LAB MANUAL

DATA MINING AND WAREHOUSING

EXPERIMENT NO: 1

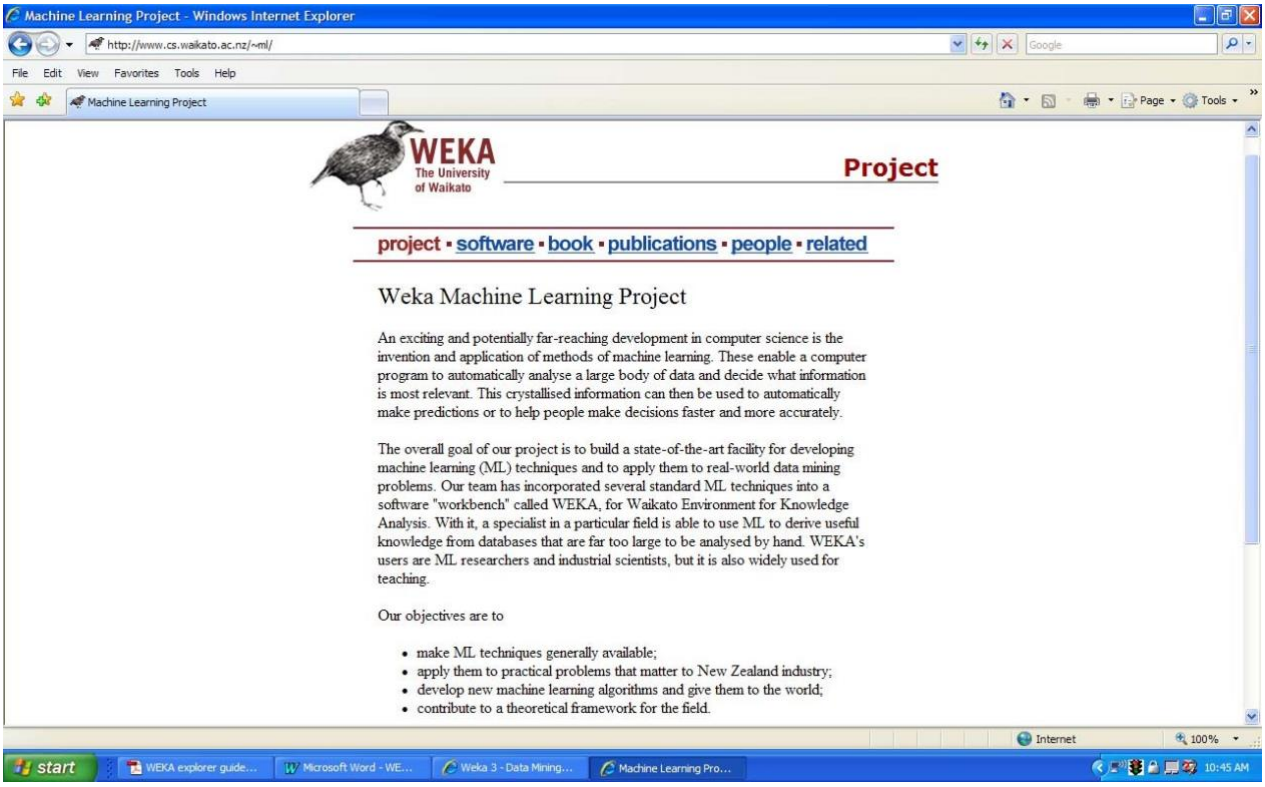
Aim : Introduction to WEKA tool.

Introduction

WEKA, formally called Waikato Environment for Knowledge Learning, is a computer program that was developed at the University of Waikato in New Zealand for the purpose of identifying information from raw data gathered from agricultural domains. WEKA supports many different standard data mining tasks such as data preprocessing, classification, clustering, regression, visualization and feature selection. The basic premise of the application is to utilize a computer application that can be trained to perform machine learning capabilities and derive useful information in the form of trends and patterns. WEKA is an open-source application that is freely available under the GNU general public license agreement. Originally written in C the WEKA application has been completely rewritten in Java and is compatible with almost every computing platform. It is user friendly with a graphical interface that allows for quick set up and operation. WEKA operates on the predication that the user data is available as a flat file or relation, this means that each data object is described by a fixed number of attributes that usually are of a specific type, normal alpha-numeric or numeric values. The WEKA application allows novice users a tool to identify hidden information from database and file systems with simple to use options and visual interfaces.

Installation

The program information can be found by conducting a search on the Web for WEKA Data Mining or going directly to the site at www.cs.waikato.ac.nz/~ml/WEKA . The site has a very large amount of useful information on the program's benefits and background. New users might find some benefit from investigating the user manual for the program. The main WEKA site has links to this information as well as past experiments for new users to refine the potential uses that might be of particular interest to them. When prepared to download the software it is best to select the latest application from the selection offered on the site. The format for downloading the application is offered in a self-installation package and is a simple procedure that provides the complete program on the end users machine that is ready to use when extracted.



Opening the program

Once the program has been loaded on the user's machine it is opened by navigating to the programs start option and that will depend on the user's operating system. Figure 1 is an example of the initial opening screen on a computer with Windows XP.

Figure 1 Chooser screen

There are four options available on this initial screen.

- ♦ Simple CLI- provides users without a graphic interface option the ability to execute commands from a terminal window.
- ♦ Explorer- the graphical interface used to conduct experimentation on raw data
- ♦ Experimenter- this option allows users to conduct different experimental variations on data sets and perform statistical manipulation
- ♦ Knowledge Flow-basically the same functionality as Explorer with drag and drop functionality. The advantage of this option is that it supports incremental learning from previous results

While the options available can be useful for different applications the remaining focus of the user manual will be on the Experimenter option through the rest of the user guide.

After selecting the Experimenter option the program starts and provides the user with a separate graphical interface.

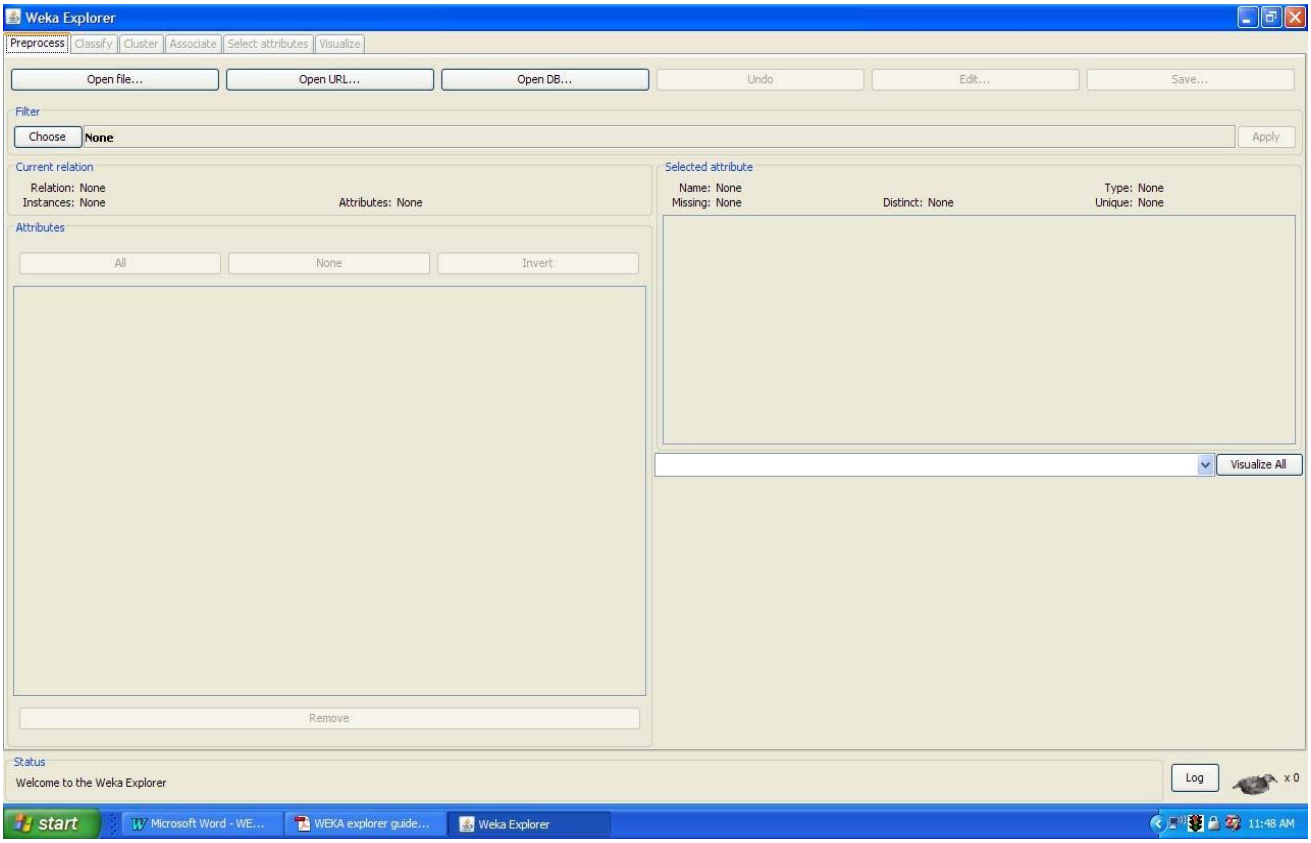


Figure 2

Figure 2 shows the opening screen with the available options. At first there is only the option to select the Preprocess tab in the top left corner. This is due to the necessity to present the data set to the application so it can be manipulated. After the data has been preprocessed the other tabs become active for use.

There are six tabs:

1. Preprocess- used to choose the data file to be used by the application
2. Classify- used to test and train different learning schemes on the preprocessed data file under experimentation
3. Cluster- used to apply different tools that identify clusters within the data file
4. Association- used to apply different rules to the data file that identify association within the data
5. Select attributes-used to apply different rules to reveal changes based on selected attributes inclusion or exclusion from the experiment
6. Visualize- used to see what the various manipulation produced on the data set in a 2D format, in scatter plot and bar graph output

Once the initial preprocessing of the data set has been completed the user can move between the tab options to perform changes to the experiment and view the results in real time. This provides the benefit of having the ability to move from one option to the next so that when a condition becomes exposed it can be placed in a different environment to be visually changed instantaneously.

Preprocessing

In order to experiment with the application the data set needs to be presented to WEKA in a format that the

program understands. There are rules for the type of data that WEKA will accept. There are three options for presenting data into the program.

- ♦ Open File- allows for the user to select files residing on the local machine or recorded medium
- ♦ Open URL- provides a mechanism to locate a file or data source from a different location specified by the user
- ♦ Open Database- allows the user to retrieve files or data from a database source provided by the user

There are restrictions on the type of data that can be accepted into the program. Originally the software was designed to import only ARFF files, newer versions allow different file types such as CSV, C4.5 and serialized instance formats. The extensions for these files include .csv, .arff, .names, .bsi and .data. Figure 3 shows an example of selection of the file weather.arff.

