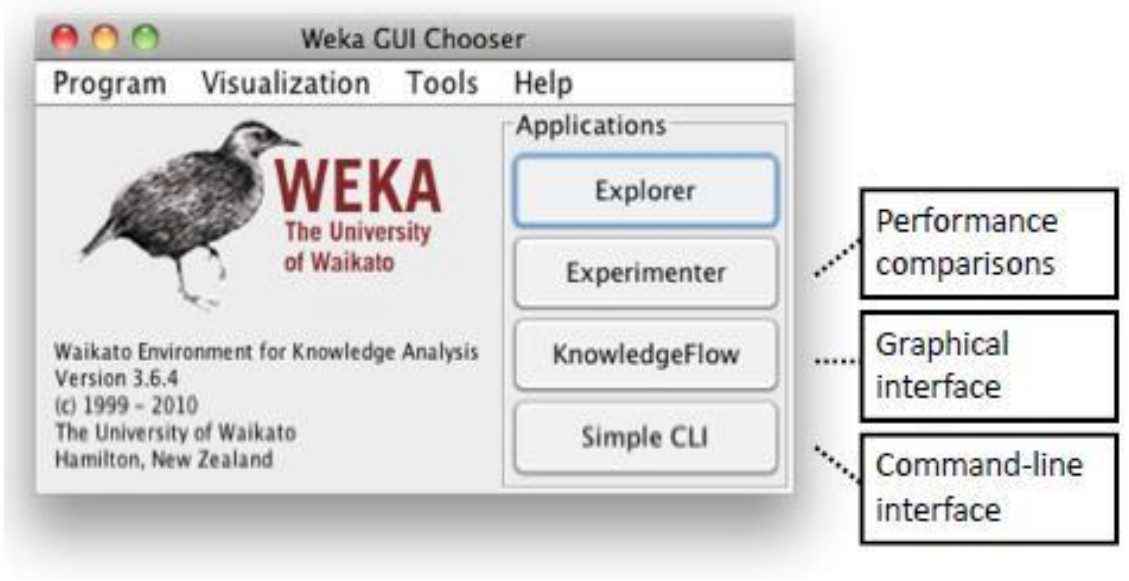
- Weka (Waikato Environment for Knowledge Analysis) is a collection of machine learning algorithms for data mining tasks written in Java.
- Weka is open source software issued under the GNU General Public License.
- Developed at the University of Waikato, New Zealand.
- The algorithms can either be applied 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.

Installation

- **Sources**:
- Software is available to download from: http://www.cs.waikato.ac.nz/ml/weka/
  - If you are interested in modifying/extending weka there is a developer version that includes the source code
- To complete exercise data repository should be also downloaded from http://mlearn.ics.uci.edu/MLRepository.html

Launching Weka

- The Weka GUI Chooser (class weka.gui.GUIChooser) provides a starting point for launching Weka's main GUI applications and supporting tools.

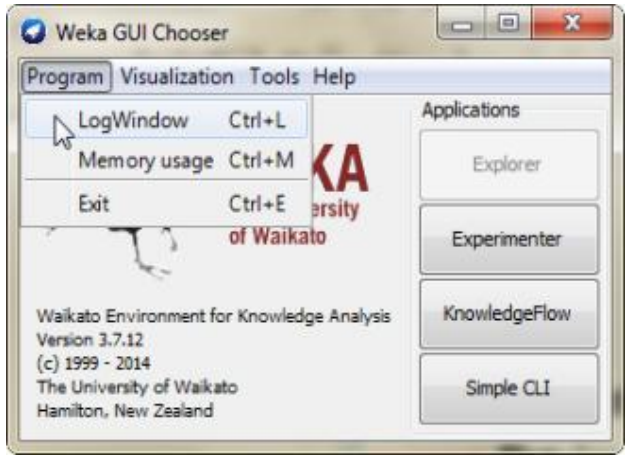


GUI

- The GUI Chooser consists of four **menus** and four **Applications** buttons—one for each of the four major Weka applications.
- Menus:
  - Program
  - Visualization
  - Tools
  - Help
- Applications:
  - Explorer
  - Experimenter
  - KnowledgeFlow
  - Simple CLI

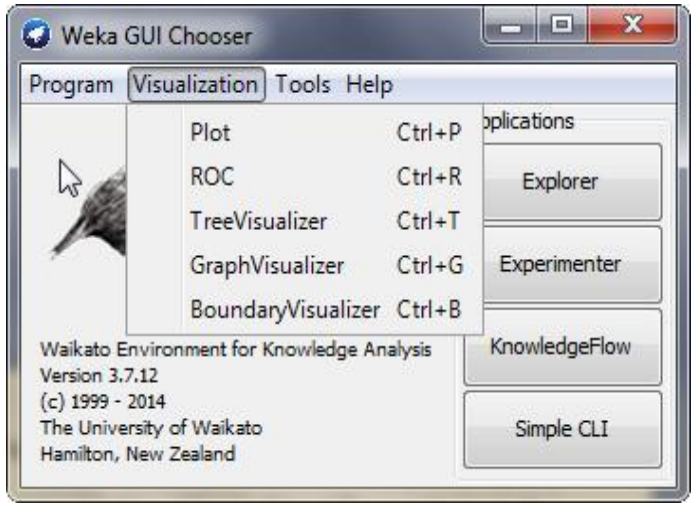
Menu Program

- **Program**
  - **LogWindow** Opens a log window that captures all that is printed to stdout
  - **Memory Usage**
  - **Exit** Closes WEKA



Menu Visualization

- **Visualization:**
  - **Plot:** For plotting a 2D plot of a dataset.
  - **ROC:** Displays a previously saved ROC curve.
  - **TreeVisualizer:** For displaying directed graphs, e.g., a decision tree.
  - **GraphVisualizer:** Visualizes XML BIF or DOT format graphs, e.g., for Bayesian networks.
  - **BoundaryVisualizer:** Allows the visualization of classifier decision boundaries in two dimensions.



Menu : Tools

- **Tools** Other useful applications.
  - **Package manager:** A graphical interface to Weka's package management system.
  - **ArffViewer:** An MDI application for viewing ARFF files
  - **SqlViewer:** Represents an SQL worksheet, for querying databases via JDBC.
  - **Bayes net editor:** An application for editing, visualizing and learning Bayes nets.

WEKA : Applications

- The buttons can be used to start the following WEKA major application:
- **Explorer**

  An environment for exploring data with WEKA

- **Experimenter**

  An environment for performing experiments and conducting statistical tests between learning schemes.

