**Experiment No. 6**

**Aim:** Perform data Pre-processing task and demonstrate performing

Classification, Clustering, Association algorithm on data sets using

data mining tool (WEKA/R tool).

**Theory:**

The Weka GUI Chooser (class weka.gui.GUIChooser) provides a starting

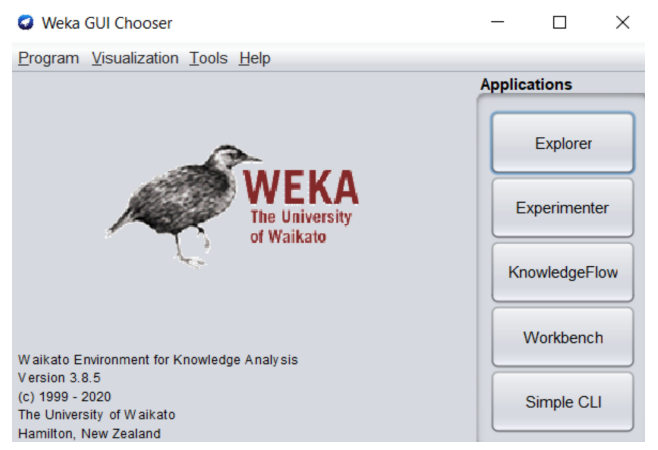
point for launching Weka‘s main GUI applications and supporting tools. If

one prefersa MDI (―multiple document interface‖) appearance, then

this is provided by an alternative launcher called ―Main‖ (class

weka.gui.Main). The GUI Chooser consists of four buttons—one for each

of the four major Weka applications— and four menus.



The buttons can be used to start the following applications:

Explorer - An environment for exploring data with WEKA

a) Click on ―explorer button to bring up the explorer window.

b) Make sure the ―preprocess tab is highlighted.

c) Open a new file by clicking on ―Open New file‖ and choosing

a file with ―.arff‖ extension from the ―Data‖ directory.

d) Attributes appear in the window below.

e) Click on the attributes to see the visualization on the right.

f) Click ―visualize all‖ to see them all.

Experimenter - An environment for performing experiments

and conducting statistical tests between learning schemes.

a) Experimenter is for comparing results.

b) Under the ―set up tab click ―New.

c) Click on ―Add New under ―Data frame. Choose a couple

of arff format files from ―Data‖ directory one at a time.

d) Click on ―Add New under ―Algorithm frame. Choose

several algorithms, one at a time by clicking ―OK in the

window and ―Add New.

e) Under the ―Run tab click ―Start‖

f) Wait for WEKA to finish.

g) Under ―Analyses tab click on ―Experiment to see results.

Knowledge Flow - 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 - Provides a simple command-line interface that

allows direct execution of WEKA commands for operating

systems that do not provide their own command line interface.

Navigate the options available in the WEKA (ex. Select

attributes panel, Preprocess panel, classify panel, Cluster

panel, Associate panel and Visualize panel)

When the Explorer is first started only the first tab is active;

the others are greyed out. This is because it is necessary to

open (and potentially pre-process) a data set before starting to

explore the data.

The tabs are as follows:

1. Preprocess. Choose and modify the data being acted on.

2. Classify. Train and test learning schemes that classify or perform

regression.

3. Cluster. Learn clusters for the data.

4. Associate. Learn association rules for the data.

5. Select attributes. Select the most relevant attributes in the

data.

6. Visualize. View an interactive 2D plot of the data. Once the tabs

are active, clicking on them flicks between different screens, on

which the respective actions can be performed. The bottom area of

the window (including the status box, the log button, and the Weka

bird) stays visible regardless of which section you are in

Loading Data:

The first four buttons at the top of the preprocess section

enable you to load data into WEKA:

1. Open file.... Brings up a dialog box allowing you to browse

for the datafile on the local file system.

2. Open URL.... Asks for a Uniform Resource Locator address

for where the data is stored.

3. Open DB.....Reads data from a database. (Note that to make

this work you might have to edit the file in

weka/experiment/DatabaseUtils.props.)

4. Generate.... Enables you to generate artificial data from a

variety ofDataGenerators.