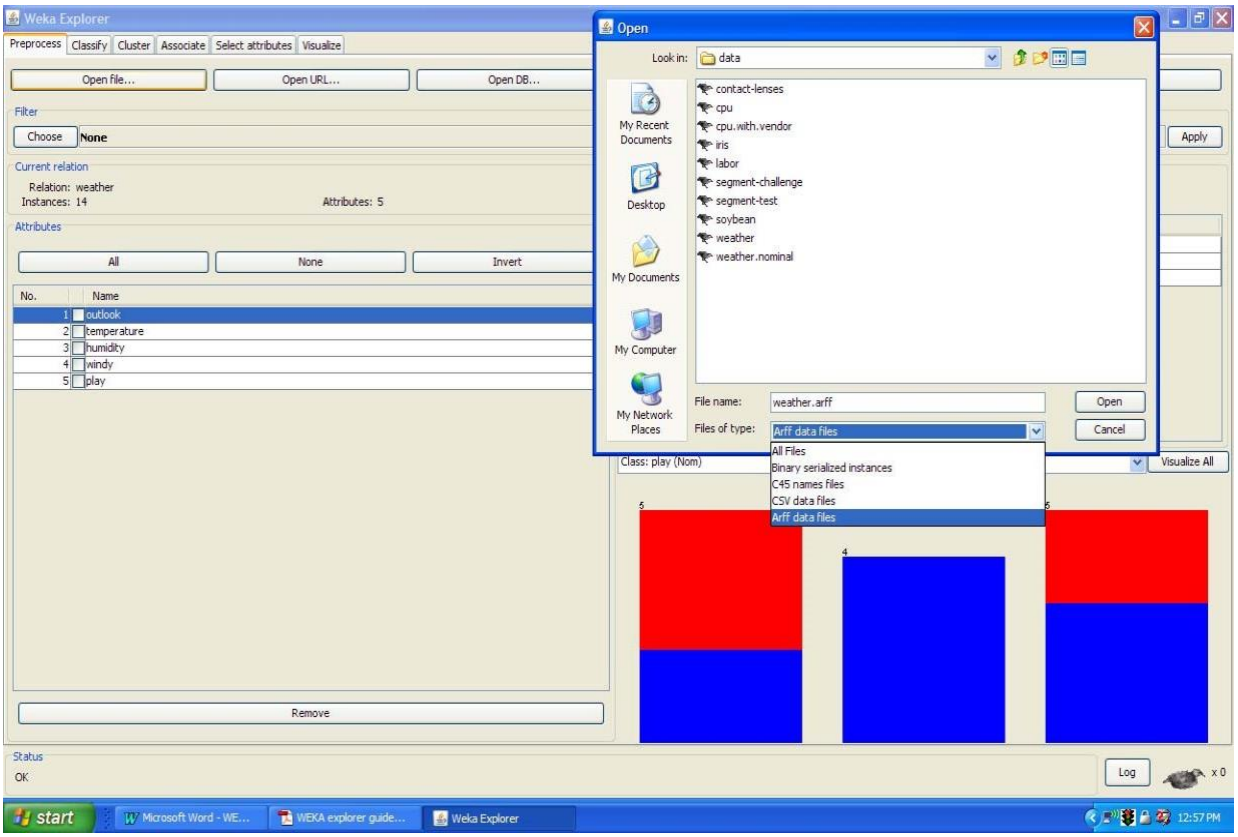


Figure 3

Once the initial data has been selected and loaded the user can select options for refining the experimental data. The options in the preprocess window include selection of optional filters to apply and the user can select or remove different attributes of the data set as necessary to identify specific information. The ability to pick from the available attributes allows users to separate different parts of the data set for clarity in the experimentation. The user can modify the attribute selection and change the relationship among the different attributes by deselecting different choices from the original data set. There are many different filtering options available within the preprocessing window and the user can select the different options based on need and type of data present.

Classify

The user has the option of applying many different algorithms to the data set that would in theory produce a representation of the information used to make observation easier. It is difficult to identify which of the options would provide the best output for the experiment. The best approach is to independently apply a mixture of the available choices and see what yields something close to the desired results. The Classify tab is where the user selects the classifier choices. Figure 4 shows some of the categories.

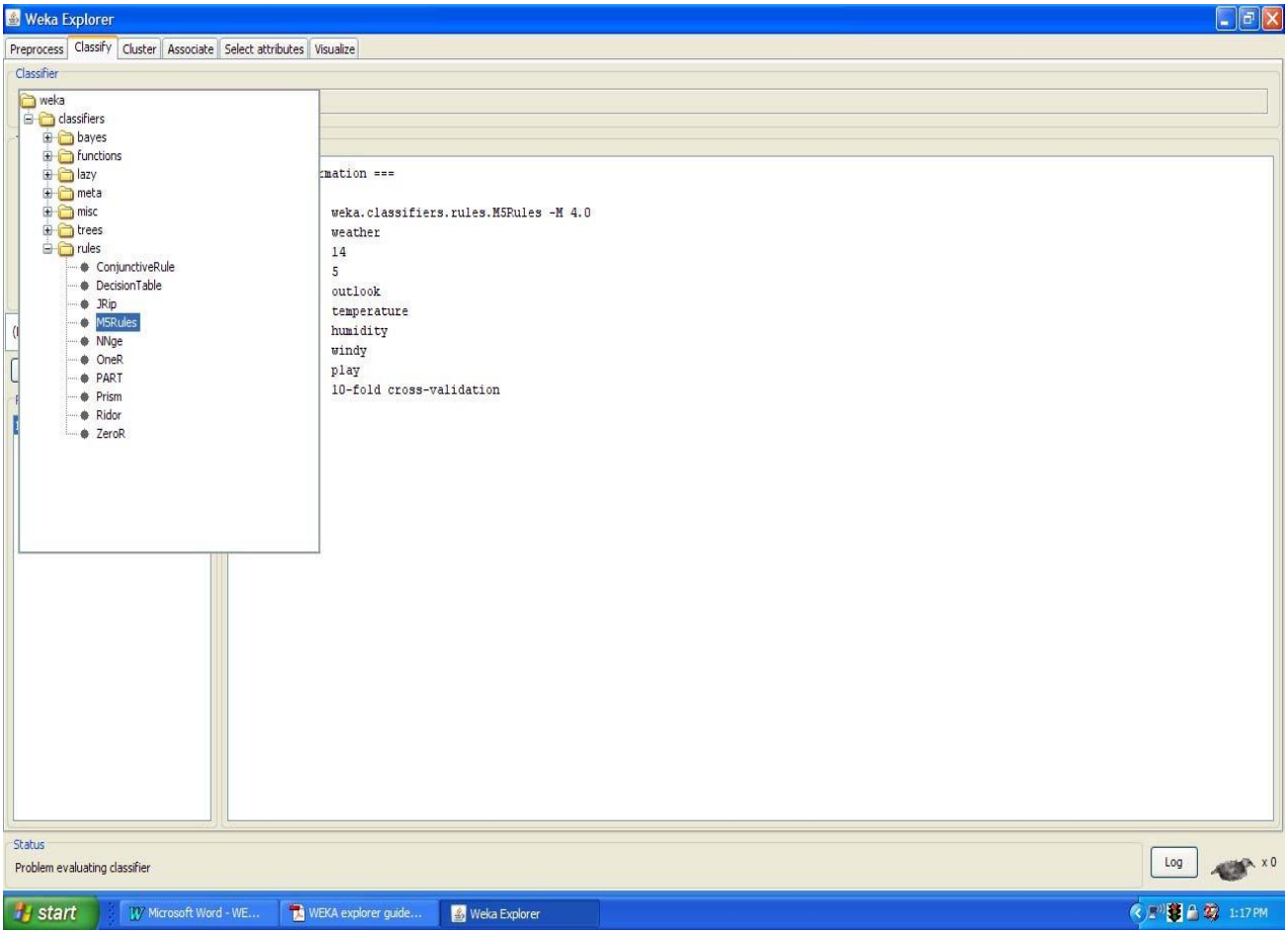


Figure 4

Again there are several options to be selected inside of the classify tab. Test option gives the user the choice of using four different test mode scenarios on the data set:

1. Use training set
2. Supplied training set
3. Cross validation
4. Split percentage

There is the option of applying any or all of the modes to produce results that can be compared by the user. Additionally inside the test options toolbox there is a dropdown menu so the user can select various items to apply that depending on the choice can provide output options such as saving the results to file or specifying the random seed value to be applied for the classification.

The classifiers in WEKA have been developed to train the data set to produce output that has been classified based on the characteristics of the last attribute in the data set. For a specific attribute to be used the option must be selected by the user in the options menu before testing is performed. Finally the results have been calculated and they are shown in the text box on the lower right. They can be saved in a file and later retrieved for comparison at a later time or viewed within the window after changes and different results have been derived.

Cluster

The Cluster tab opens the process that is used to identify commonalties or clusters of occurrences within the data set and produce information for the user to analyze. There are a few options within the cluster window that are similar to those described in the classifier tab. They are use training set, supplied test set, percentage split. The fourth option is classes to cluster evaluation, which compares how well the data compares with a pre-assigned class within the data. While in cluster mode users have the option of ignoring some of the attributes from the data set. This can be useful if there are specific attributes causing the results to be out of range or for large data sets. Figure 5 shows the Cluster window and some of its options.

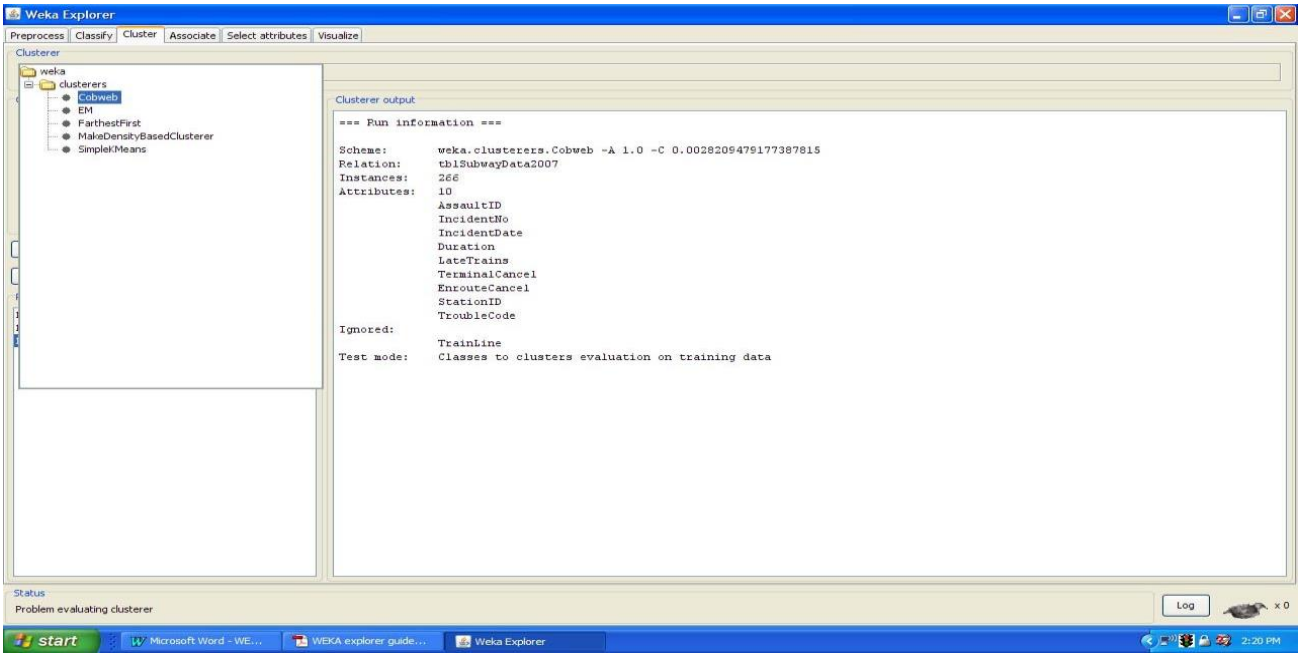


Figure 5

Associate

The associate tab opens a window to select the options for associations within the data set. The user selects one of the choices and presses start to yield the results. There are few options for this window and they are shown in Figure 6 below.

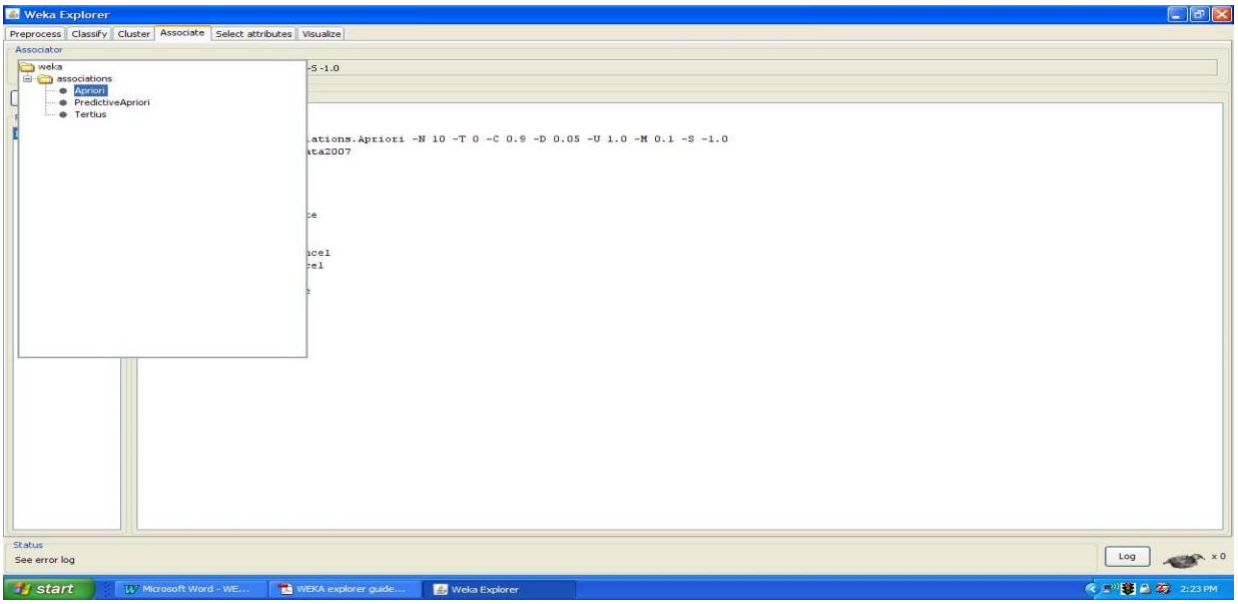


Figure 6

Select Attributes

The next tab is used to select the specific attributes used for the calculation process. By default all of the available attributes are used in the evaluation of the data set. If the use wanted to exclude certain categories of the data they would deselect those specific choices from the list in the cluster window. This is useful if some of the attributes are of a different form such as alphanumeric data that could alter the results. The software searches through the selected attributes to decide which of them will best fit the desired calculation. To perform this, the user has to select two options, an attribute evaluator and a search method. Once this is done the program evaluates the data based on the sub set of the attributes then

performs the necessary search for commonality with the date. Figure 7 shows the opinions of attribute evaluation. Figure 8 shows the options for the search method.