- **KnowledgeFlow**

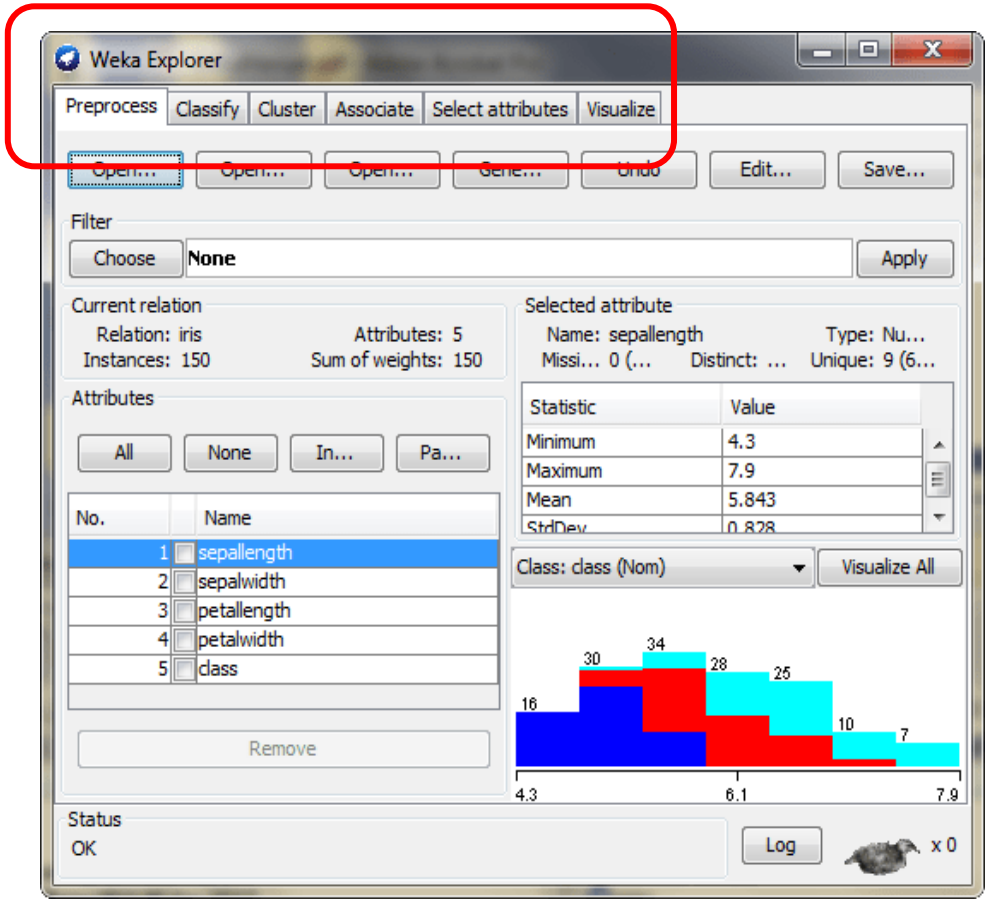
  This environment supports essentially the same functions as the Explorer but with a drag-and-drop interface. One advantage is that it supports incremental learning.

- **SimpleCLI**

Applications : Explorer

Explorer is an environment for exploring data with WEKA

- **Preprocess**. Choose and modify the data being acted on.
- **Classify**. Train and test learning schemes that classify or perform regres- sion.
- **Cluster.** Learn clusters for the data.
- **Associate**. Learn association rules for the data.
- **Select attributes**. Select the most relevant attributes in the data.
- **Visualize**. View an interactive 2D plot of the data.



Explorer: Preprocessing

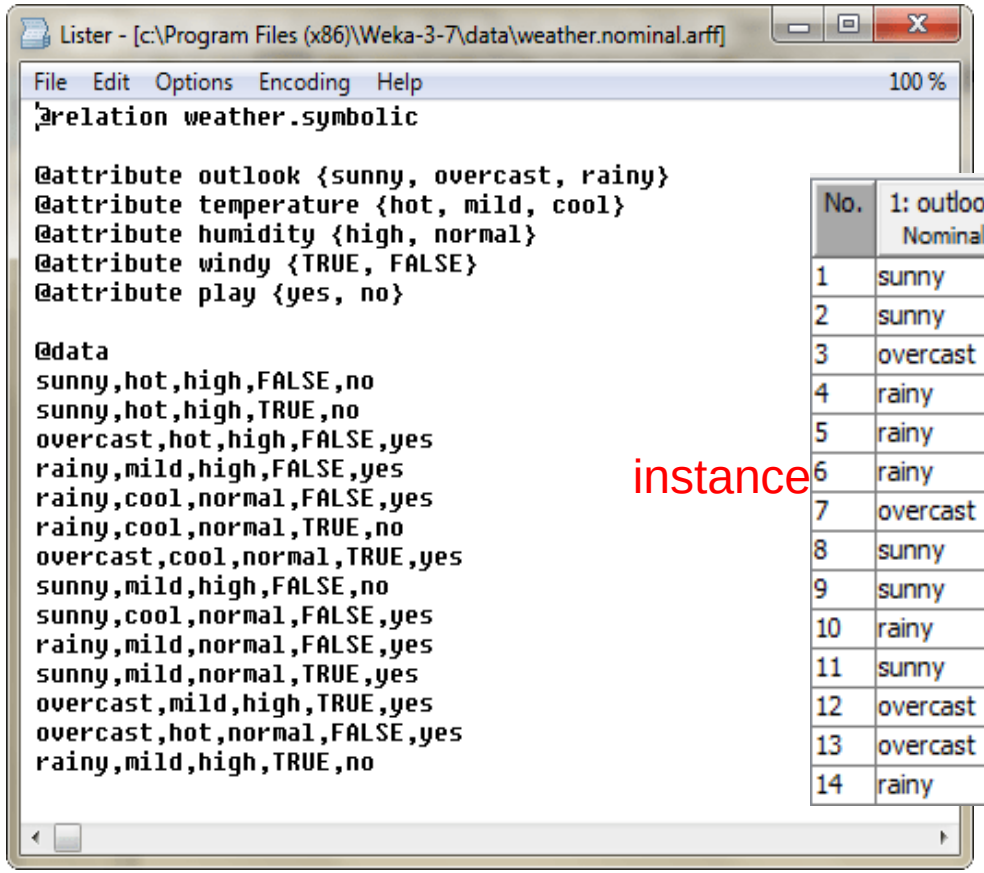
- The first four buttons at the top of the preprocess section enable you to load data into WEKA:
  - **Open file**.... Brings up a dialog box allowing you to browse for the datafile on the local file system.
  - **Open URL**.... Asks for a Uniform Resource Locator address for wherethe data is stored.
  - **Open DB**.... Reads data from a database. (Note that to make this workyou might have to edit the file in weka/experiment/DatabaseUtils.props.)
  - **Generate**.... Enables you to generate artificial data from a variety ofDataGenerators.

Supported data format

- Using the **Open file** button you can read files in a variety of formats:
  - WEKA's ARFF format, CSV format, C4.5 format, or serialized Instances format.
  - ARFF files typically have a .arff extension, CSV files a .csv extension, C4.5 files a .data and .names extension, and serialized Instances objects a .bsi extension.
- Data can also be read from a URL or from an SQL database (using JDBC)

arrf file format

- Uses flat text files to describe the data



attribute

instance

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

A more thorough description is available here
http://www.cs.waikato.ac.nz/~ml/weka/arff.html