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,

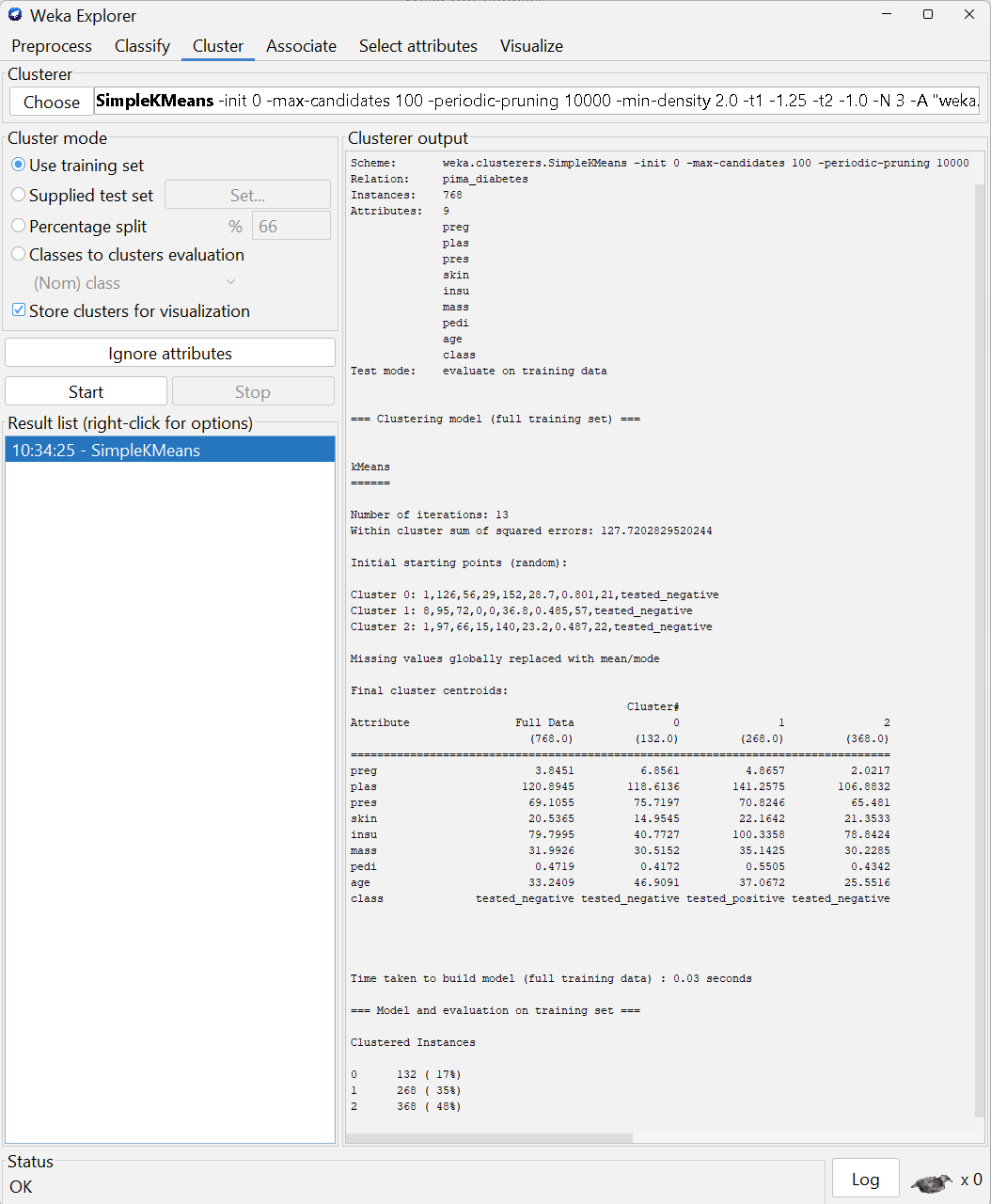
CSV files a .csv extension, C4.5 files a .data and .names

extension, and serialized Instances objects a .bsiextension.