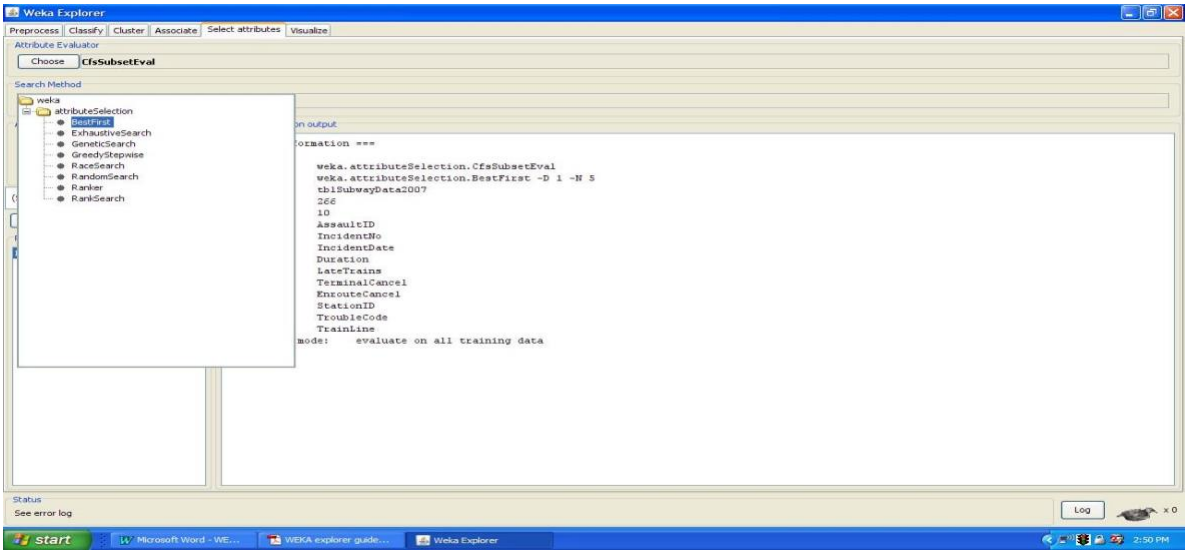


Figure 7

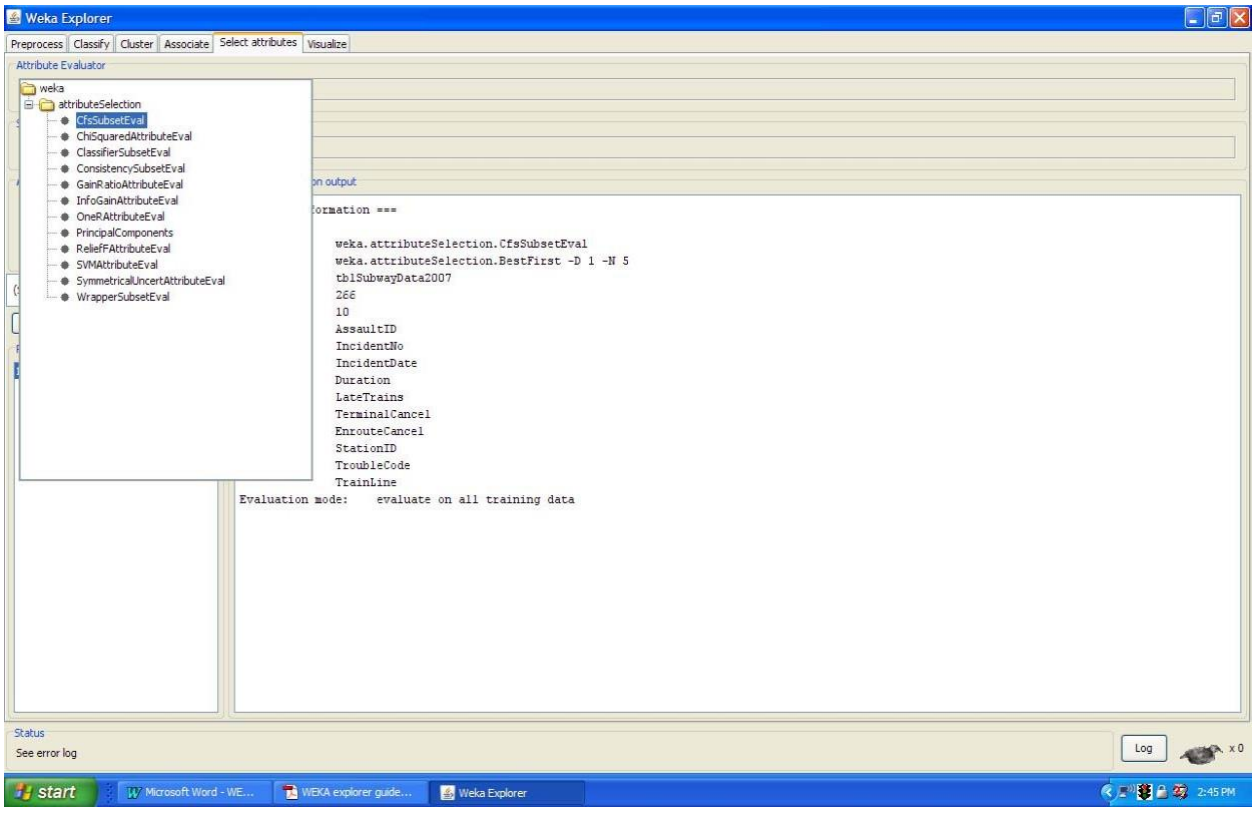


Figure 8

Visualization

The last tab in the window is the visualization tab. Within the program calculations and comparisons have occurred on the data set. Selections of attributes and methods of manipulation have been chosen. The final piece of the puzzle is looking at the information that has been derived throughout the process. The user can now actually see the fruit of their efforts in a two-dimensional representation of the information. The first screen that the user sees when they select the visualization option is a matrix of plots representing the different attributes within the data set plotted against the other attributes. If necessary, there is a scroll bar to view all of the produced

plots. The user can select a specific plot from the matrix to view its contents for analyzation. A grid pattern of the plots allows the user to select the attribute positioning to their liking and for better understanding. Once a specific plot has been selected the user can change the attributes from one view to another providing flexibility. Figure 9 shows the plot matrix view.

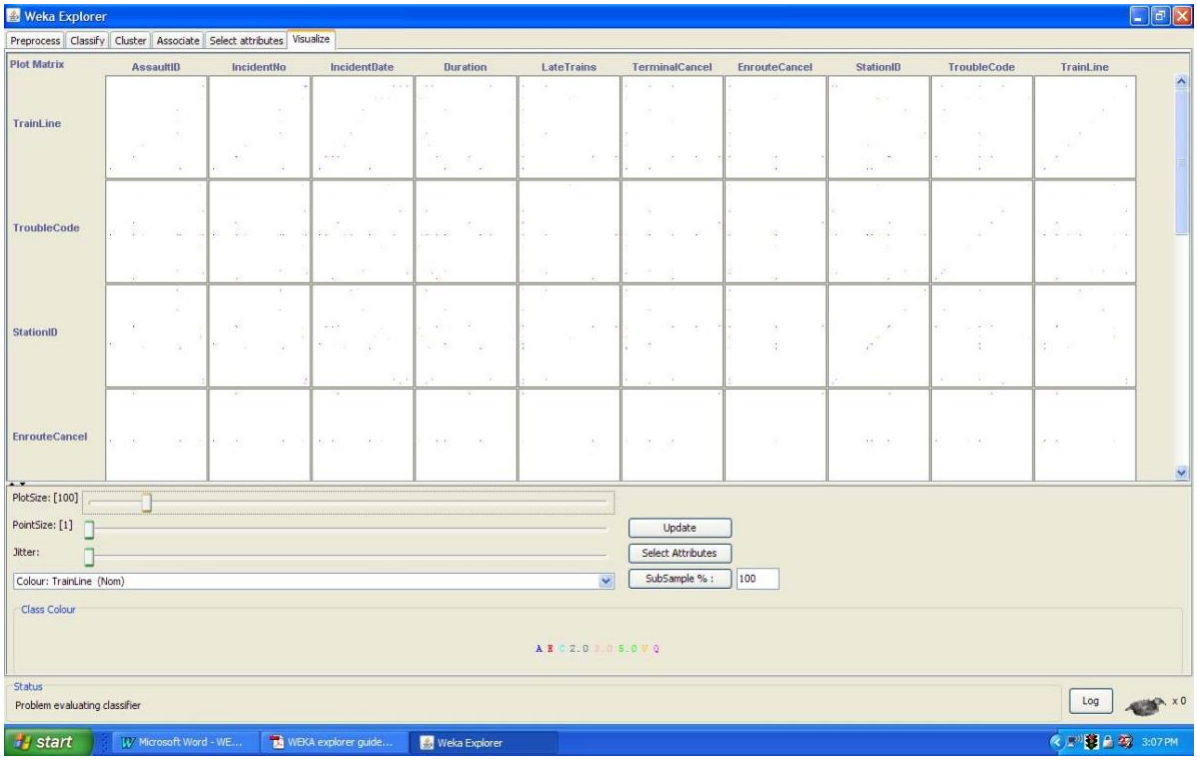


Figure 9

The scatter plot matrix gives the user a visual representation of the manipulated data sets for selection and analysis. The choices are the attributes across the top and the same from top to bottom giving the user easy access to pick the area of interest. Clicking on a plot brings up a separate window of the selected scatter plot. The user can then look at a visualization of the data of the attributes selected and select areas of the scatter plot with a selection window or by clicking on the points within the plot to identify the point's specific information. Figure 10 shows the scatter plot for two attributes and the points derived from the data set. There are a few options to view the plot that could be helpful to the user. It is formatted similar to an X/Y graph yet it can show any of the attribute classes that appear on the main scatter plot matrix. This is handy when the scale of the attribute is unable to be ascertained in one axis over the other. Within the plot the points can be adjusted by utilizing a feature called jitter. This option moves the individual points so that in the event of close data points users can reveal hidden multiple occurrences within the initial plot. Figure 11 shows an example of this point selection and the results the user sees.