arrf file format

- Heterogen data types

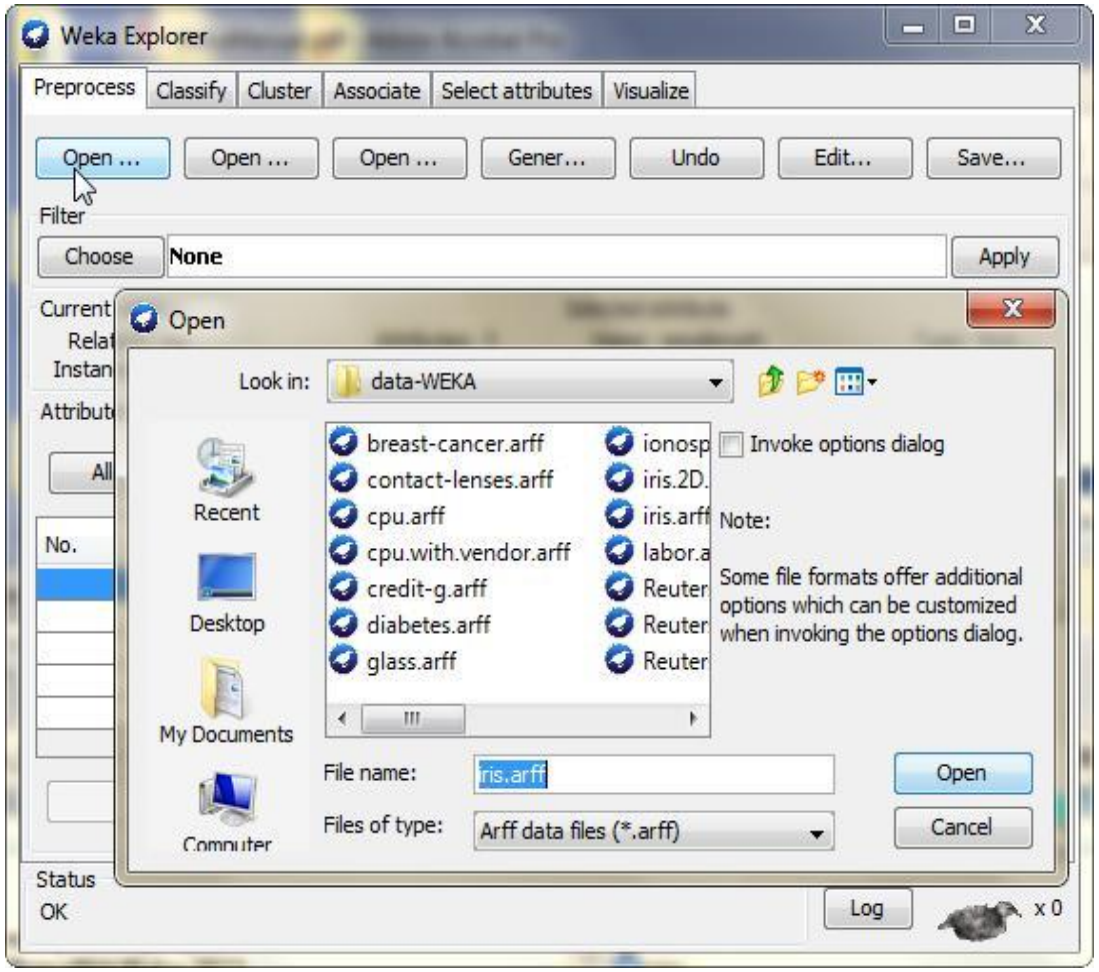
Lister - [c:\Users\user\Documents_data-WEKA\weather.nu...

File Edit Options Encoding Help 100 %

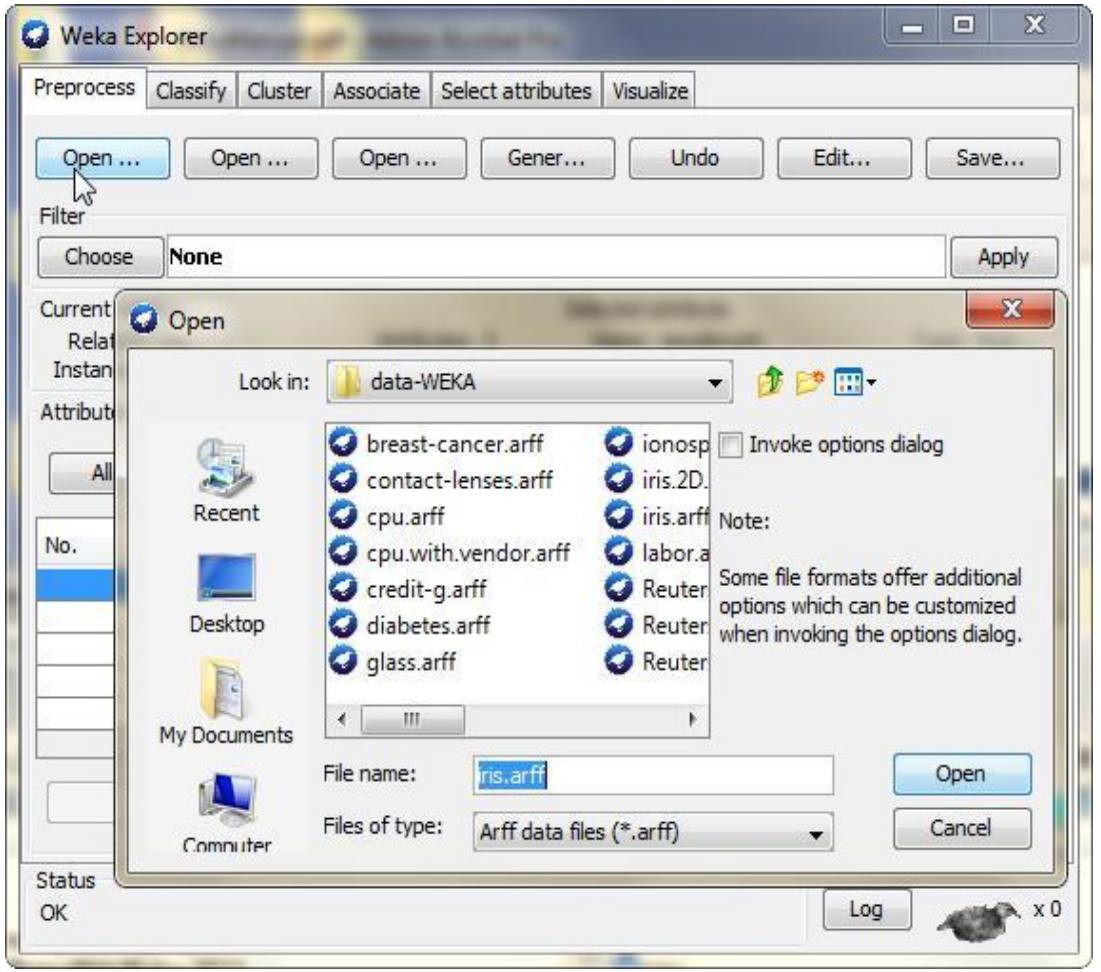
```
@relation weather

@attribute outlook {sunny, overcast, rainy}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@data
sunny,85,85,FALSE,no
sunny,80,90,TRUE,no
overcast,83,86,FALSE,yes
rainy,70,96,FALSE,yes
rainy,68,80,FALSE,yes
rainy,65,70,TRUE,no
overcast,64,65,TRUE,yes
sunny,72,95,FALSE,no
sunny,69,70,FALSE,yes
rainy,75,80,FALSE,yes
sunny,75,70,TRUE,yes
overcast,72,90,TRUE,yes
overcast,81,75,FALSE,yes
rainy,71,91,TRUE,no
```

No.	1: outlook Nominal	2: temperature Numeric	3: humidity Numeric	4: windy Nominal	5: **play** Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

