Experiment No : 5

Name : Dhiraj Ravindra Bodake
Roll No : 18141216

Title : Exploring weka explorer for loading & fillening file, Training & testing learning schema.

Theory : Weka is an open source collection of data mining take task which you can utilize in no. of different ways. It comes with graphical user interface but can also be called from your own Java Code.

A] Opening file from local file system.
Click on 'open file' button select CSV file & if you like to convert this file into ARFF format, click on 'save' button. WEKA automatically creat ARFF file from your CSV file.

B] Reading Data from databases.
- Data can be read from SQL database using JDBC. Click on 'open DB' button. 'Generic Object Editor' appears on screen.
- To read data from database. Click on 'open' button & select database from file system.

- once data is loaded, WEKA recognize attribute in 'Attribute' window.
- Note that data can be loaded from CSV file as well as because some databases have the ability to convert only into CSV format.

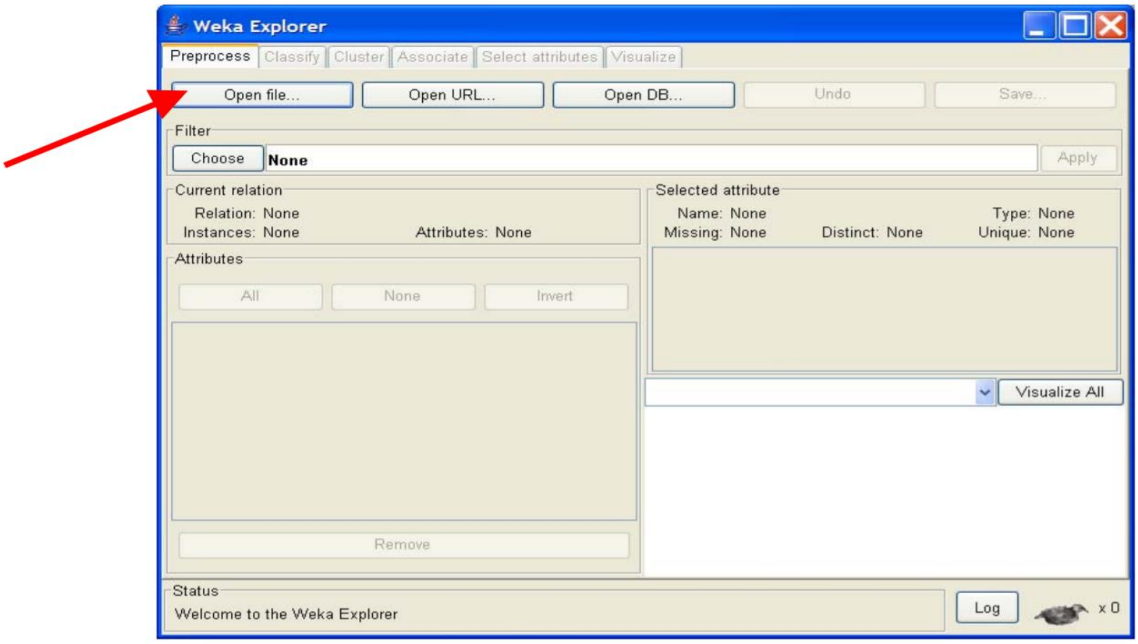
c] Filtering data in WEKA.

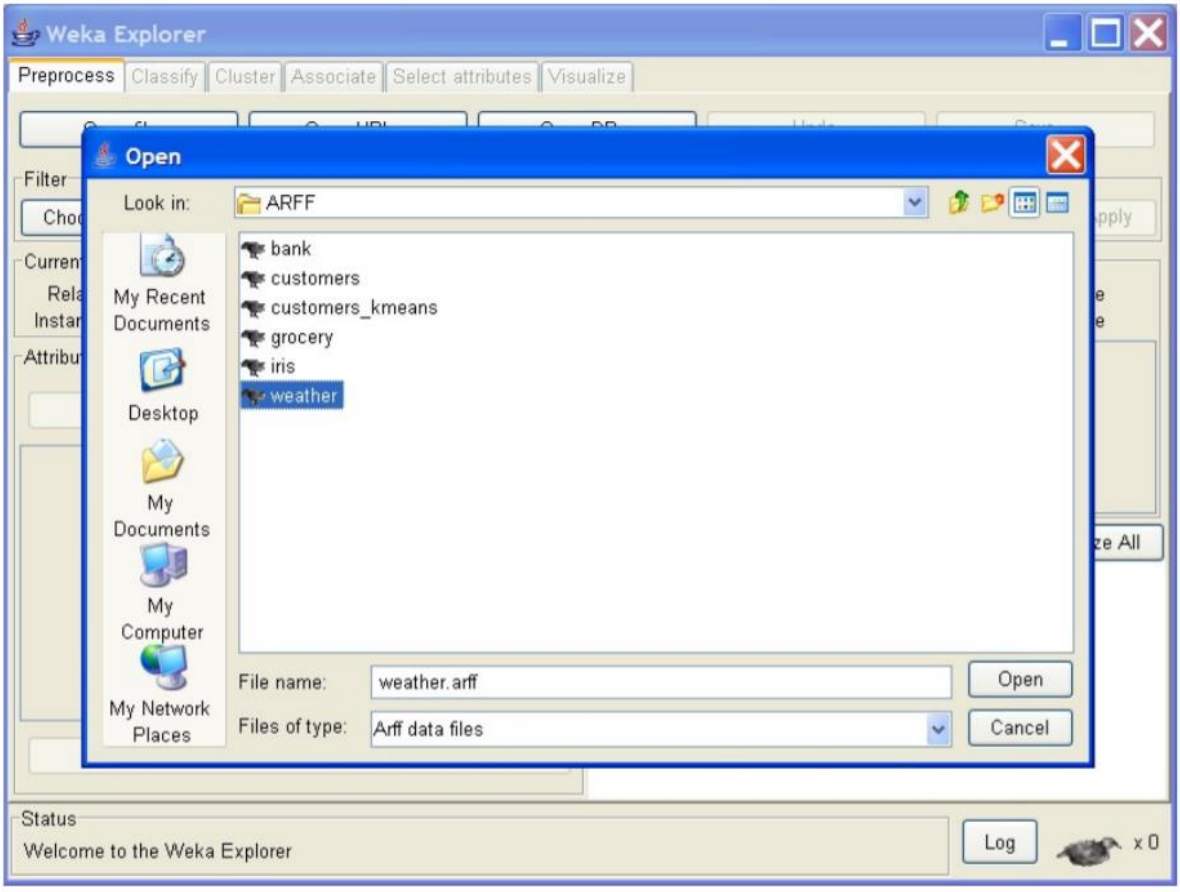
Pre-processing tools in WEKA are called filters. WEKA contains filter for device normalization, resampling, transformation and combination of attributes some techniques such as association rule mining, can only be performed on categoneal data.