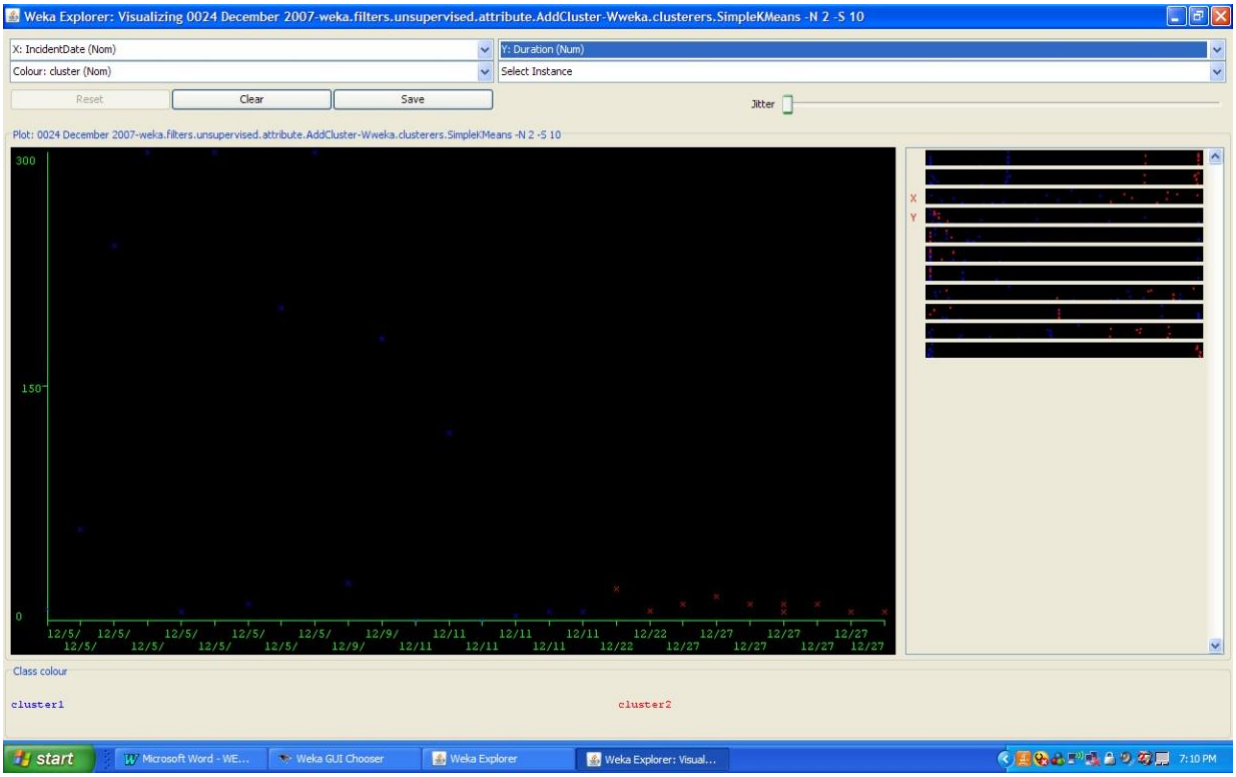


Figure 10

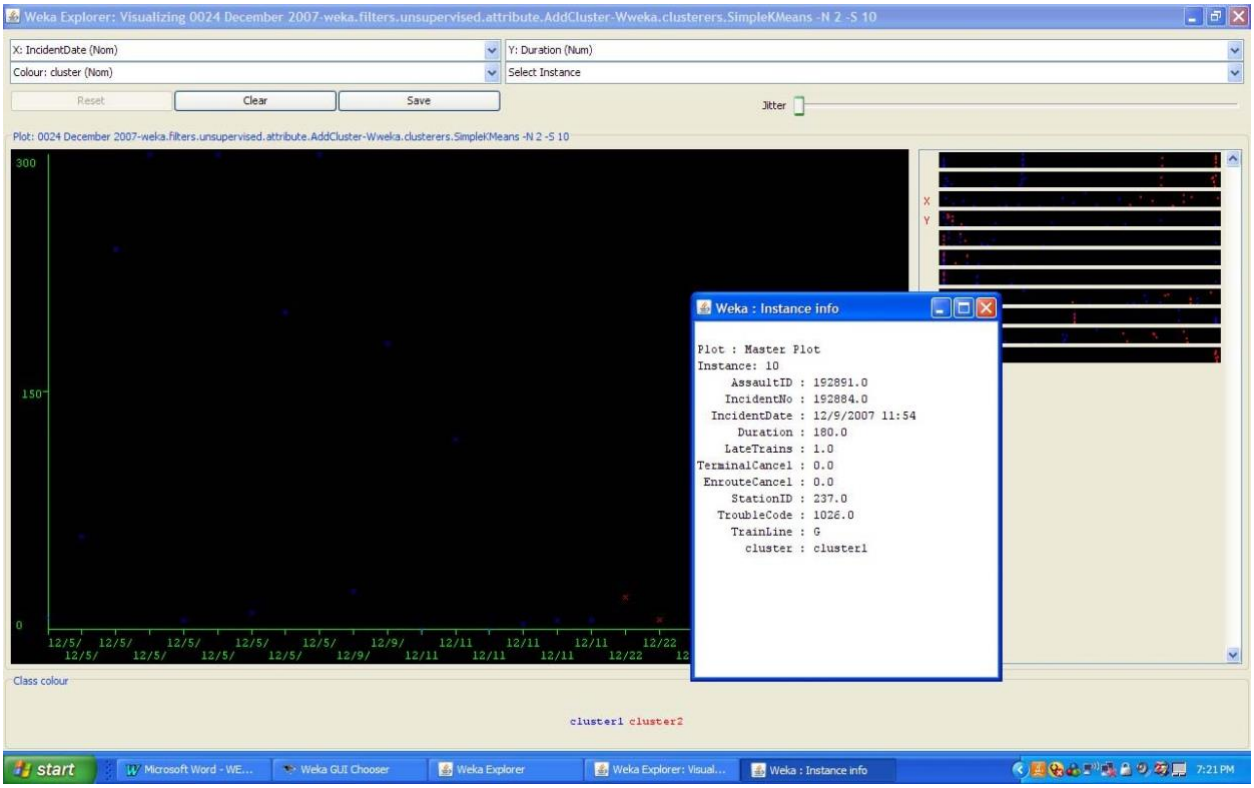


Figure 11

There are a few options to manipulate the view for the identification of subsets or to separate the data points on the plot.

- ♦ Polyline---can be used to segment different values for additional visualization clarity on the plot. This is useful when there are many data points represented on the graph.
- ♦ Rectangle--this tool is helpful to select instances within the graph for copying or clarification.
- ♦ Polygon—Users can connect points to segregate information and isolate points for reference.

This user guide is meant to assist users in their efforts to become familiar with some of the features within the Explorer portion of the WEKA data mining software application and is used for informational purposes only. It is a summary of the user information found on the programs main web site. For a more comprehensive and in depth version users can visit the main site http://www.cs.waikato.ac.nz/~ml/WEKA for examples and FAQ's about the program.

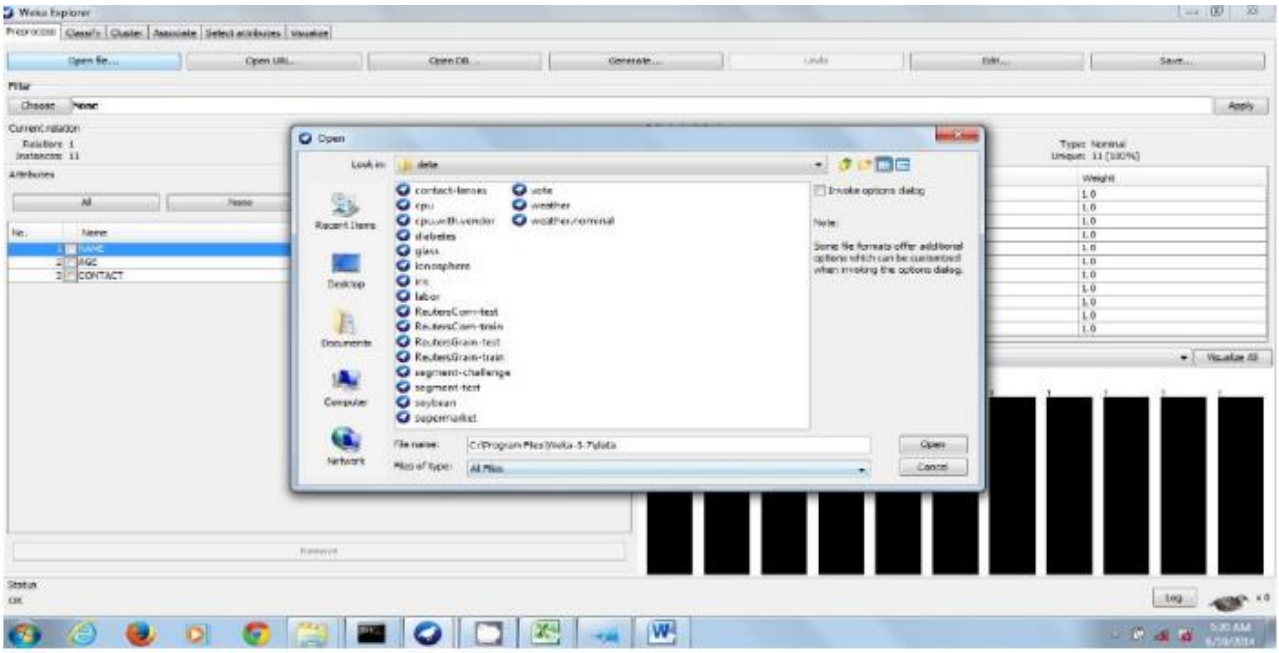
Experiment-2

Aim: To Create Data in .arff format

Procedure: Open MS Excel.

Steps

Create a new worksheet with respective headings and data.
Save the file with .csv extension
Open Preprocessor tab in explorer
Click open the file button and browse the file to open.
Load the desired .csv file using open File tab.



Click SAVE shown… dialog box opens save with extension as .arff

EXPERIMENT NO: 3

Aim:

Create an Employee Table with the help of Data Mining Tool WEKA.

Description:

We need to create an Employee Table with training data set which includes attributes like name, id, salary, experience, gender, phone number.

Procedure:

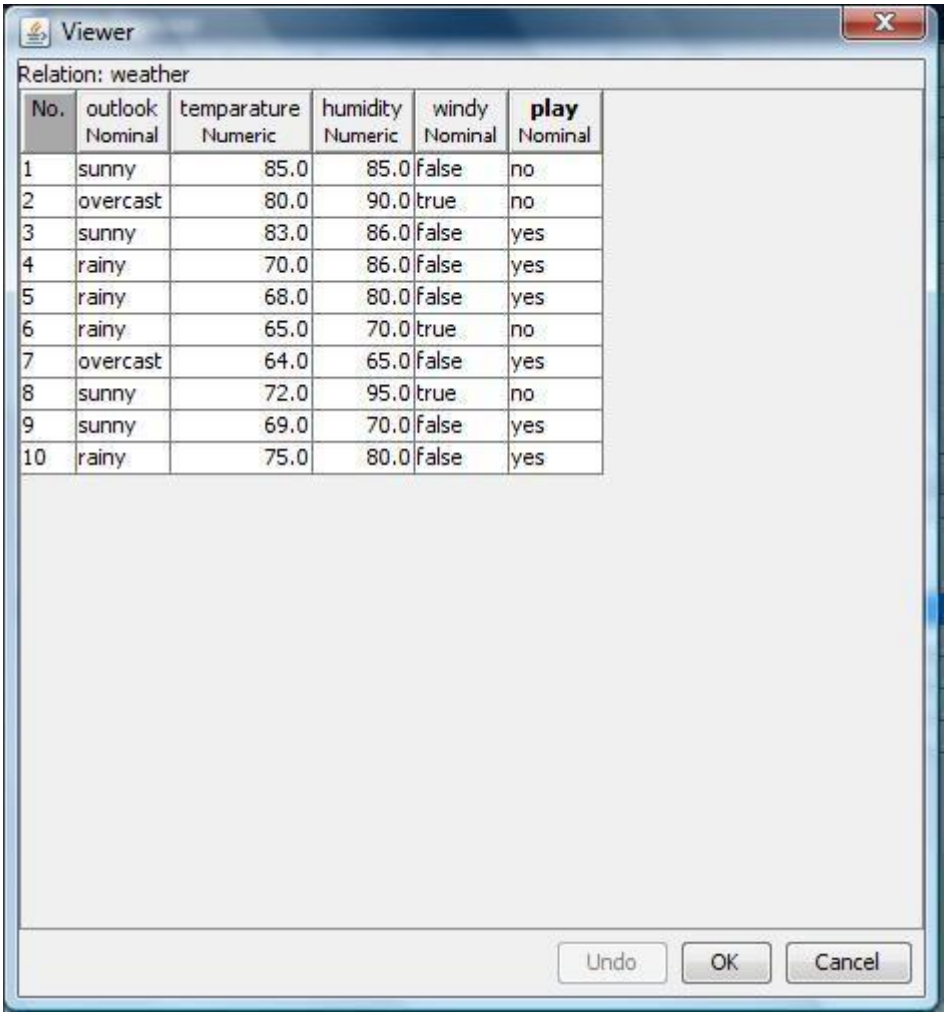
Steps:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Employee Table.

```
@relation employee
@attribute name {x,y,z,a,b}
@attribute id numeric
@attribute salary {low,medium,high}
@attribute exp numeric
@attribute gender {male,female}
@attribute phone numeric

@data
x,101,low,2,male,250311
y,102,high,3,female,251665
z,103,medium,1,male,240238
a,104,low,5,female,200200
b,105,high,2,male,240240
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows employee table on weka.

Training Data Set → Weather Table



No.	outlook Nominal	temparature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	false	no
2	overcast	80.0	90.0	true	no
3	sunny	83.0	86.0	false	yes
4	rainy	70.0	86.0	false	yes
5	rainy	68.0	80.0	false	yes
6	rainy	65.0	70.0	true	no
7	overcast	64.0	65.0	false	yes
8	sunny	72.0	95.0	true	no
9	sunny	69.0	70.0	false	yes
10	rainy	75.0	80.0	false	yes

Result:

This program has been successfully executed.

EXPERIMENT NO:4

Aim:

Create a Weather Table with the help of Data Mining Tool WEKA.

Description:

We need to create a Weather table with training data set which includes attributes like outlook, temperature, humidity, windy, play.

Procedure:

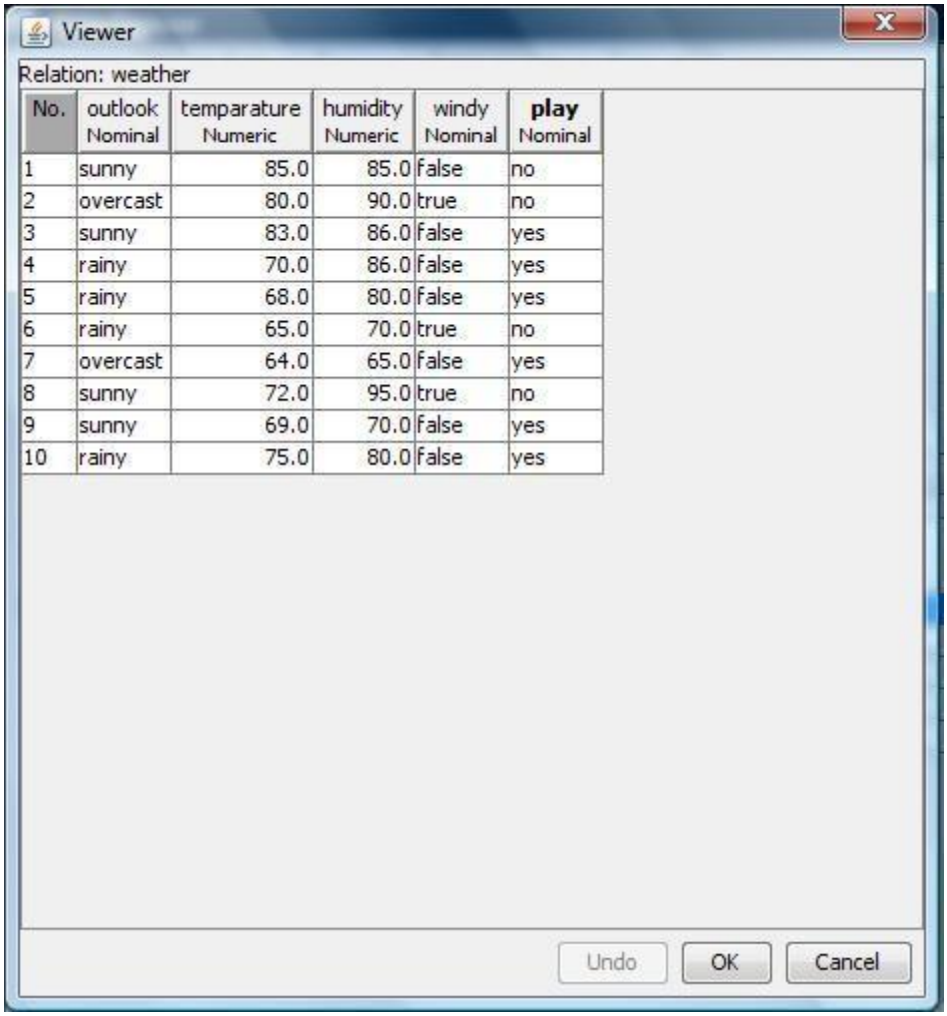
Steps:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Weather Table.

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no}

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows weather table on weka.

Training Data Set → Weather Table



Result:

This program has been successfully executed.

EXPERIMENT NO:5

Aim:

Apply Pre-Processing techniques to the training data set of Weather Table

Description:

Real world databases are highly influenced to noise, missing and inconsistency due to their queue size so the data can be pre-processed to improve the quality of data and missing results and it also improves the efficiency.

There are 3 pre-processing techniques they are:

1) Add
2) Remove
3) Normalization

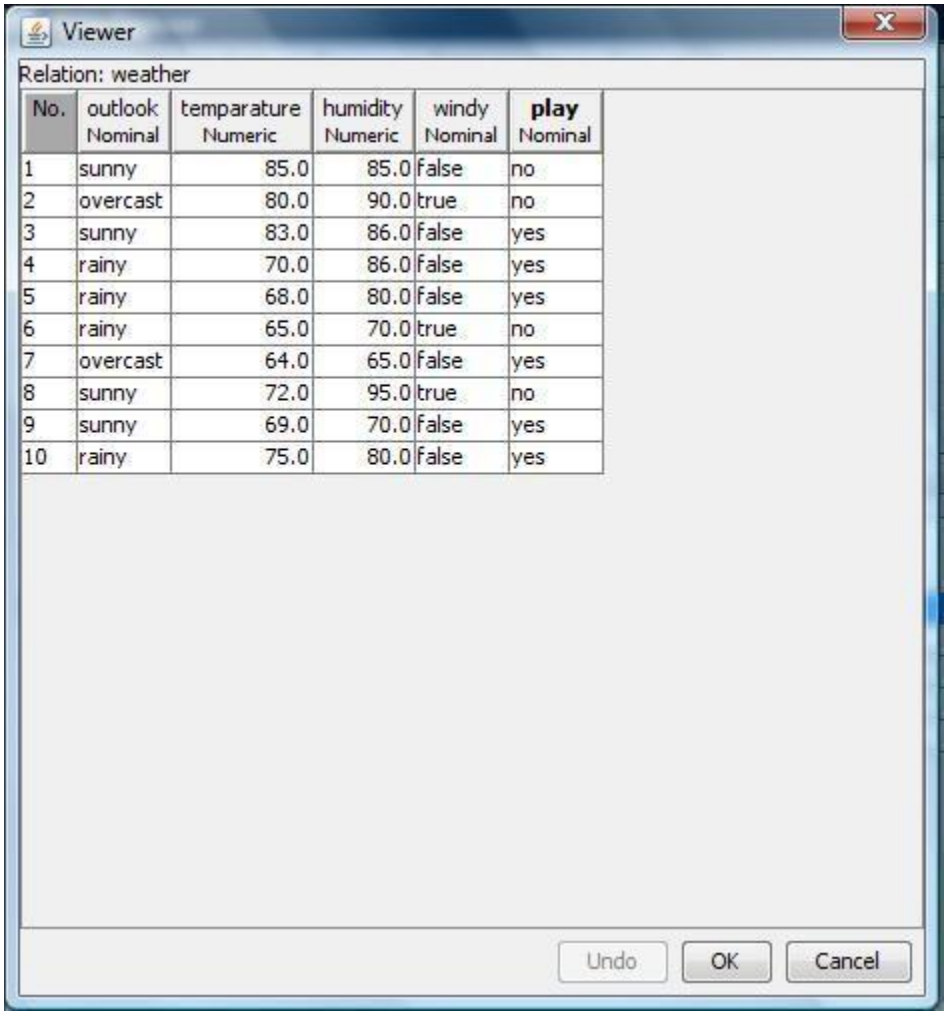
Creation of Weather Table:

Procedure:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Weather Table.

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no}

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows weather table on weka.

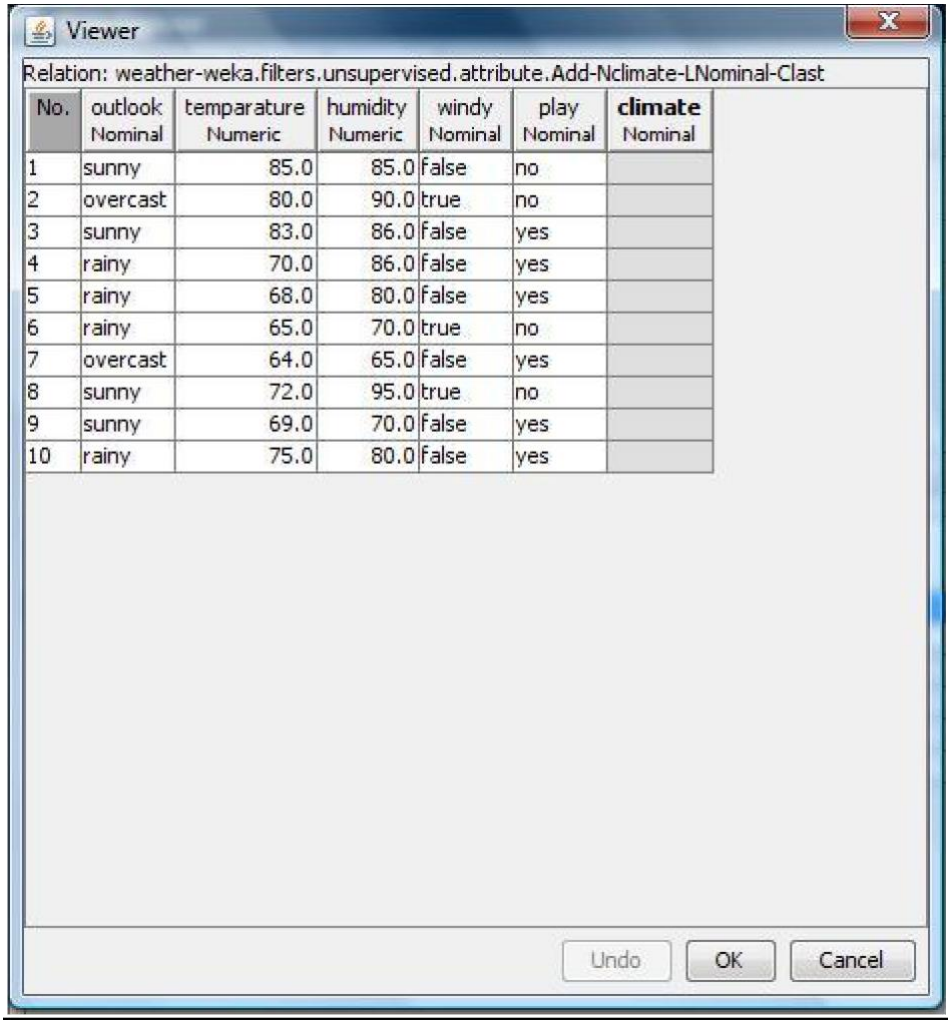


Add → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Weather.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Add**.
9) A new window is opened.
10) In that we enter attribute index, type, data format, nominal label values for **Climate**.
11) Click on **OK**.
12) Press the **Apply button**, then a new attribute is added to the Weather Table.
13) **Save** the file.
14) Click on the **Edit button**, it shows a new Weather Table on Weka.

Weather Table after adding new attribute CLIMATE:

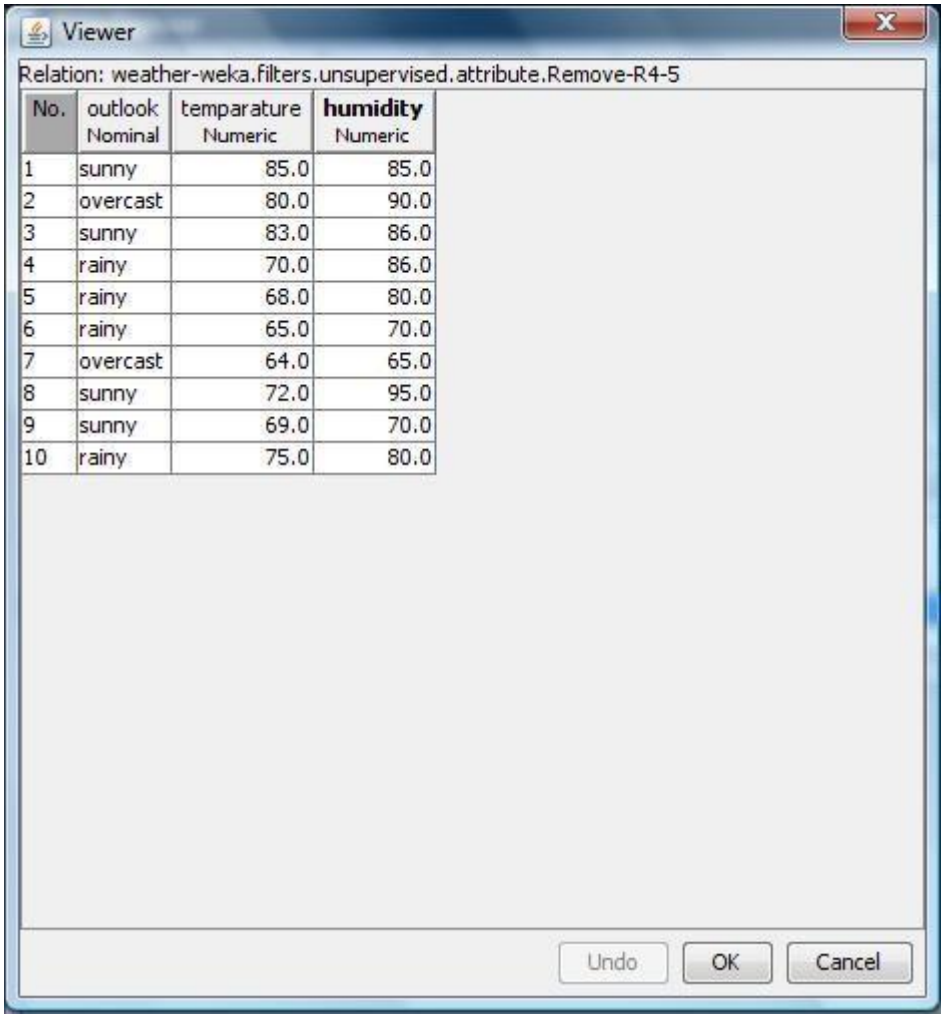


Remove → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Weather.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Remove**.
9) Select the attributes **windy, play** to Remove.
10) Click **Remove button** and then **Save**.
11) Click on the **Edit button**, it shows a new Weather Table on Weka.

Weather Table after removing attributes WINDY, PLAY:

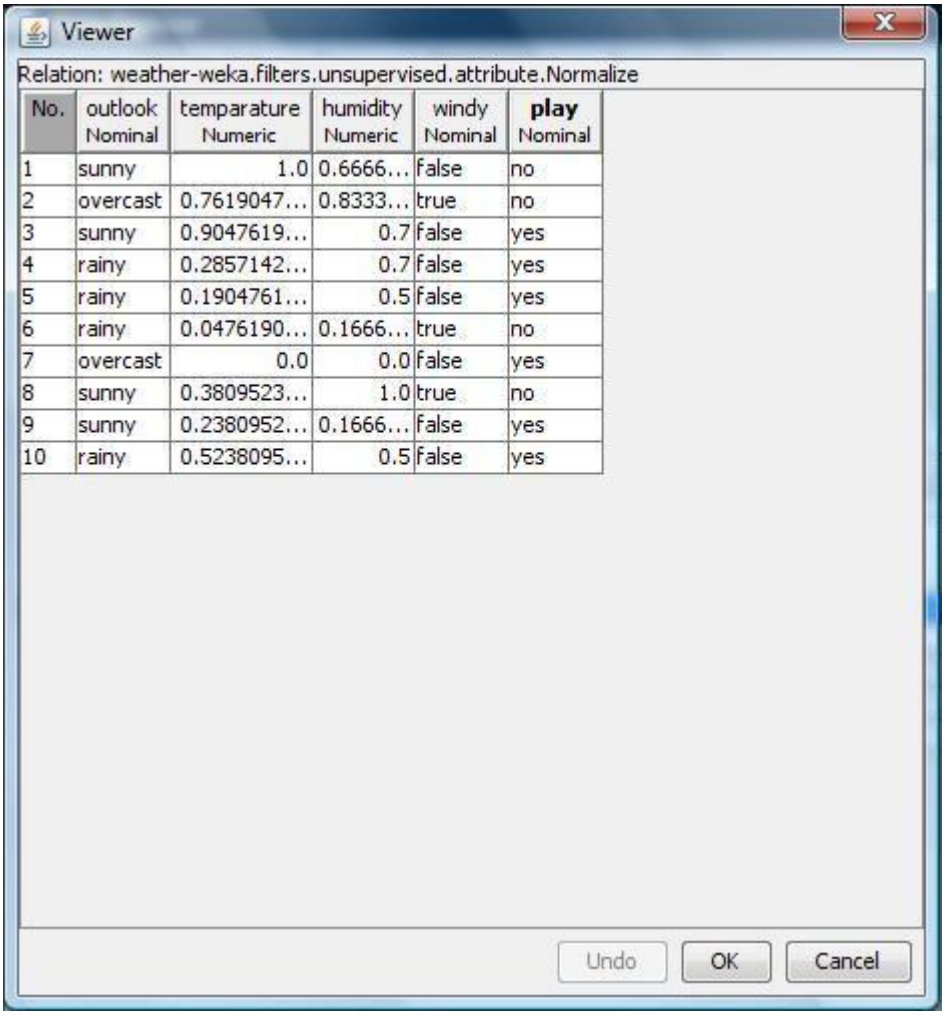


Normalize → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Weather.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Normalize**.
9) Select the attributes **temparature, humidity** to Normalize.
10) Click on **Apply button** and then **Save**.
11) Click on the **Edit button**, it shows a new Weather Table with normalized values on Weka.

Weather Table after Normalizing TEMPARATURE, HUMIDITY:



Result:

This program has been successfully executed.

EXPERIMENT NO:6

Aim:

Apply Pre-Processing techniques to the training data set of Employee Table

Description:

Real world databases are highly influenced to noise, missing and inconsistency due to their queue size so the data can be pre-processed to improve the quality of data and missing results and it also improves the efficiency.

There are 3 pre-processing techniques they are:

1) Add
2) Remove
3) Normalization

Creation of Employee Table:

Procedure:

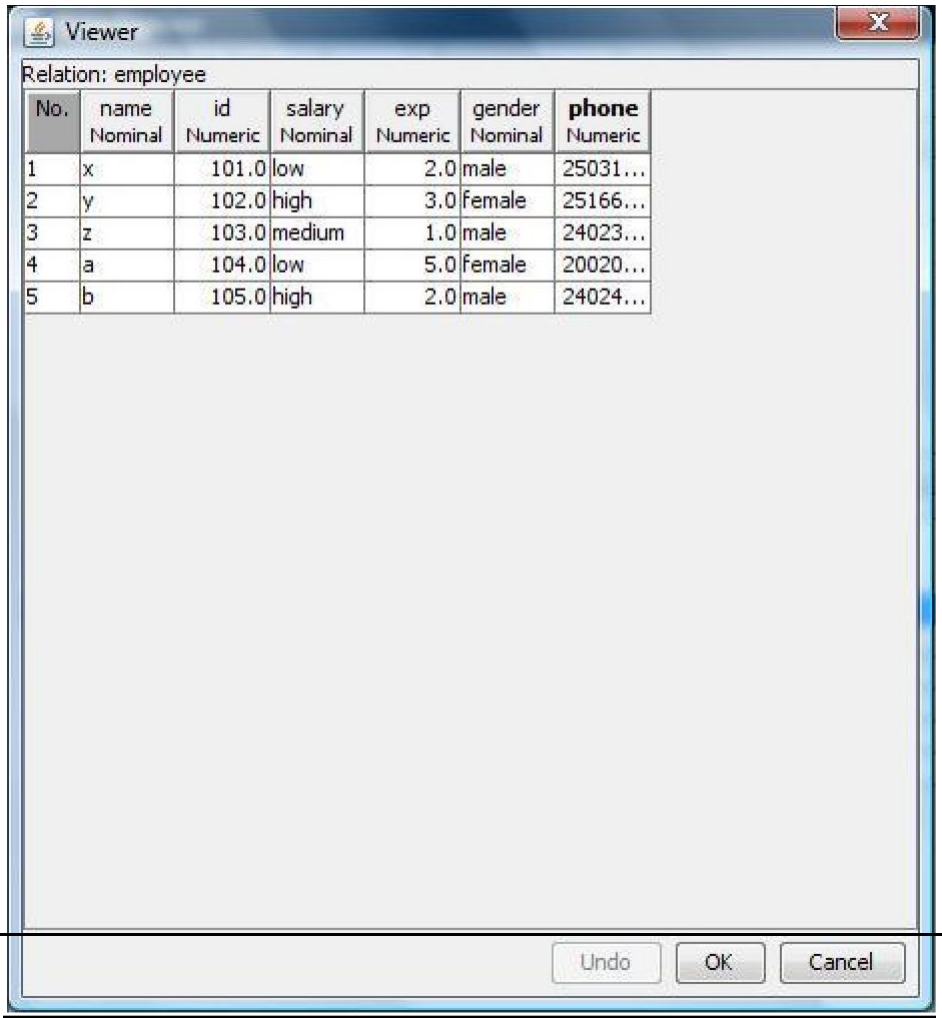
1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Employee Table.