Open File


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open ...
Open ...
Open ...
Gener...
Undo
Edit...
Save...
Filter
Choose
None
Apply
Open
Look in:
data-WEKA
breast-cancer.arff
contact-lenses.arff
cpu.arff
cpu.with.vendor.arff
credit-g.arff
diabetes.arff
glass.arff
Invoke options dialog
Note:
Some file formats offer additional options which can be customized when invoking the options dialog.
Recent
Desktop
My Documents
Computer
File name:
iris.arff
Files of type:
Arff data files (*.arff)
Open
Cancel
Status
OK
Log
x 0

Current Relations

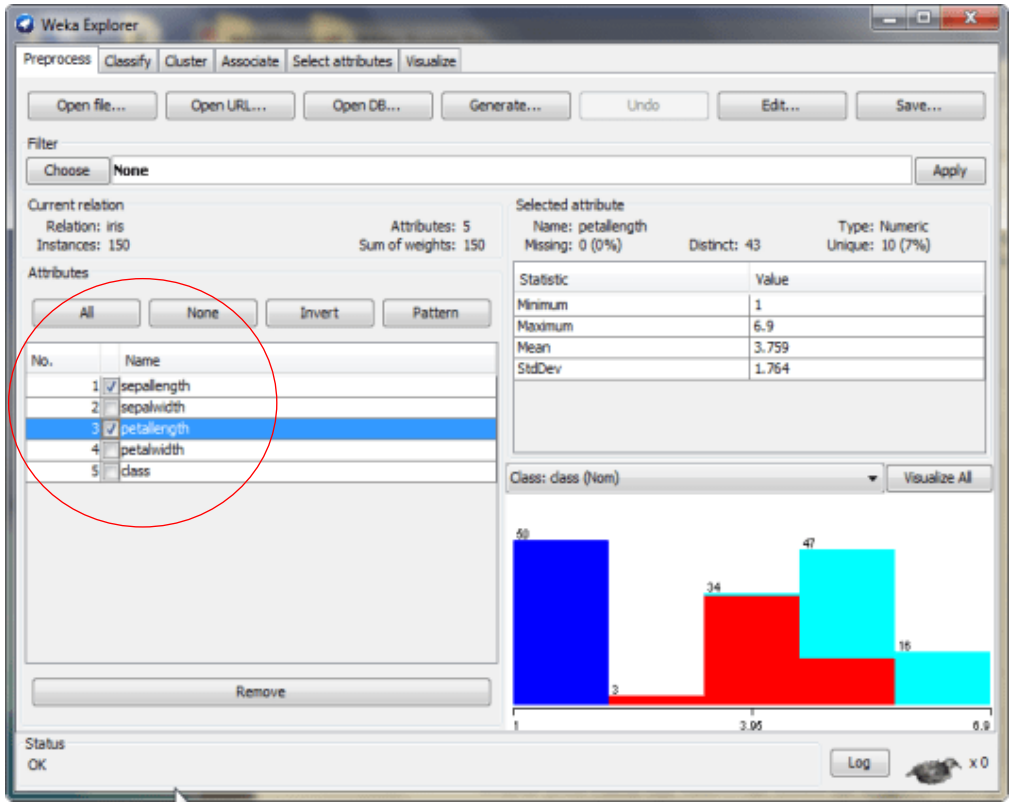


Working With Attributes

- Below the Current relation box is a box titled **Attributes**.
- There are four buttons, and beneath them is a list of the attributes in the current relation.
- The list has three columns:
  - **No.** A number that identifies the attribute in the order they are specified in the data file.
  - Selection **tick boxes**. These allow you select which attributes are present in the relation.
  - **Name**. The name of the attribute, as it was declared in the data file.

Working With Attributes

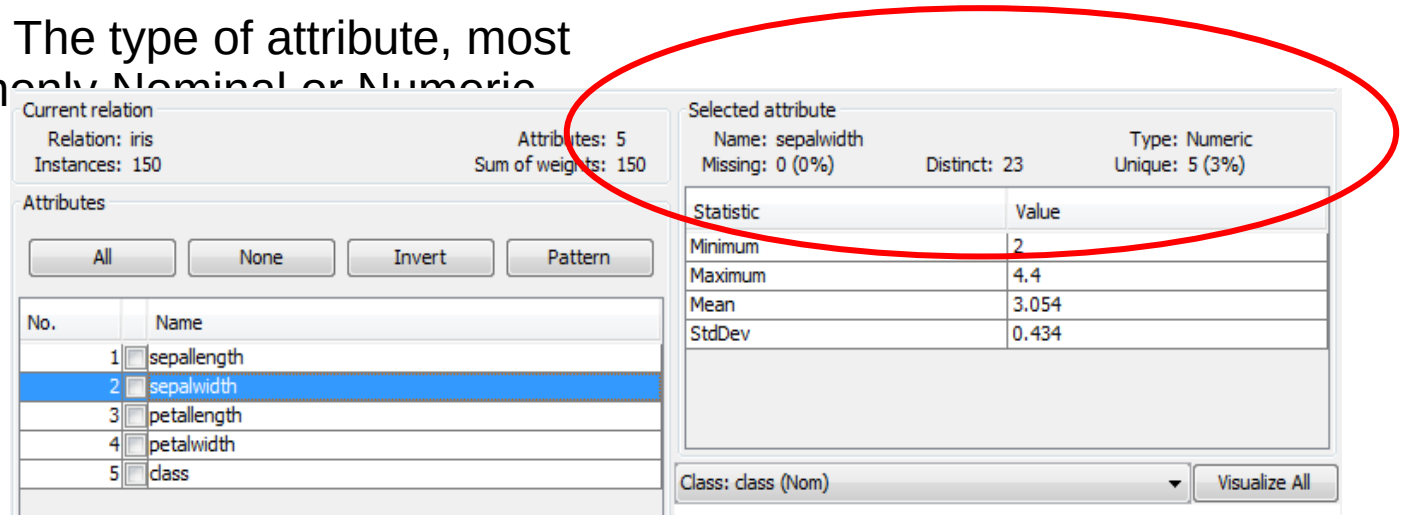
- Selection **tick boxes**. These allow you select which attributes are present in the relation