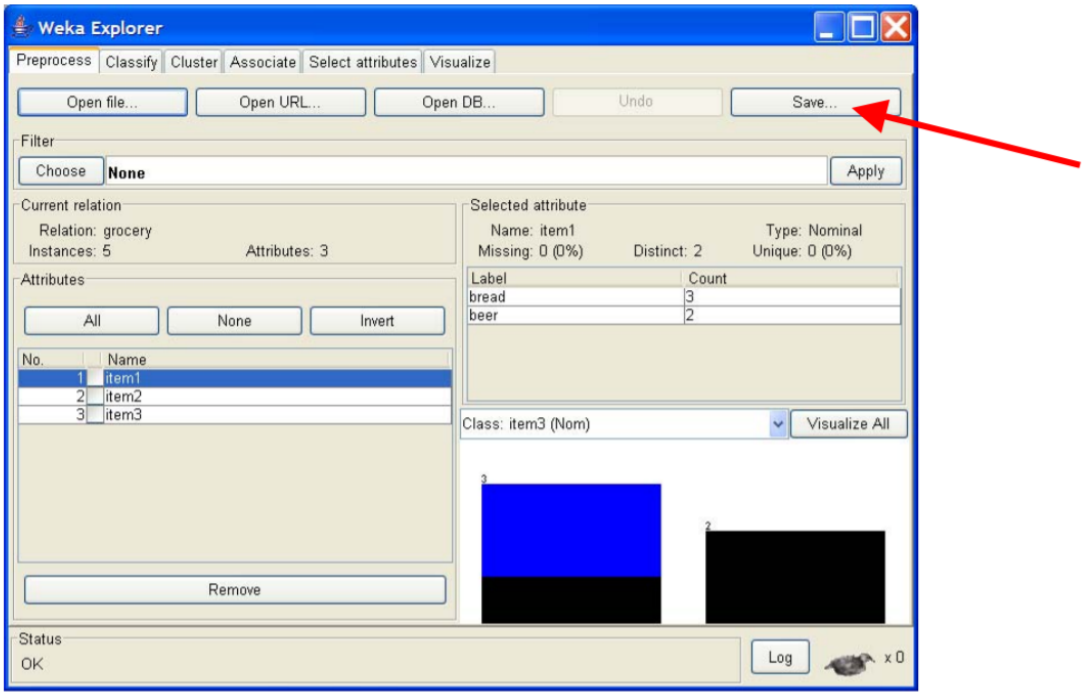
Ex: Consider weather.arff data set. There are two attributes that need to be converted 'temperature' & 'Humidity'.

- You can describe by removing keyword 'sumeric' as type of 'temperature' attribute and replace it with 'nominal' values.
- In 'Filters' window in WEKA explorer Click on 'choose' button, window changes to reflect available options.

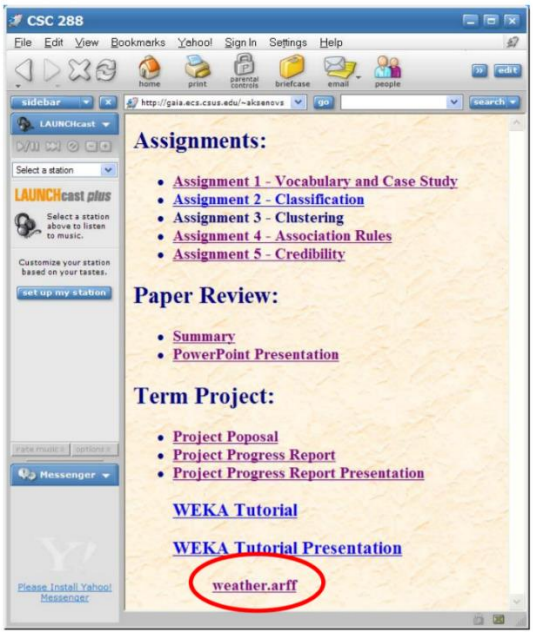
- select temperature value, look at 'selected attribute' box. 'Type' field show attribute type has changed from numeric to nominal. The list also changes.
- On right click on filter, 'generic object of editor' dialog box comes on your screen. It sets you to choose filter configure option.
- Clicking 'more' button brings up an 'Information' window describing what different options can do.