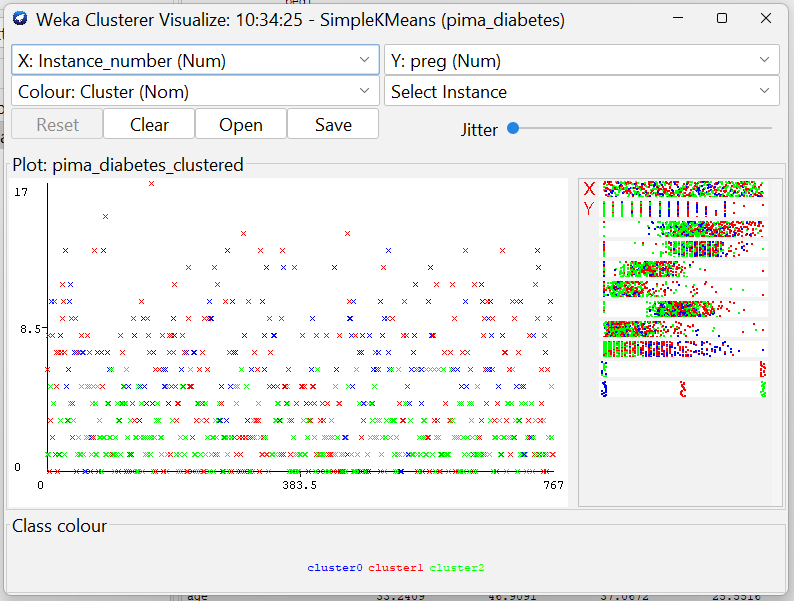
**Output:**

**K-means Clustering:**

**Clustering:**

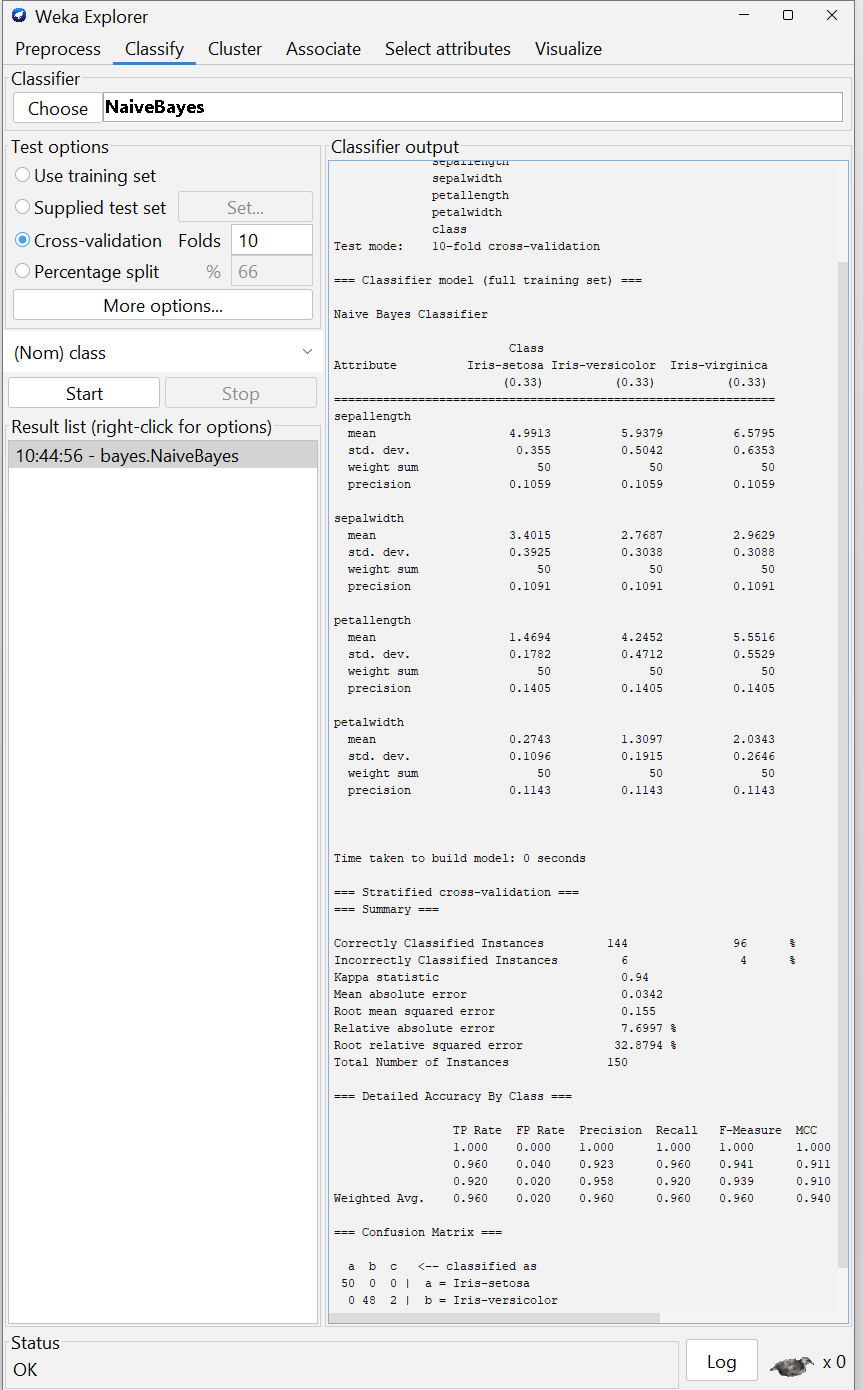
****

**Visualization:**