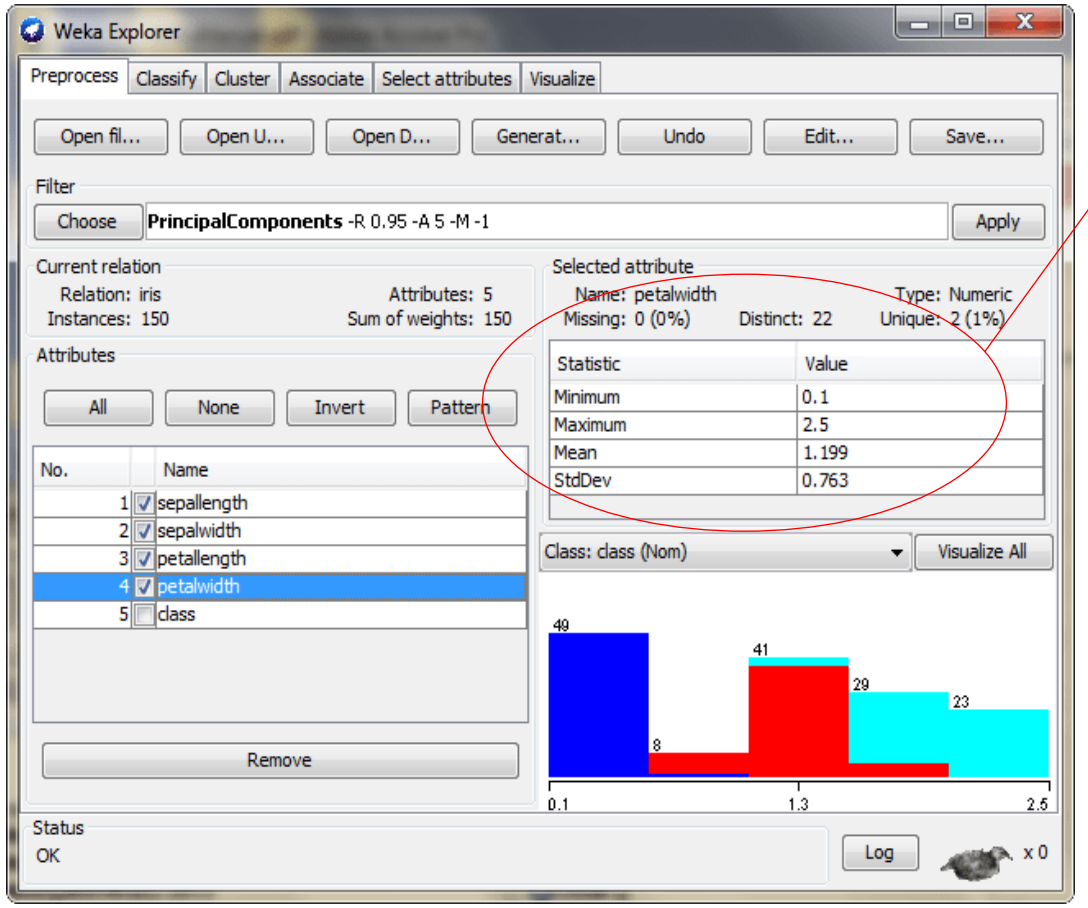
Working With Attributes ...

When you click on different rows in the list of attributes, the fields change in the box to the right titled Selected attribute.

- **Name**. The name of the attribute, the same as that given in the attribute list.
- **Type**. The type of attribute, most commonly Nominal or Numeric



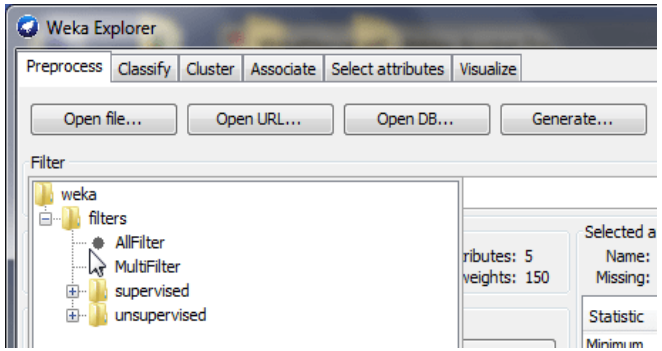
Working With Attributes ...



Statistical description of data

Working with Filters

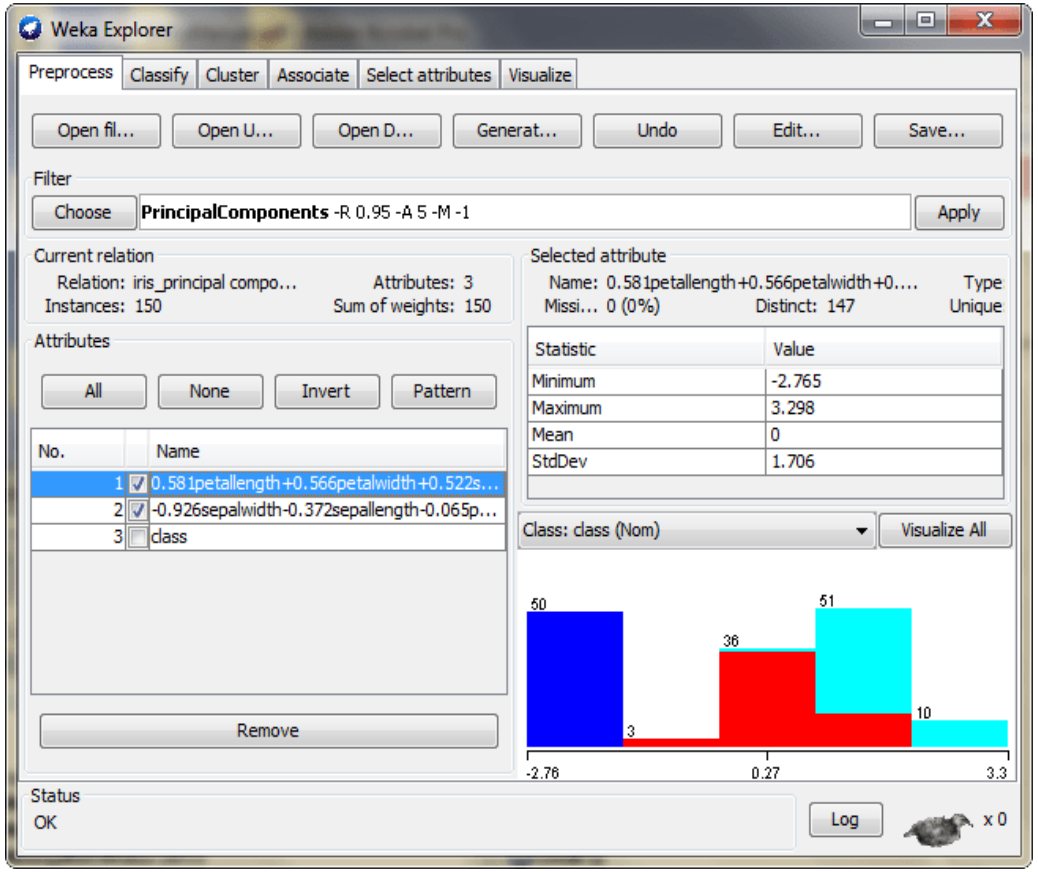
- **Preprocess:** Allows filters to be defined that transform the data in various ways.
- The **Filter** box is used to set up the filters that are required. At the left of the Filter box is a **Choose** button. By clicking this button it is possible to select one of the filters in WEKA.



- Once a filter has been selected, its name and options are shown in the field next to the Choose button.