```
@relation employee
@attribute name {x,y,z,a,b}
@attribute id numeric
@attribute salary {low,medium,high}
@attribute exp numeric
@attribute gender {male,female}
@attribute phone numeric

@data
x,101,low,2,male,250311
y,102,high,3,female,251665
z,103,medium,1,male,240238
a,104,low,5,female,200200
b,105,high,2,male,240240
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.

7) Explorer shows many options. In that click on **'open file'** and select the arff file

8) Click on **edit button** which shows employee table on weka.



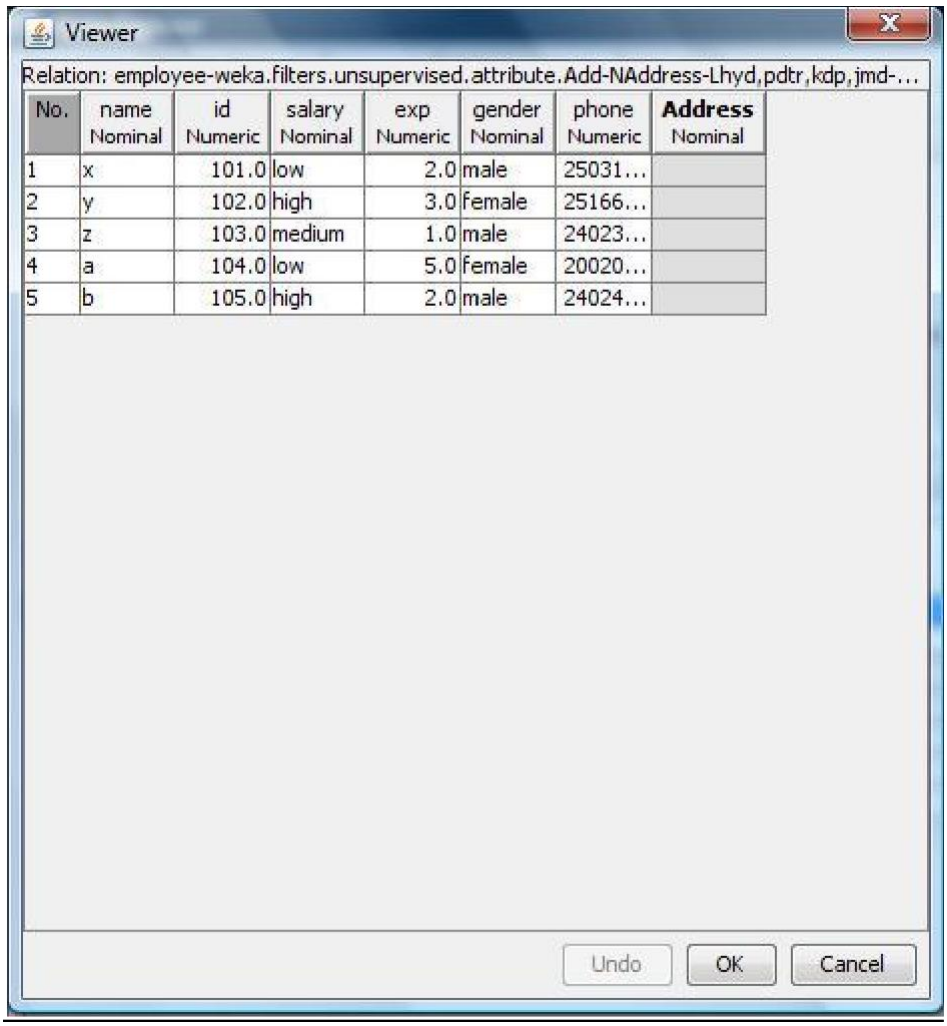
Training Data Set → Employee Table

Add → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Employee.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Add**.
9) A new window is opened.
10) In that we enter attribute index, type, data format, nominal label values for **Address**.
11) Click on **OK**.
12) Press the **Apply button**, then a new attribute is added to the Employee Table.
13) **Save** the file.
14) Click on the **Edit button**, it shows a new Employee Table on Weka.