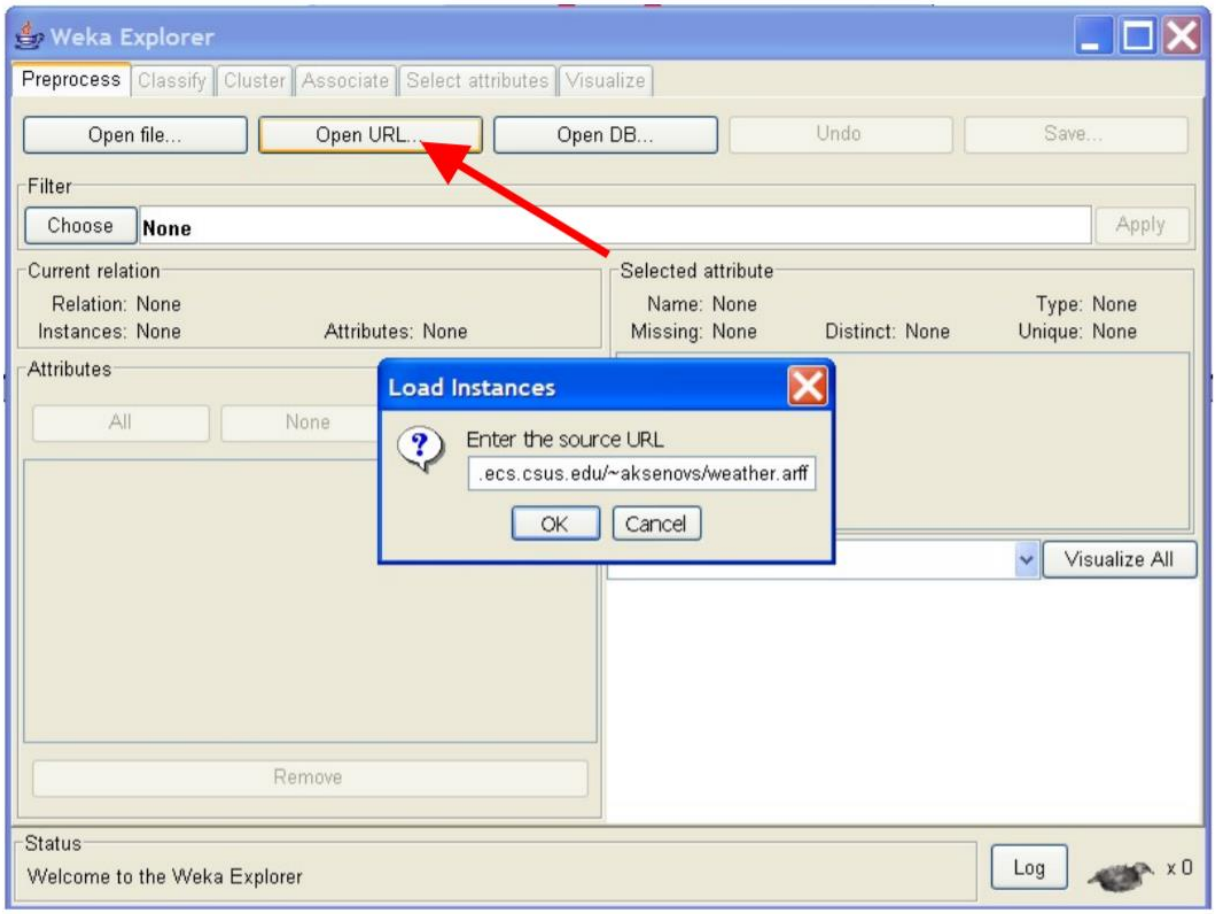

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
None
Apply
Current relation
Relation: None
Instances: None
Attributes: None
Selected attribute
Name: None
Type: None
Missing: None
Distinct: None
Unique: None
Attributes
All
None
Invert
Visualize All
Remove
Status
Welcome to the Weka Explorer
Log
x 0


Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open
Look in:
ARFF
bank
customers
customers_kmeans
grocery
iris
weather
My Recent Documents
Desktop
My Documents
My Computer
My Network Places
File name:
weather.arff
Open
Files of type:
Arff data files
Cancel
Status
Welcome to the Weka Explorer
Log
x 0

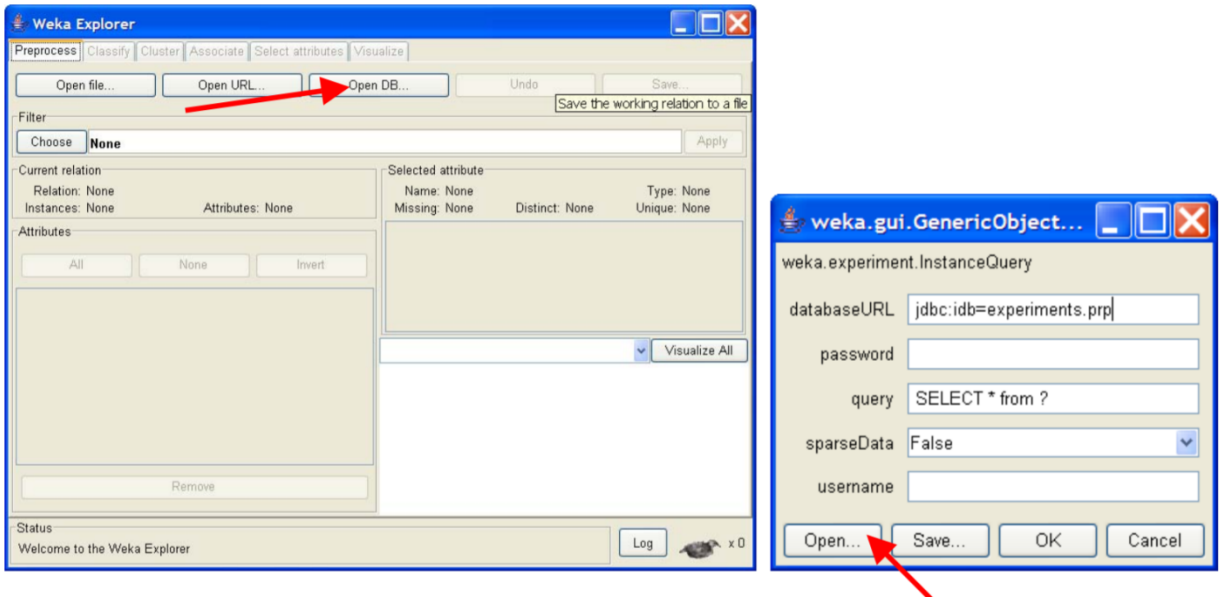


Opening file from a web site:

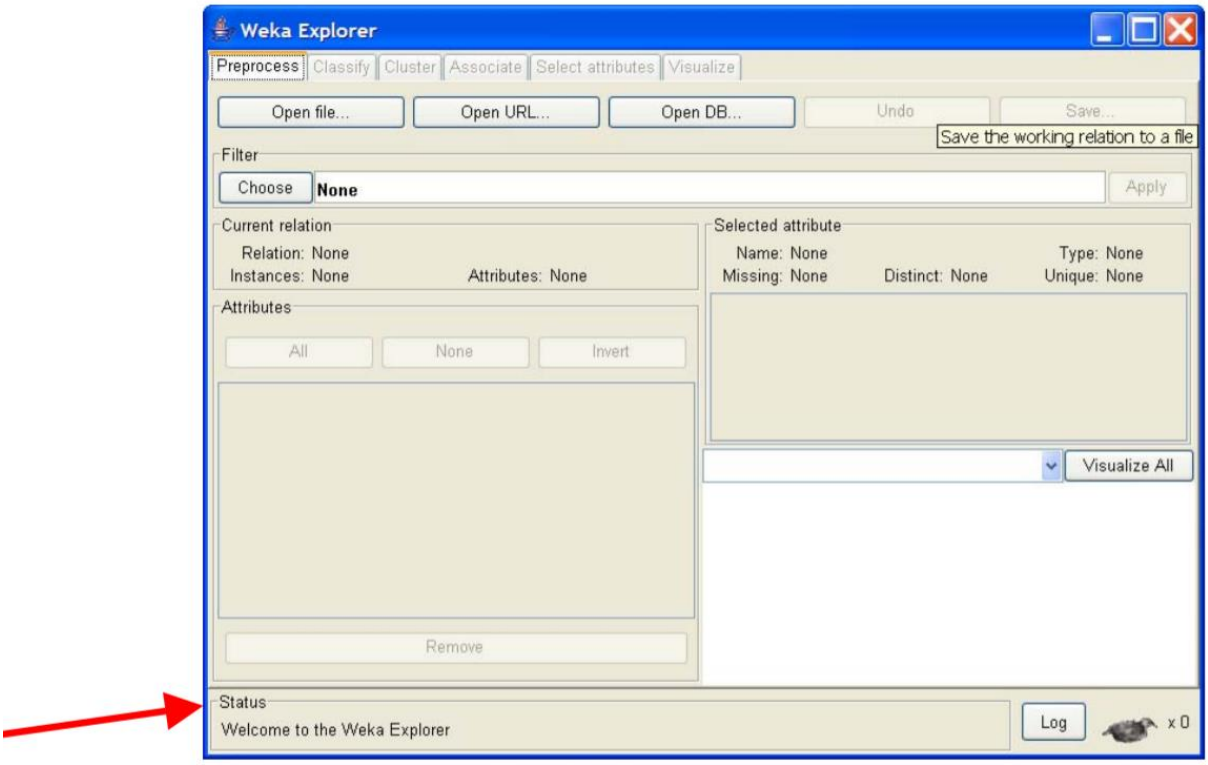


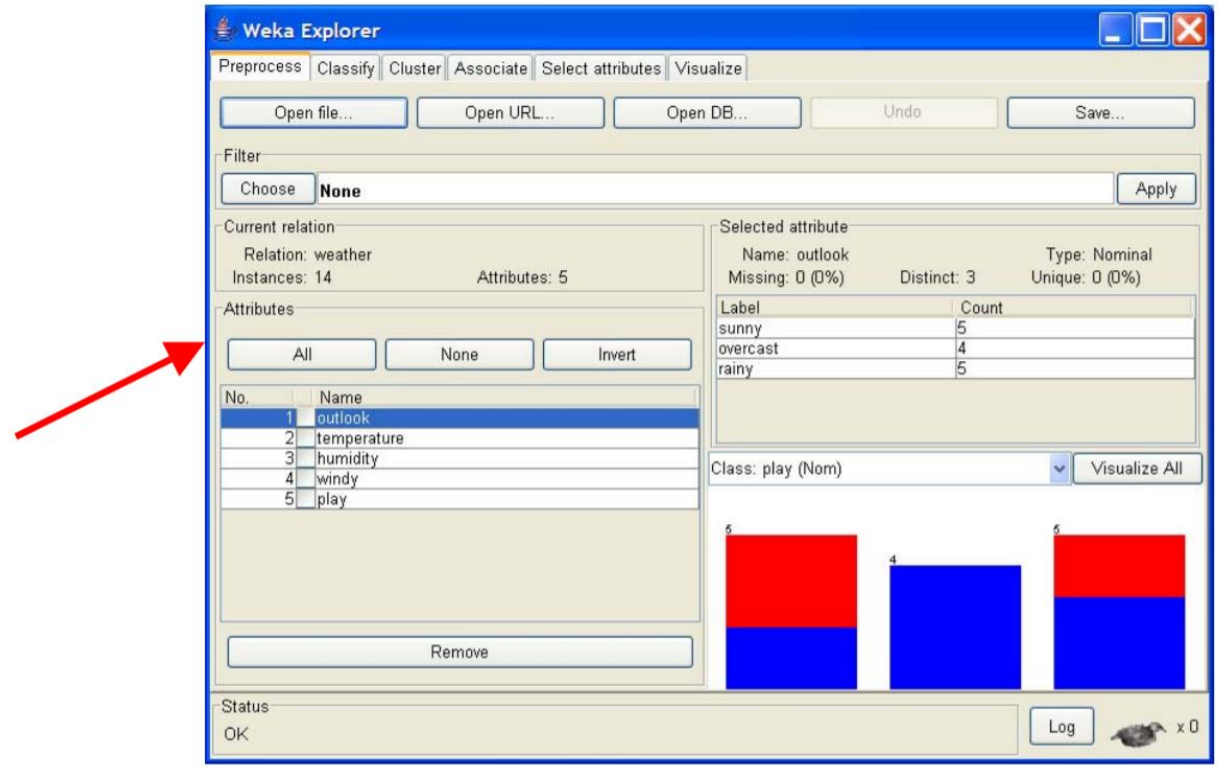


A) Reading data from a database:

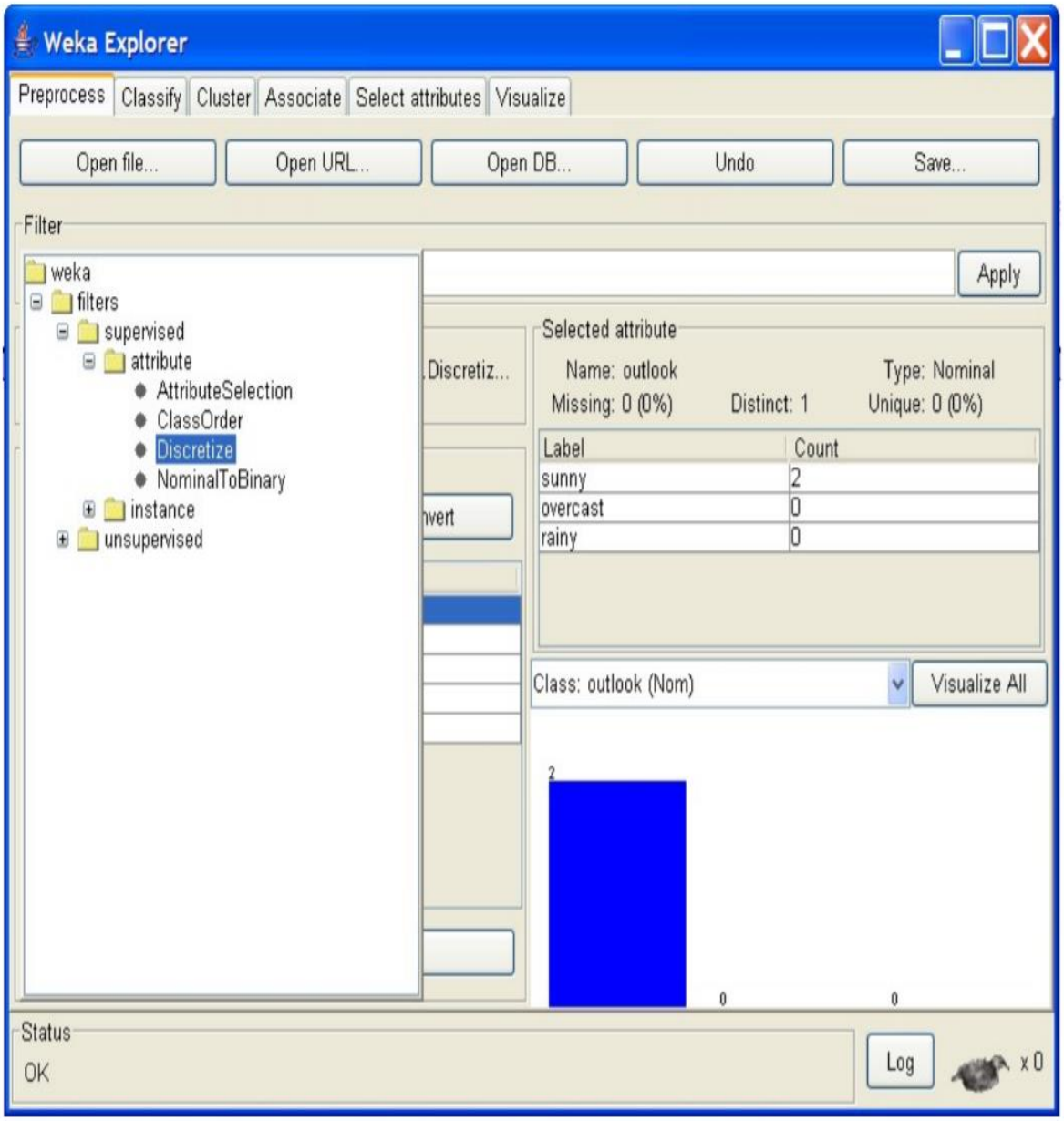


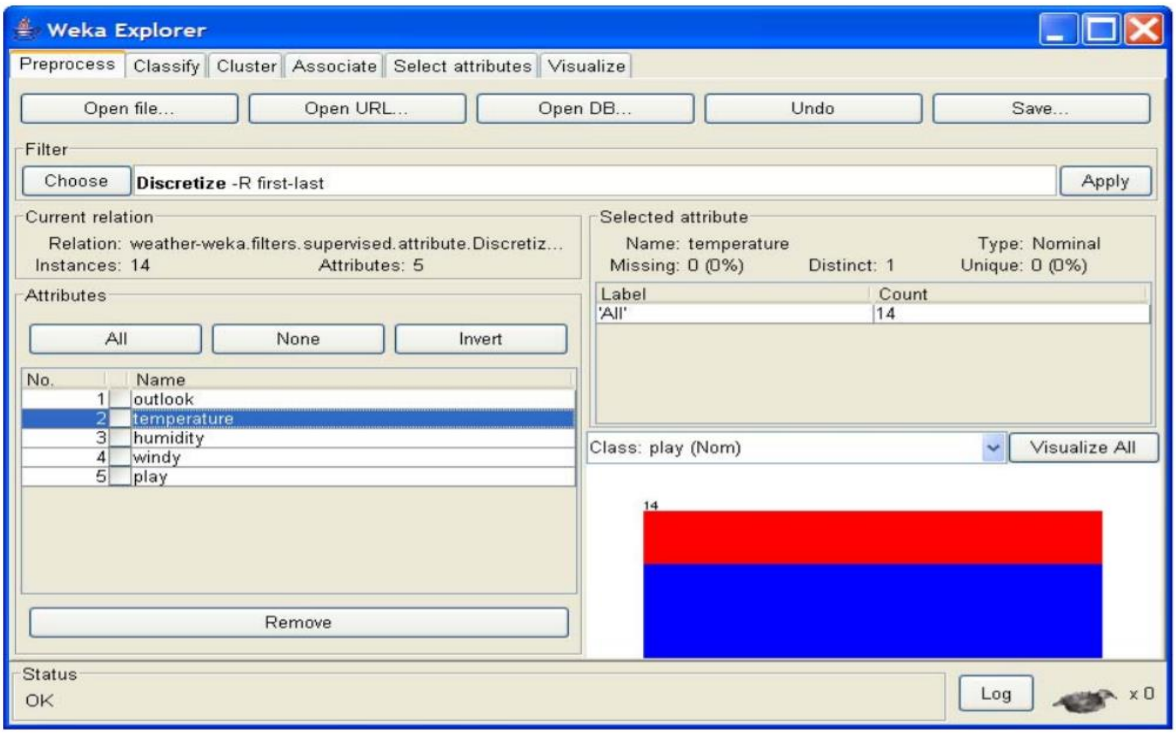
Preprocessing window:

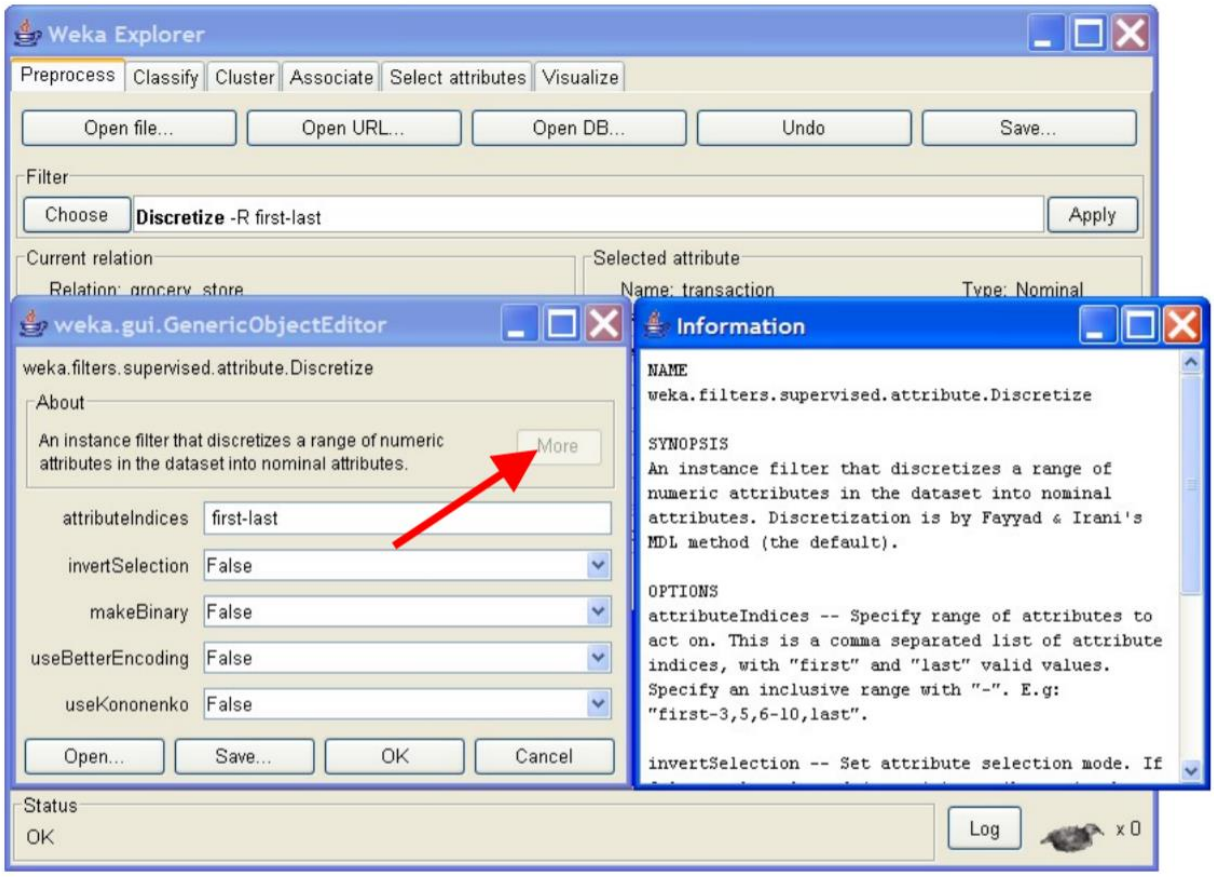




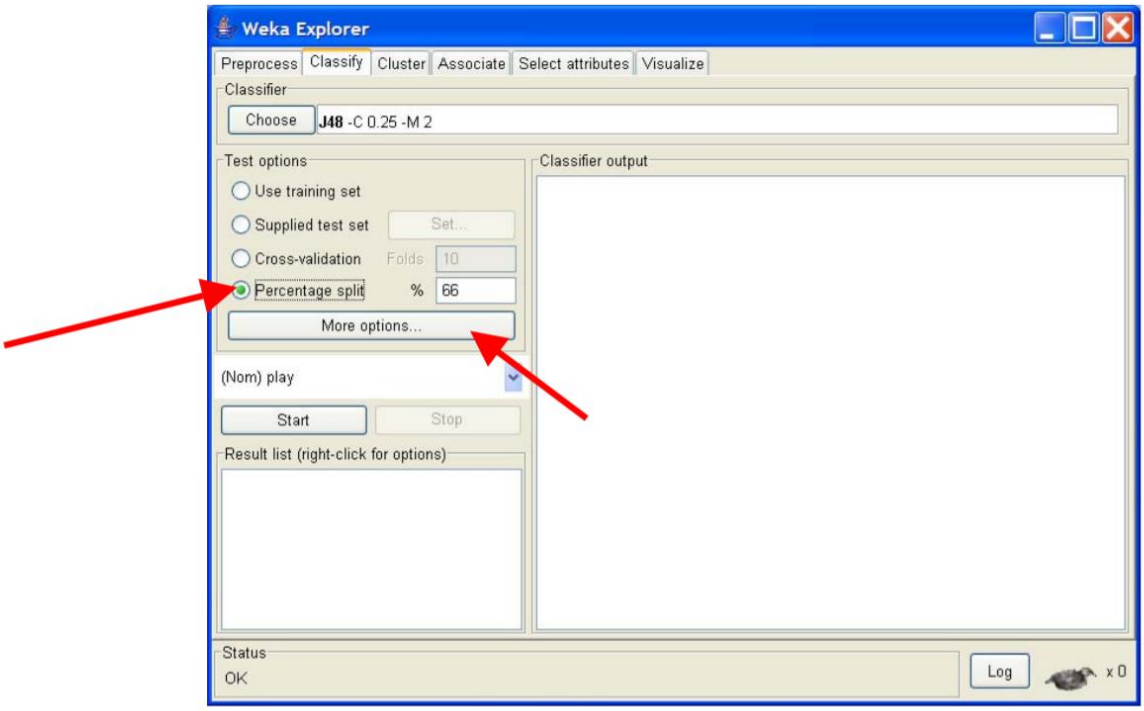
B) Filtering the data in weka:

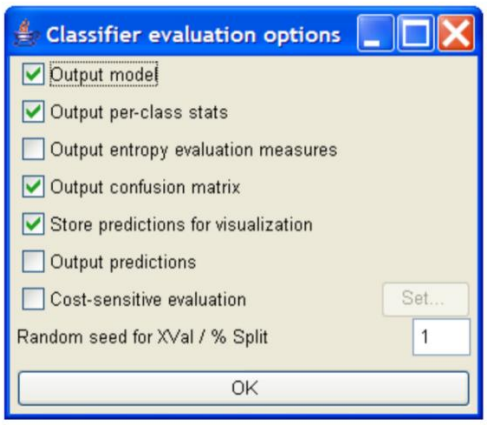


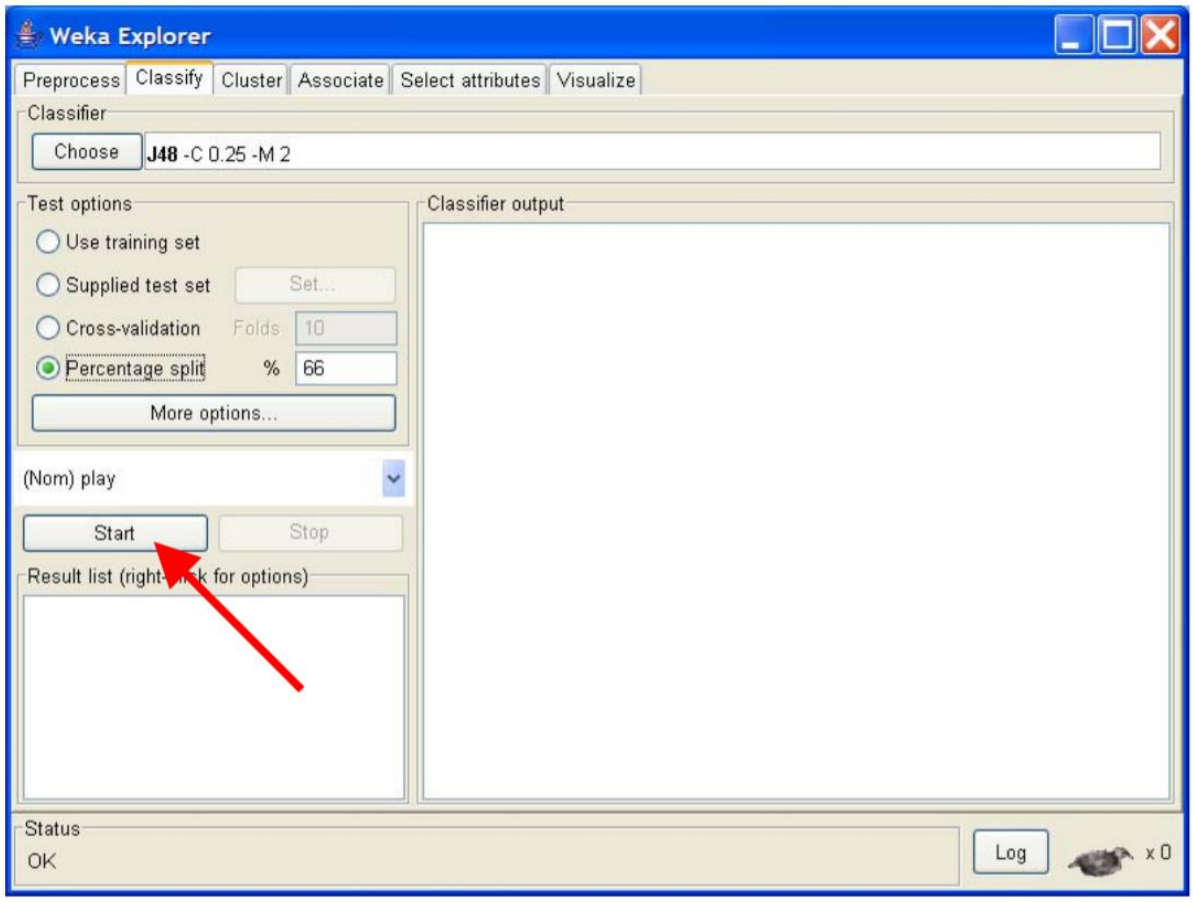

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -R first-last
Apply
Current relation
Relation: weather-weka.filters.supervised.attribute.Discretiz...
Instances: 14
Attributes: 5
Selected attribute
Name: temperature
Type: Nominal
Missing: 0 (0%)
Distinct: 1
Unique: 0 (0%)
Attributes
All
None
Invert
No.
Name
1 outlook
2 temperature
3 humidity
4 windy
5 play
Label
Count
'All'
14
Class: play (Nom)
Visualize All
14
Remove
Status
OK
Log
x 0

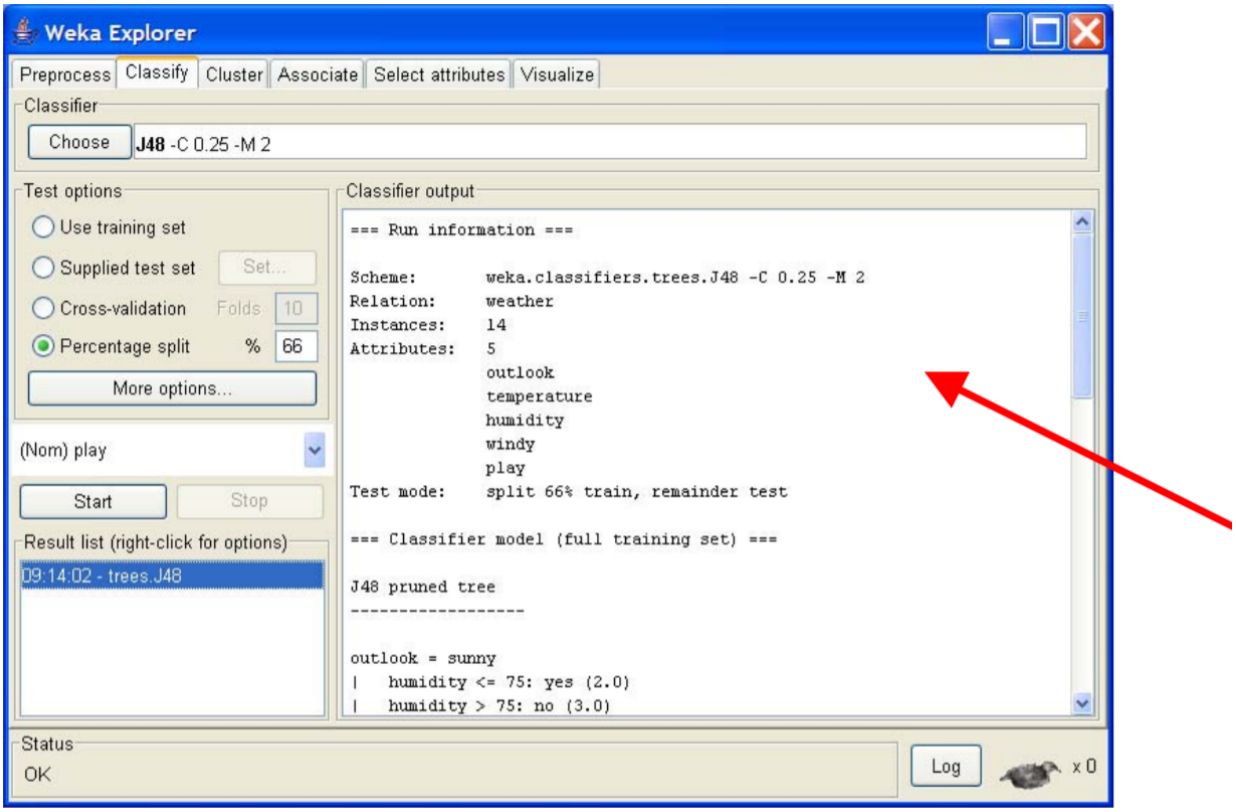

Weka Explorer
Preprocess
Classify
Cluster
Associate
Select attributes
Visualize
Open file...
Open URL...
Open DB...
Undo
Save...
Filter
Choose
Discretize -R first-last
Apply
Current relation
Selected attribute
Name: transaction
Type: Nominal
weka.gui.GenericObjectEditor
weka.filters.supervised.attribute.Discretize
About
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes.
More
attributeIndices
first-last
invertSelection
False
makeBinary
False
useBetterEncoding
False
useKononenko
False
Open...
Save...
OK
Cancel
Information
NAME
weka.filters.supervised.attribute.Discretize
SYNOPSIS
An instance filter that discretizes a range of numeric attributes in the dataset into nominal attributes. Discretization is by Fayyad & Irani's MDL method (the default).
OPTIONS
attributeIndices -- Specify range of attributes to act on. This is a comma separated list of attribute indices, with "first" and "last" valid values. Specify an inclusive range with "-". E.g: "first-3,5,6-10,last".
invertSelection -- Set attribute selection mode. If
Status
OK
Log
x 0

C) Training and testing of data:






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



Result: -

through this experiment we learned how to load data files in WEKA explorer and apply the training and testing schemas.