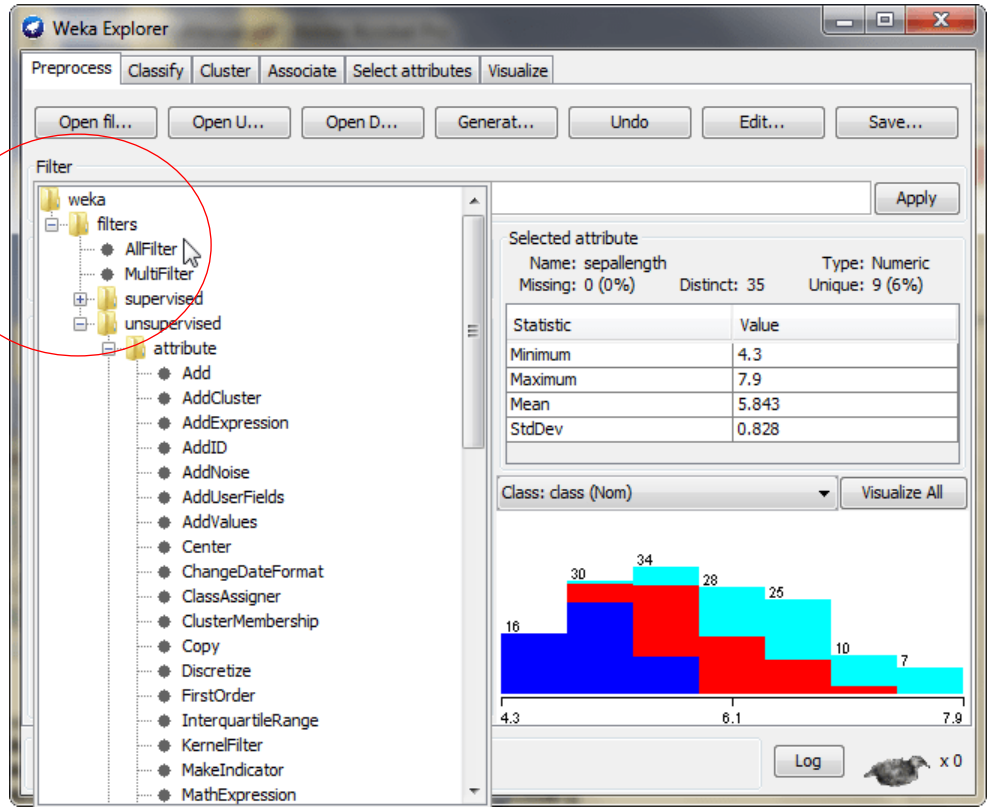
Working with Filters

- Using Principle Componen Analysis (PCA)



Working with Filters

- Filters are pre-processing tools in WEKA

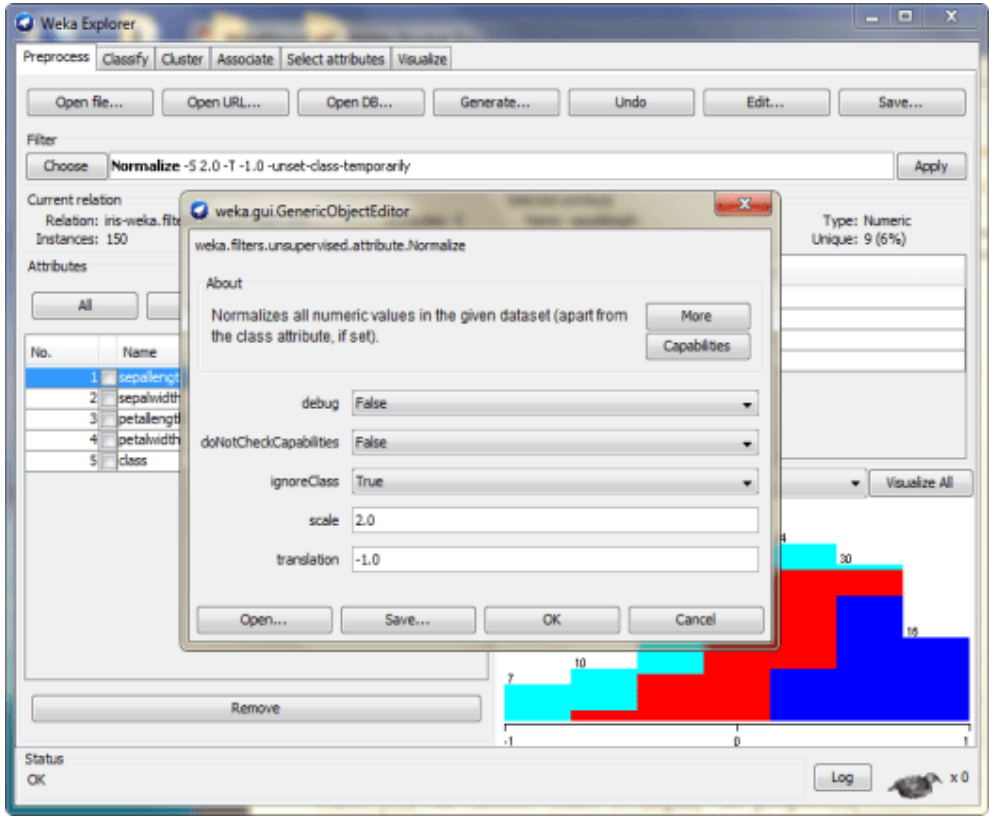


Working with Filters

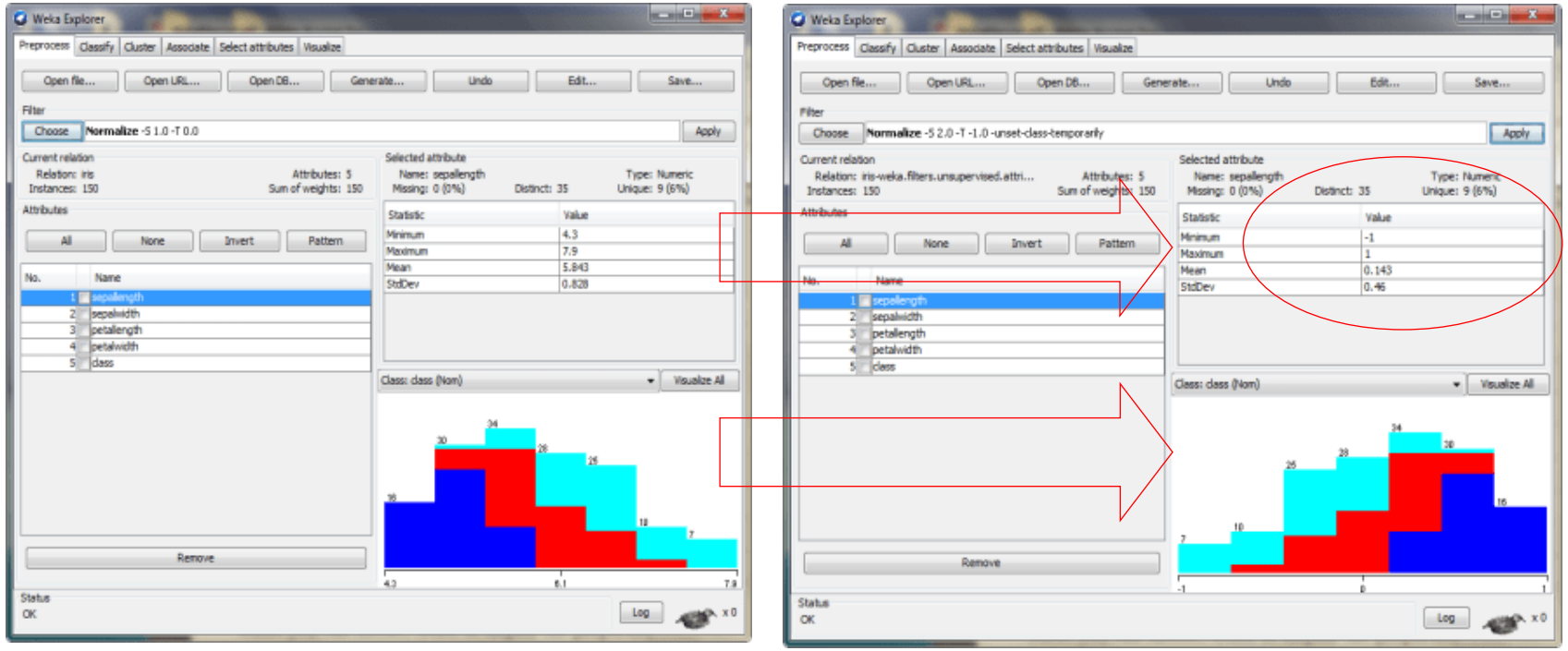
- **Applying** Filters
  - Once you have selected and configured a filter, you can apply it to the data by pressing the Apply button at the right end of the Filter panel in the Preprocess panel.
  - The Preprocess panel will then show the transformed data.
  - The result of applying filters can be saved to a new data file.
  - You can configure a filter by setting values in the **GenericObjectEditor** dialog box.