Employee Table after adding new attribute ADDRESS:

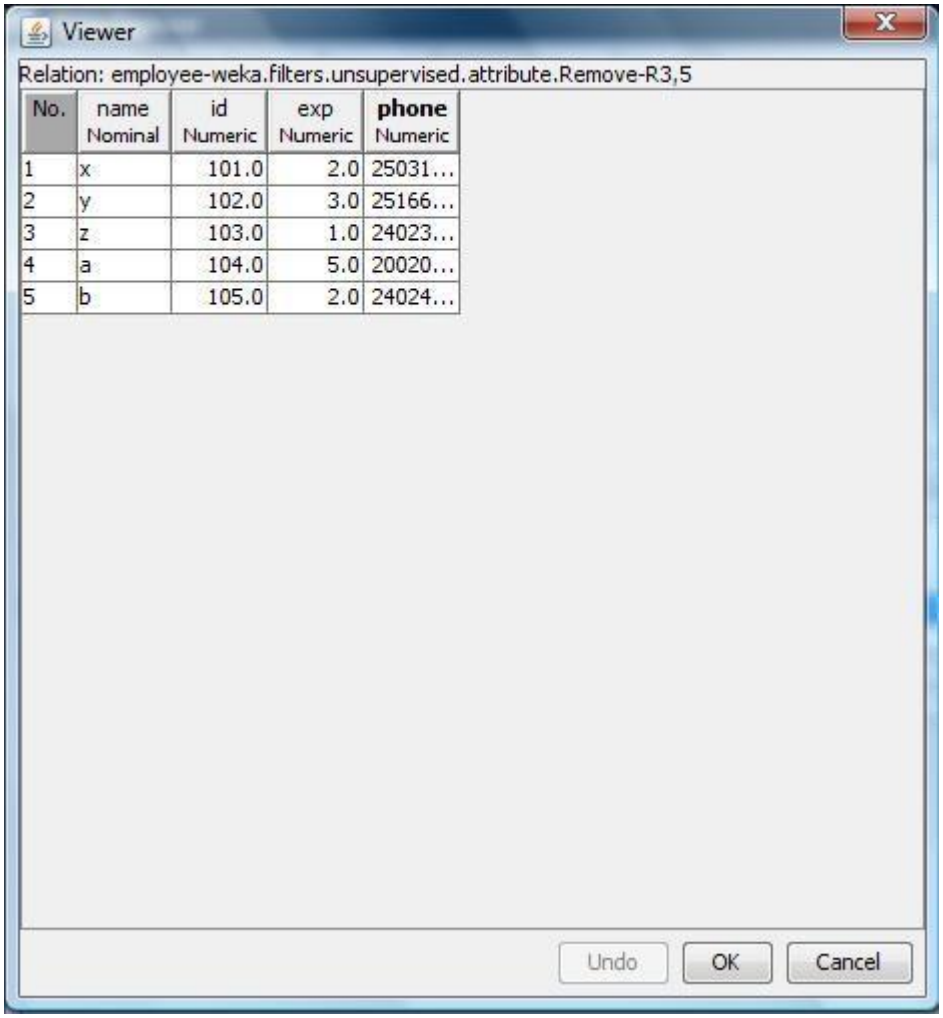


Remove → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Employee.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Remove**.
9) Select the attributes **salary, gender** to Remove.
10) Click **Remove button** and then **Save**.
11) Click on the **Edit button**, it shows a new Employee Table on Weka.

Employee Table after removing attributes SALARY, GENDER:

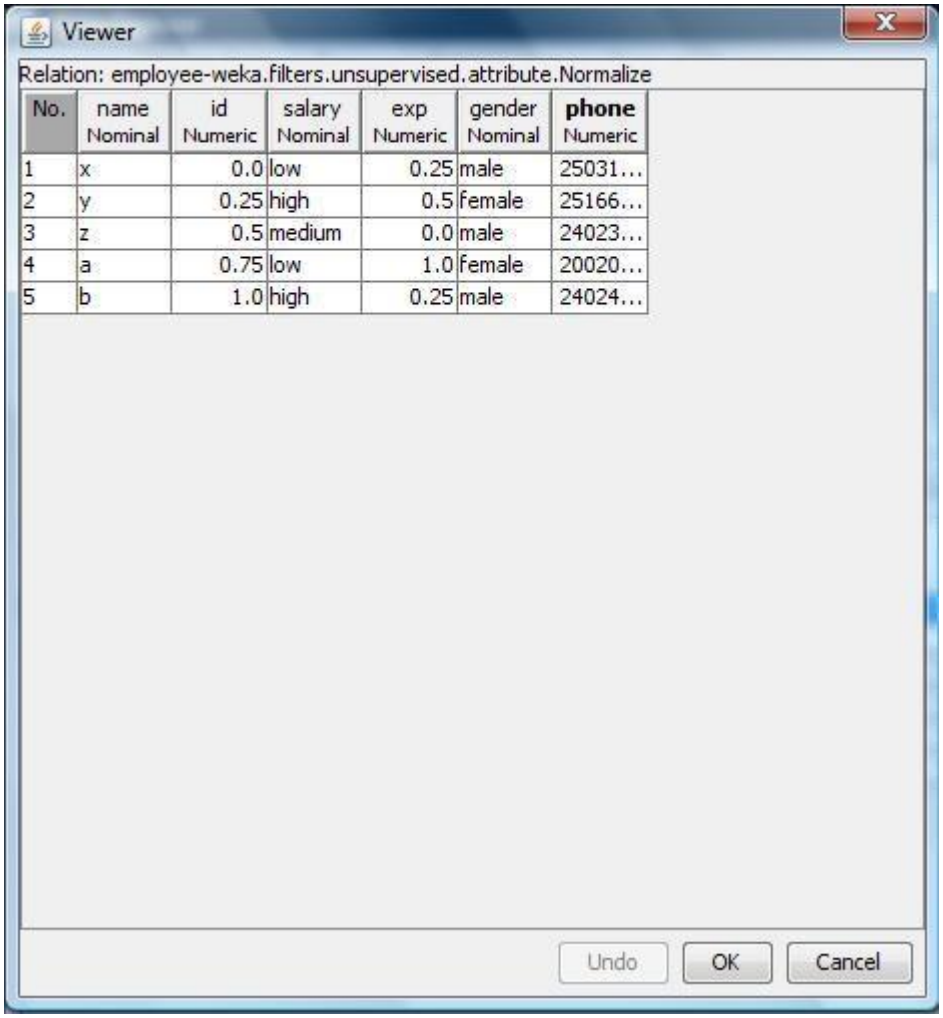


Normalize → Pre-Processing Technique:

Procedure:

1) Start → Programs → Weka-3-4 → Weka-3-4
2) Click on **explorer.**
3) Click on **open file.**
4) Select **Employee.arff** file and click on open.
5) Click on **Choose button** and select the **Filters option**.
6) In Filters, we have **Supervised** and **Unsupervised data**.
7) Click on **Unsupervised data**.
8) Select the attribute **Normalize**.
9) Select the attributes **id, experience, phone** to Normalize.
10) Click on **Apply button** and then **Save**.
11) Click on the **Edit button**, it shows a new Employee Table with normalized values on Weka.

Employee Table after Normalizing ID, EXP, PHONE:



Result:

This program has been successfully executed.

EXPERIMENT NO:7

Aim:

Normalize Weather Table data using Knowledge Flow.

Description:

The knowledge flow provides an alternative way to the explorer as a graphical front end to WEKA's algorithm. Knowledge flow is a working progress. So, some of the functionality from explorer is not yet available. So, on the other hand there are the things that can be done in knowledge flow, but not in explorer. Knowledge flow presents a dataflow interface to WEKA. The user can select WEKA components from a toolbar placed them on a layout campus and connect them together in order to form a knowledge flow for processing and analyzing the data.

Creation of Weather Table:

Procedure:

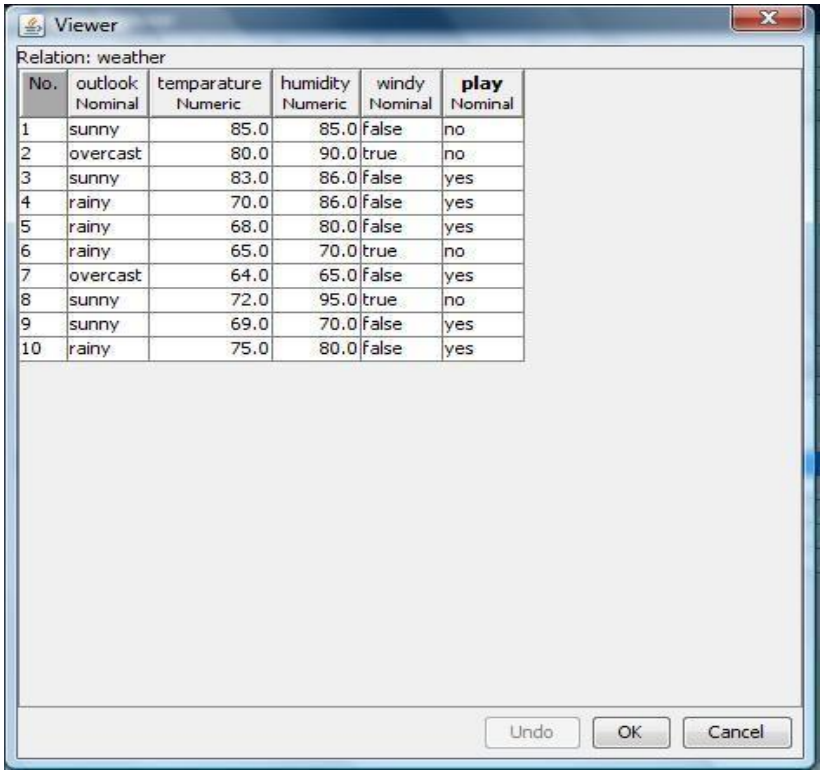
1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Weather Table.

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temparature numeric
@attribute humidity numeric
@attribute windy {true,false}
@attribute play {yes,no}

@data
sunny,85.0,85.0,false,no
overcast,80.0,90.0,true,no
sunny,83.0,86.0,false,yes
rainy,70.0,86.0,false,yes
rainy,68.0,80.0,false,yes
rainy,65.0,70.0,true,no
overcast,64.0,65.0,false,yes
sunny,72.0,95.0,true,no
sunny,69.0,70.0,false,yes
rainy,75.0,80.0,false,yes
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows Weather table on weka.