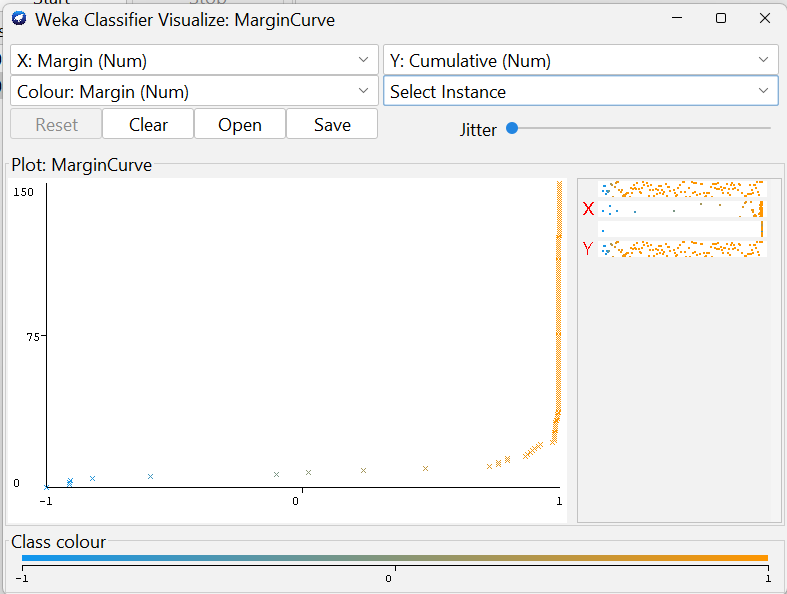
****

**Naïve Bayes Classification:**

**Classification:**

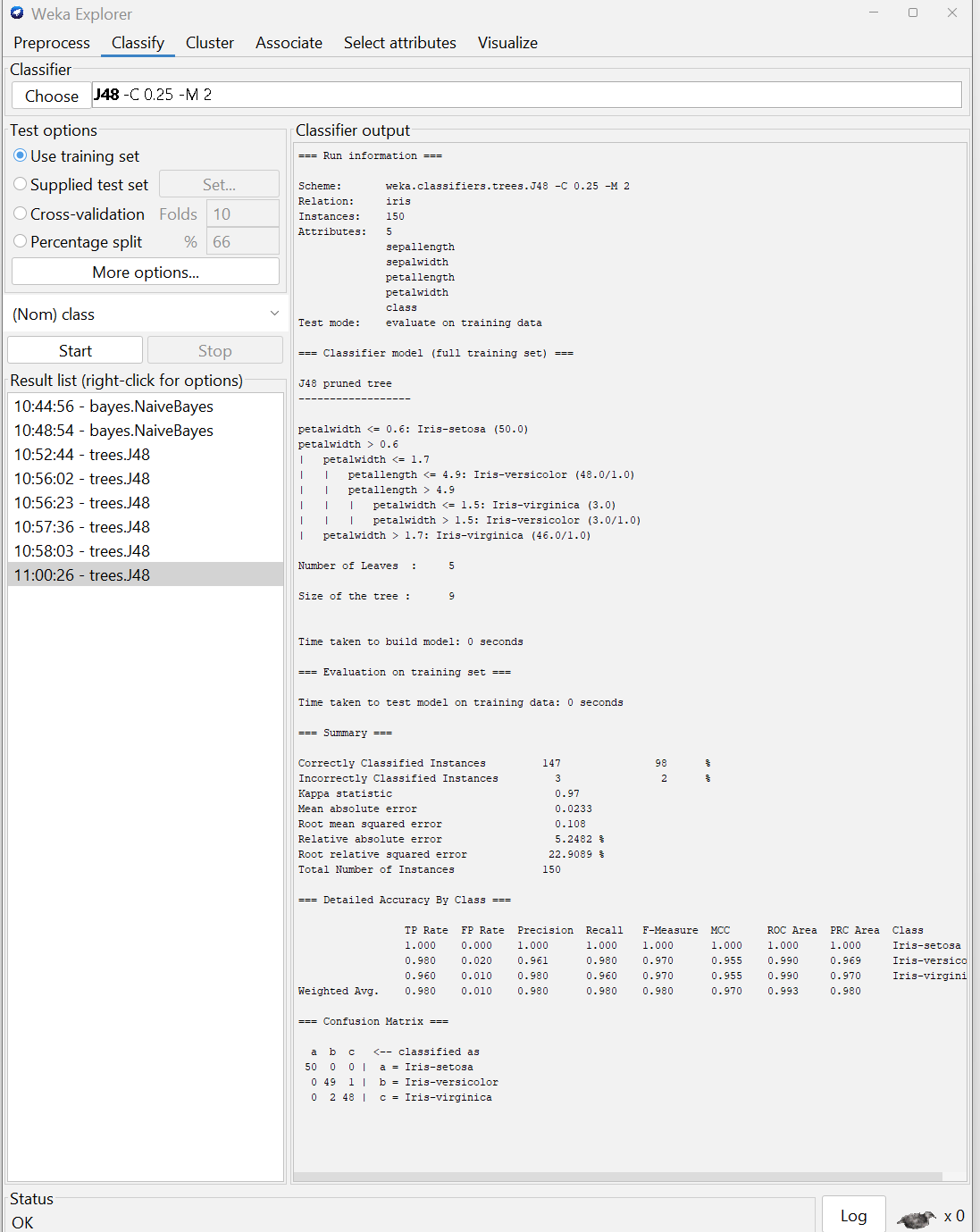