Working with Filters

- The parameters of the filter can be customized by setting values in **GenericObjectEditor** dialog box

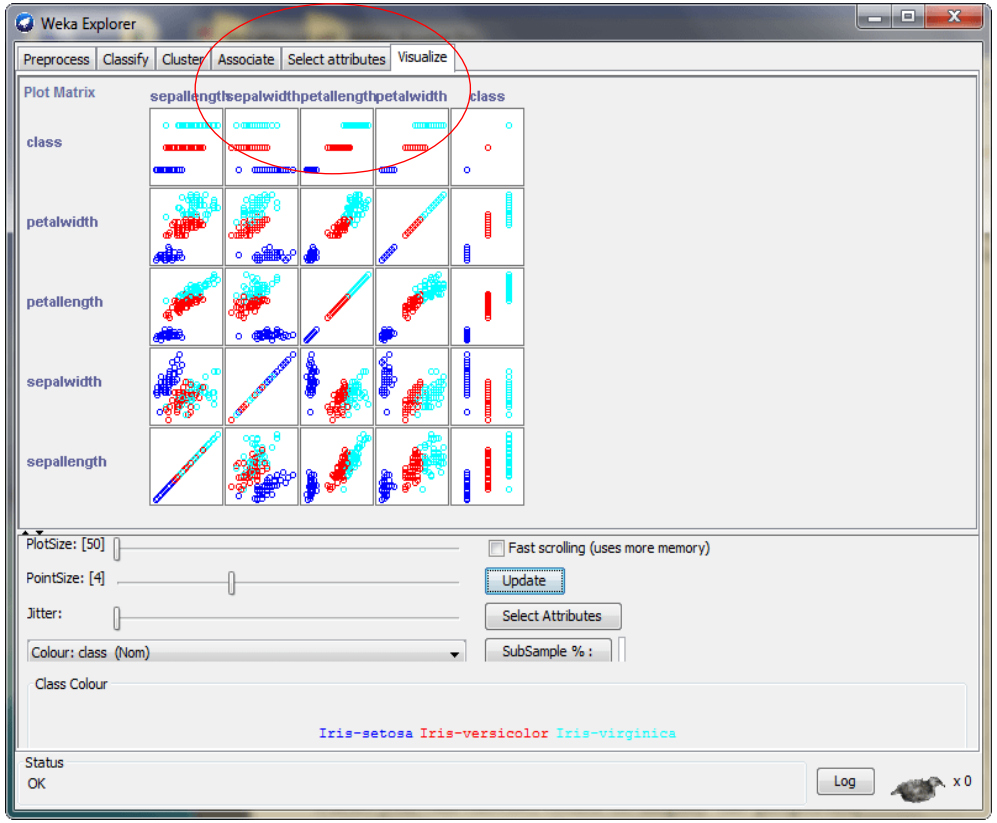


Working with Filters



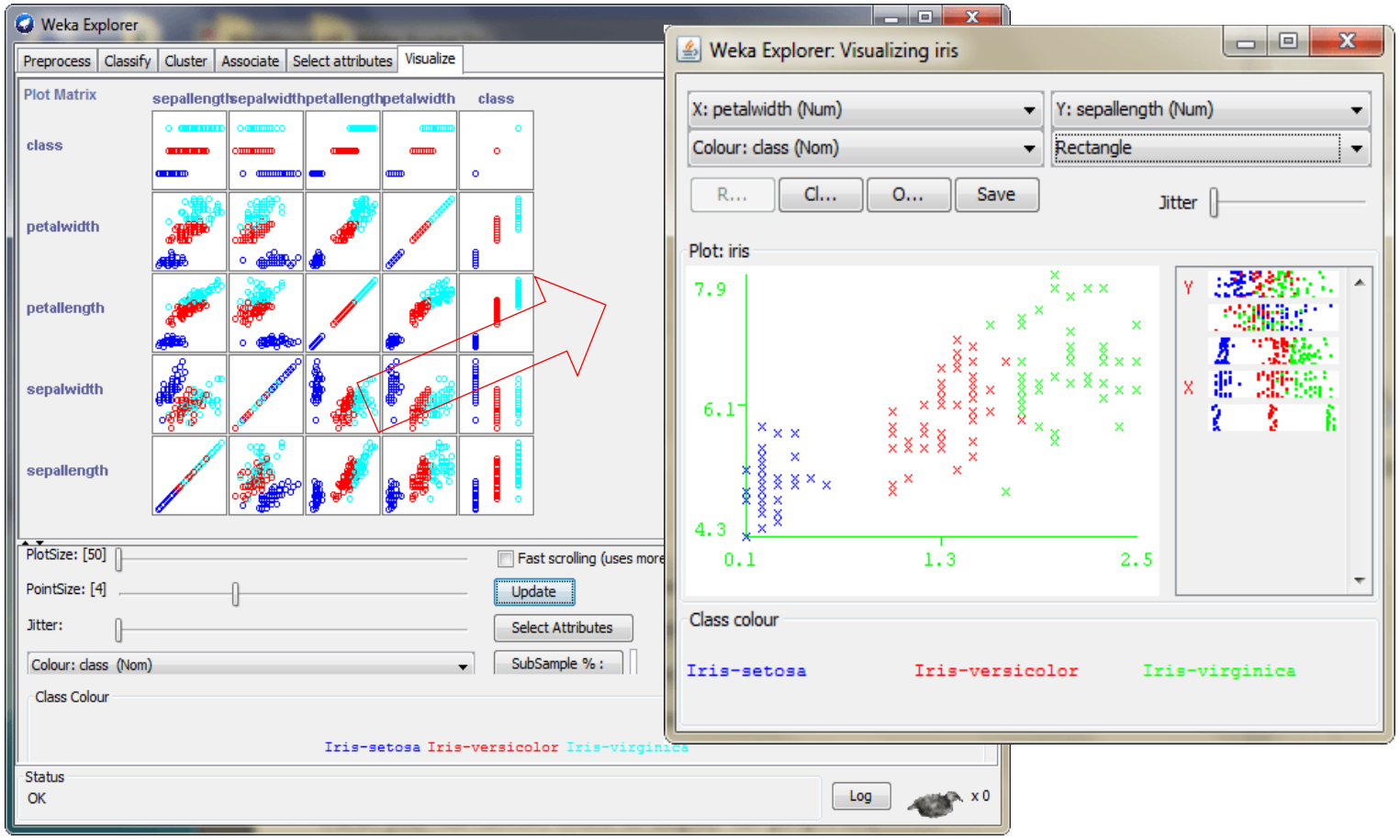
Explorer: Visualization

- **Visalize** menu



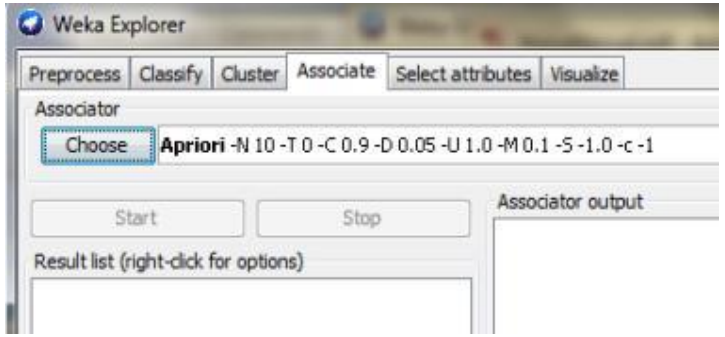
Explorer: Visualization (2)