****

**Visualization:**

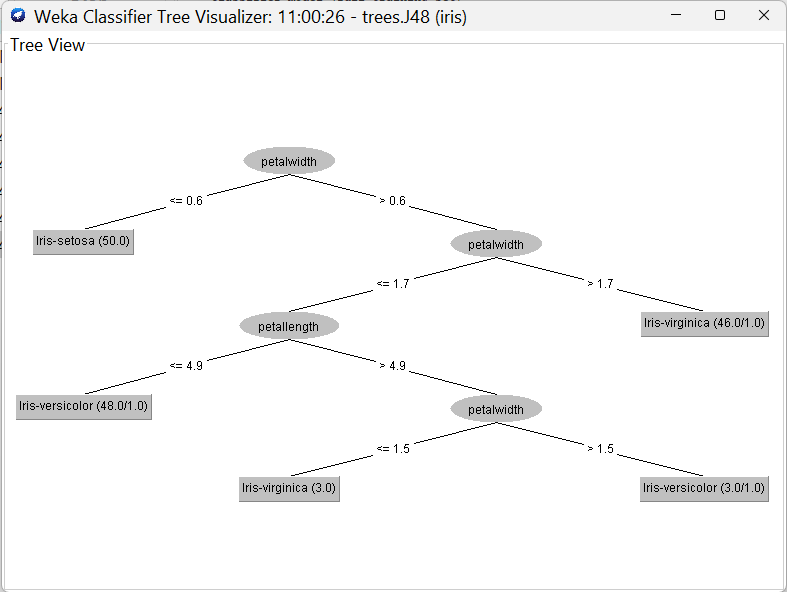
****

**Decision Tree Classification:**

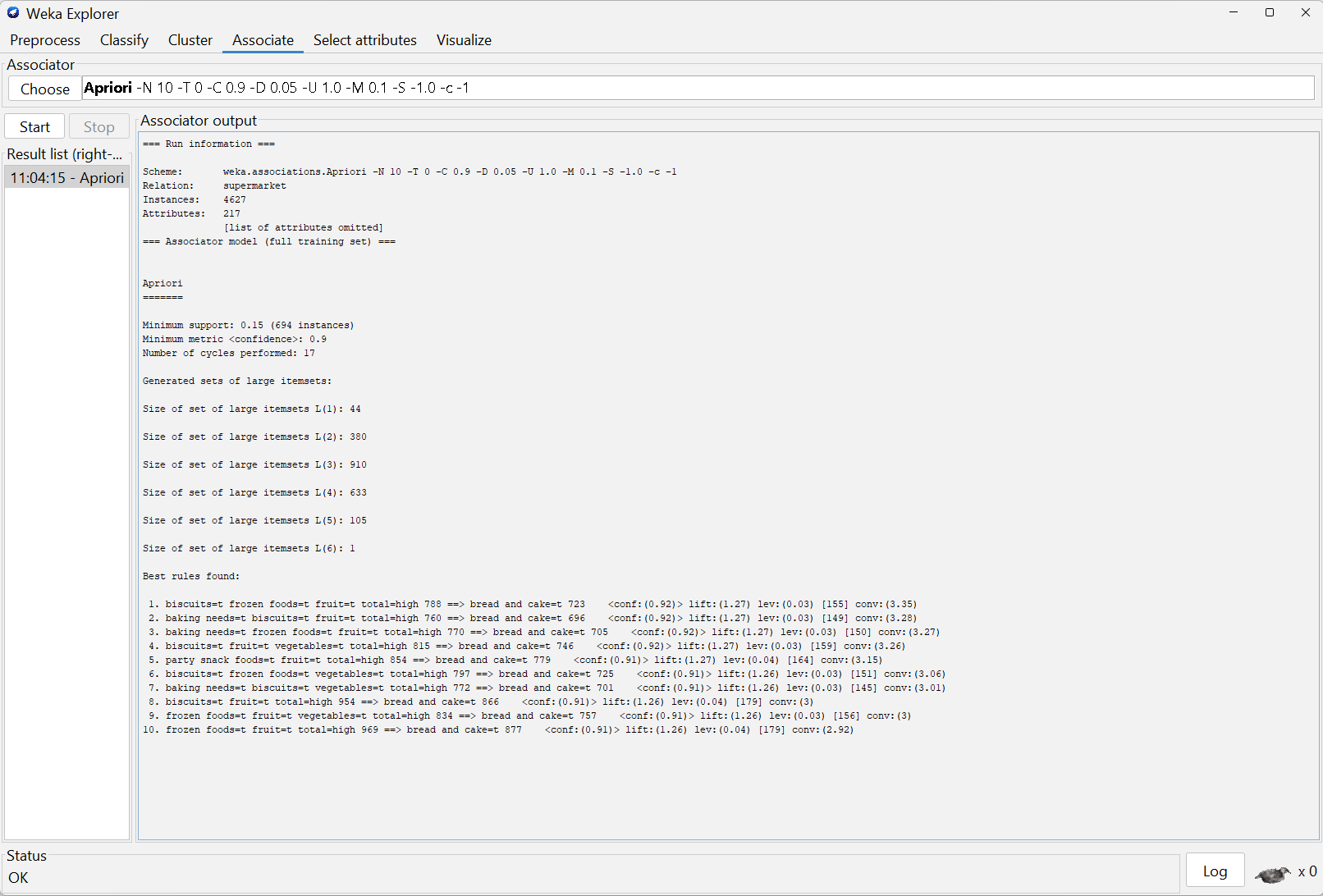
**Classification:**

****

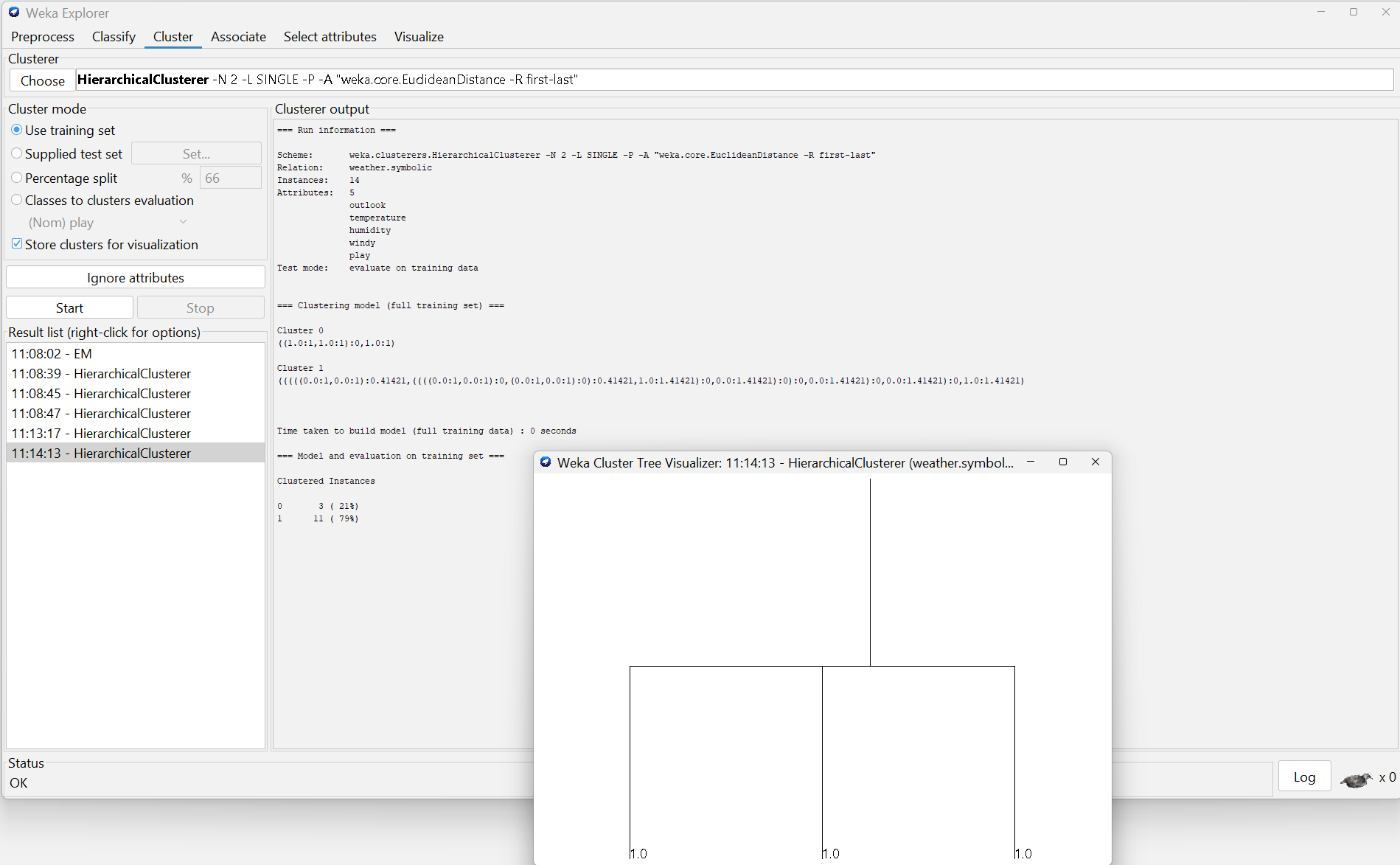
**Visualization:**

****

**Apriori:**

****

**Agglomerative:**

****

**Conclusion:** Thus, in this experiment, we have performed data Pre-processing task and demonstrated performing Classification, Clustering, Association algorithm on data sets using data mining tool (WEKA/R tool).