- Weka allows to plotting data by selecting a pair data atribute



Explorer: Associate

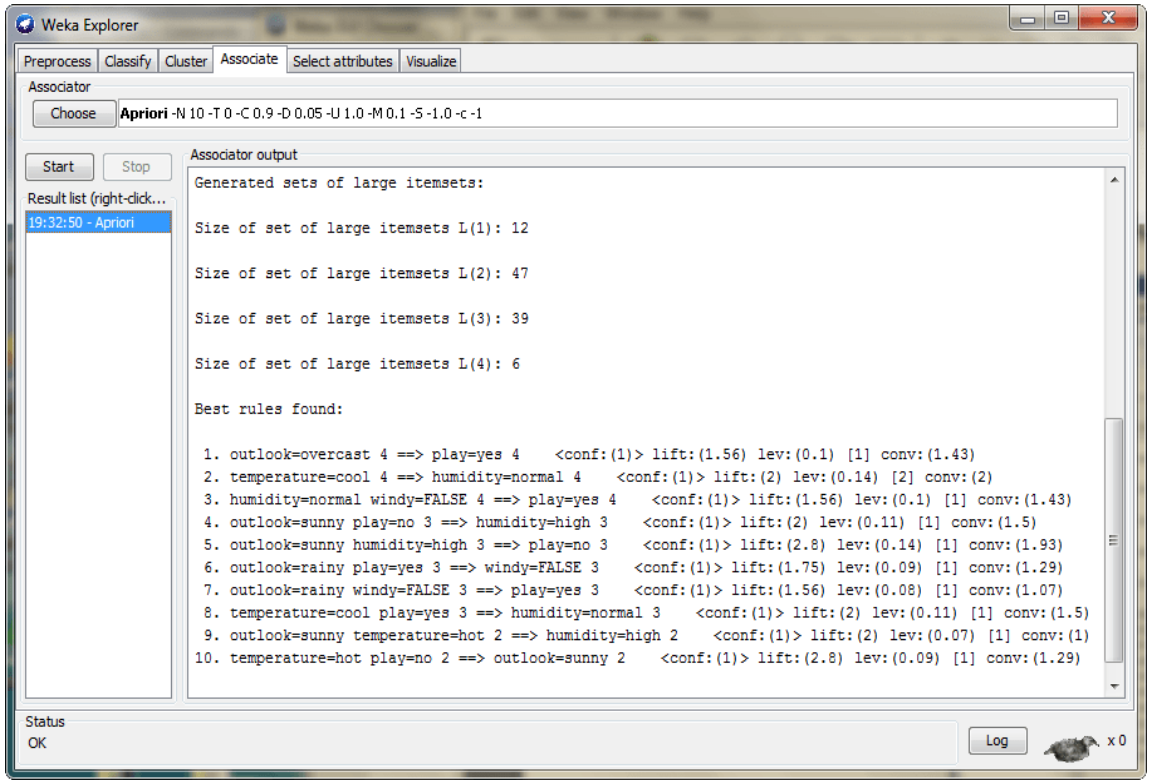
- The **Associate** panel contains schemes for learning association rules, and the learners are chosen and configured in the same way as the clusterers, filters, and classifiers.

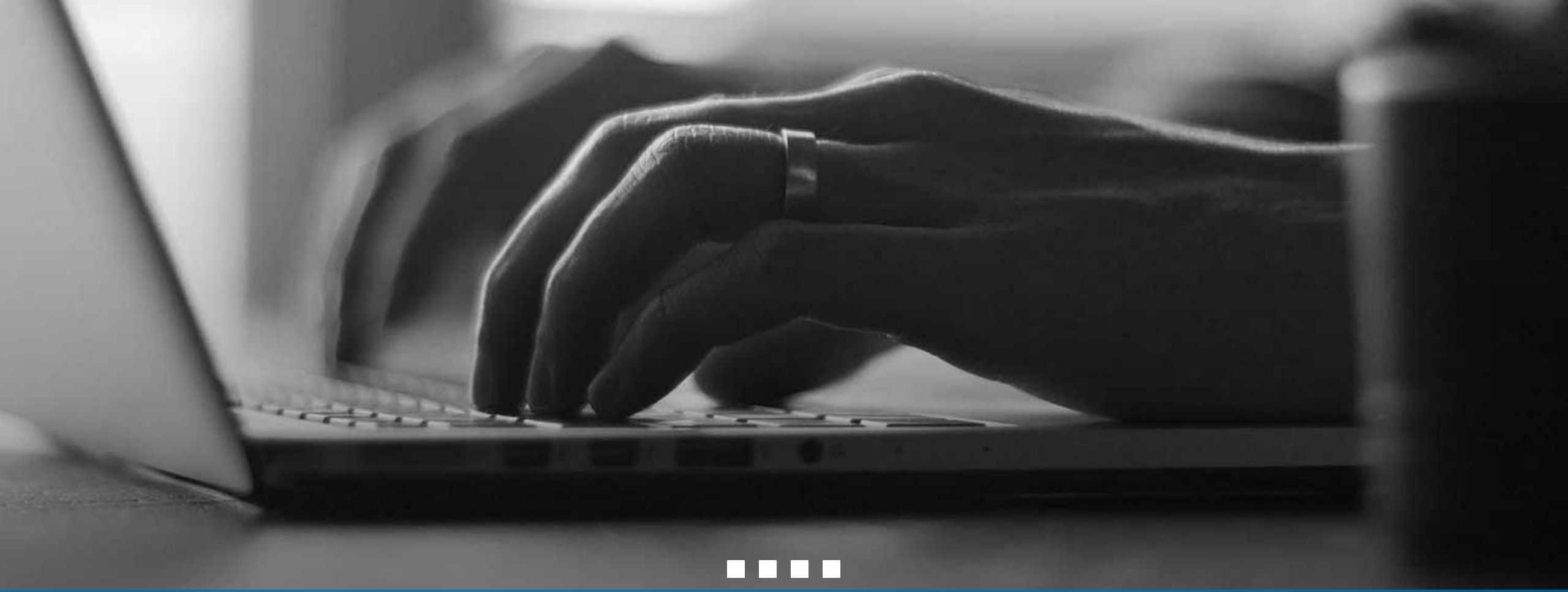


- Once appropriate parameters for the association rule learner bave been set, click the Start button.
- When complete, right-clicking on an entry in the result list allows the results to be viewed or saved.

Explorer: Associate (2)

- Experiment: Open file: weather.nominal.arff
- Choose: Apriori method





THANK YOU

Munawar, PhD – moenawar@gmail.com – www.moenawar.web.id