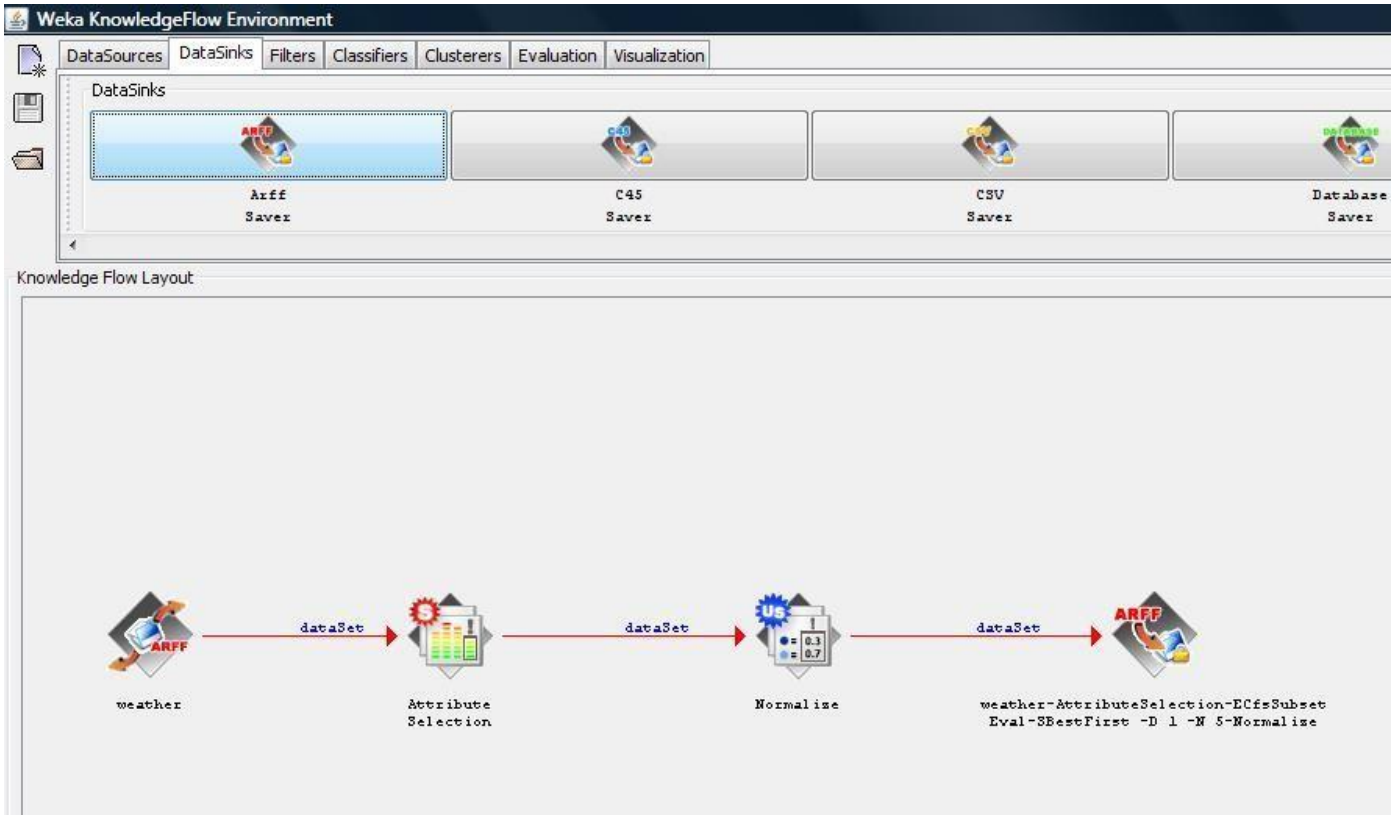
Output:

Training Data Set → Weather Table



Procedure for Knowledge Flow:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open the **Knowledge Flow**.
3) Select the **Data Source component** and **add Arff Loader** into the **knowledge layout canvas**.
4) Select the **Filters component** and **add Attribute Selection** and N**ormalize** into the knowledge layout canvas.
5) Select the **Data Sinks** component and **add Arff Saver** into the knowledge layout canvas.
6) Right click on **Arff Loader** and select **Configure option** then the new window will be opened and select **Weather.arff**
7) Right click on **Arff Loader** and select **Dataset option** then establish a link between **Arff Loader** and A**ttribute Selection**.
8) Right click on **Attribute Selection** and select **Dataset option** then establish a link between **Attribute Selection** and **Normalize**.
9) Right click on **Attribute Selection** and select **Configure option** and choose the best attribute for Weather data.
10) Right click on **Normalize** and select **Dataset option** then establish a link between **Normalize** and **Arff Saver**.
11) Right click on **Arff Saver** and select **Configure option** then new window will be opened and set the path, enter **.arff** in look in dialog box to save normalize data.
12) Right click on **Arff Loader** and click on **Start Loading option** then everything will be executed one by one.
13) Check whether output is created or not by selecting the preferred path.
14) Rename the data name as **a.arff**
15) Double click on **a.arff** then automatically the output will be opened in **MS-Excel**.



Result:

This program has been successfully executed.

EXPERIMENT NO:8

Aim:

Normalize Employee Table data using Knowledge Flow.

Description:

The knowledge flow provides an alternative way to the explorer as a graphical front end to WEKA's algorithm. Knowledge flow is a working progress. So, some of the functionality from explorer is not yet available. So, on the other hand there are the things that can be done in knowledge flow, but not in explorer. Knowledge flow presents a dataflow interface to WEKA. The user can select WEKA components from a toolbar placed them on a layout campus and connect them together in order to form a knowledge flow for processing and analyzing the data.

Creation of Employee Table:

Procedure:

1) Open Start → Programs → Accessories → Notepad

2) Type the following training data set with the help of Notepad for Employee Table.

```
@relation employee
@attribute eid numeric
@attribute ename {raj,ramu,anil,sunil,rajiv,sunitha,kavitha,suresh,ravi,ramana,ram,kavya,navya}
@attribute salary numeric
@attribute exp numeric
@attribute address {pdtr,kdp,nlr,gtr}
@data
101,raj,10000,4,pdtr
102,ramu,15000,5,pdtr
103,anil,12000,3,kdp
104,sunil,13000,3,kdp
105,rajiv,16000,6,kdp
106,sunitha,15000,5,nlr
107,kavitha,12000,3,nlr
108,suresh,11000,5,gtr
109,ravi,12000,3,gtr
110,ramana,11000,5,gtr
111,ram,12000,3,kdp
112,kavya,13000,4,kdp
113,navya,14000,5,kdp
```

3) After that the file is saved with **.arff** file format.

4) Minimize the arff file and then open Start → Programs → weka-3-4.

5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

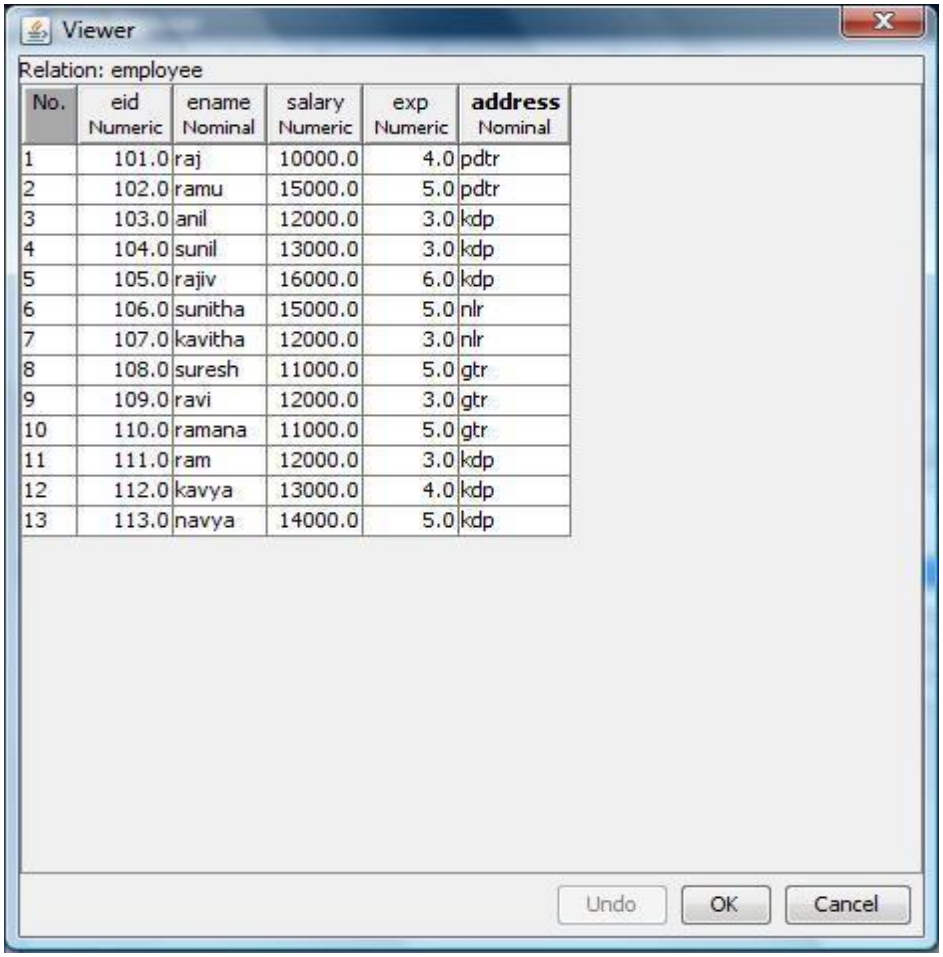
6) In that dialog box there are four modes, click on **explorer**.

7) Explorer shows many options. In that click on **'open file'** and select the arff file

8) Click on **edit button** which shows employee table on weka.

Output:

Training Data Set → Employee Table



No.	eid Numeric	ename Nominal	salary Numeric	exp Numeric	address Nominal
1	101.0	raj	10000.0	4.0	pdtr
2	102.0	ramu	15000.0	5.0	pdtr
3	103.0	anil	12000.0	3.0	kdp
4	104.0	sunil	13000.0	3.0	kdp
5	105.0	rajiv	16000.0	6.0	kdp
6	106.0	sunitha	15000.0	5.0	nlr
7	107.0	kavitha	12000.0	3.0	nlr
8	108.0	suresh	11000.0	5.0	gtr
9	109.0	ravi	12000.0	3.0	gtr
10	110.0	ramana	11000.0	5.0	gtr
11	111.0	ram	12000.0	3.0	kdp
12	112.0	kavya	13000.0	4.0	kdp
13	113.0	navya	14000.0	5.0	kdp

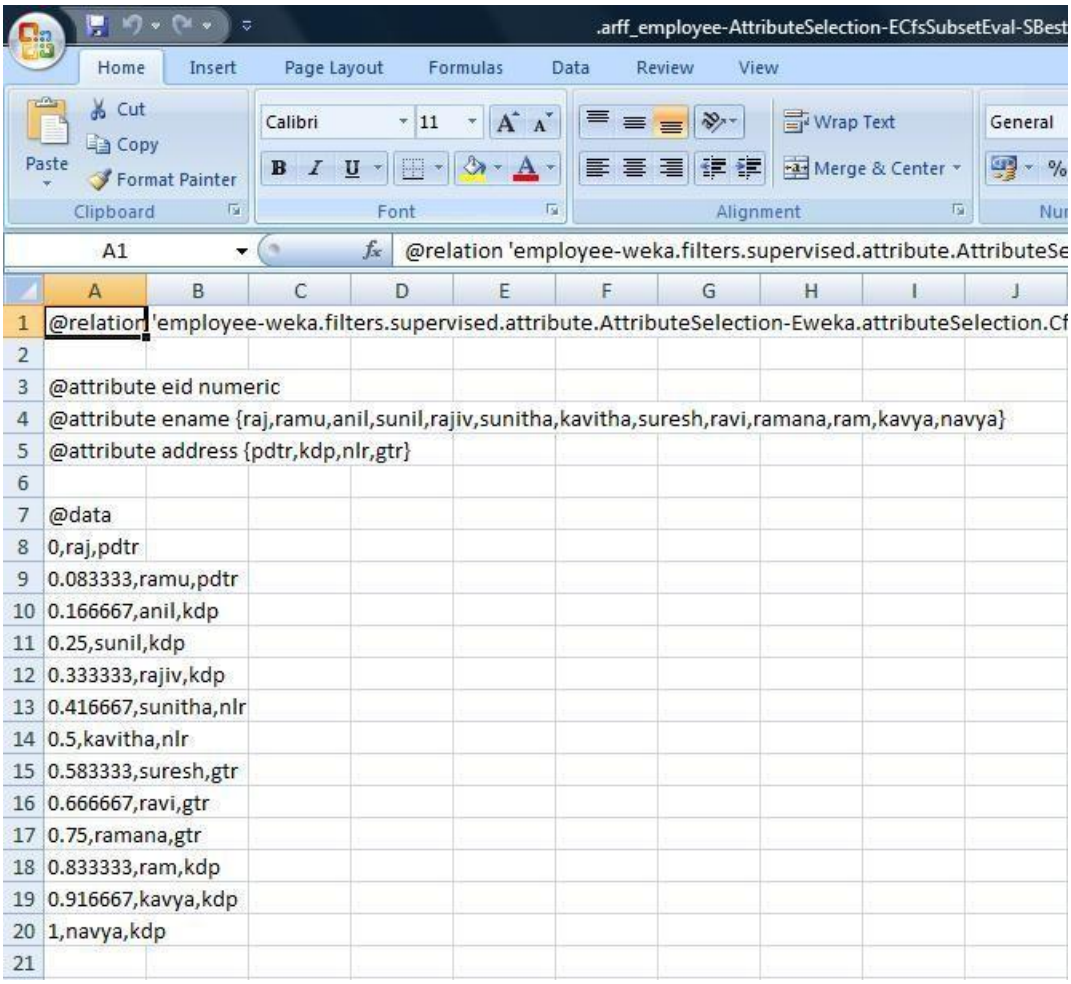
Procedure for Knowledge Flow:

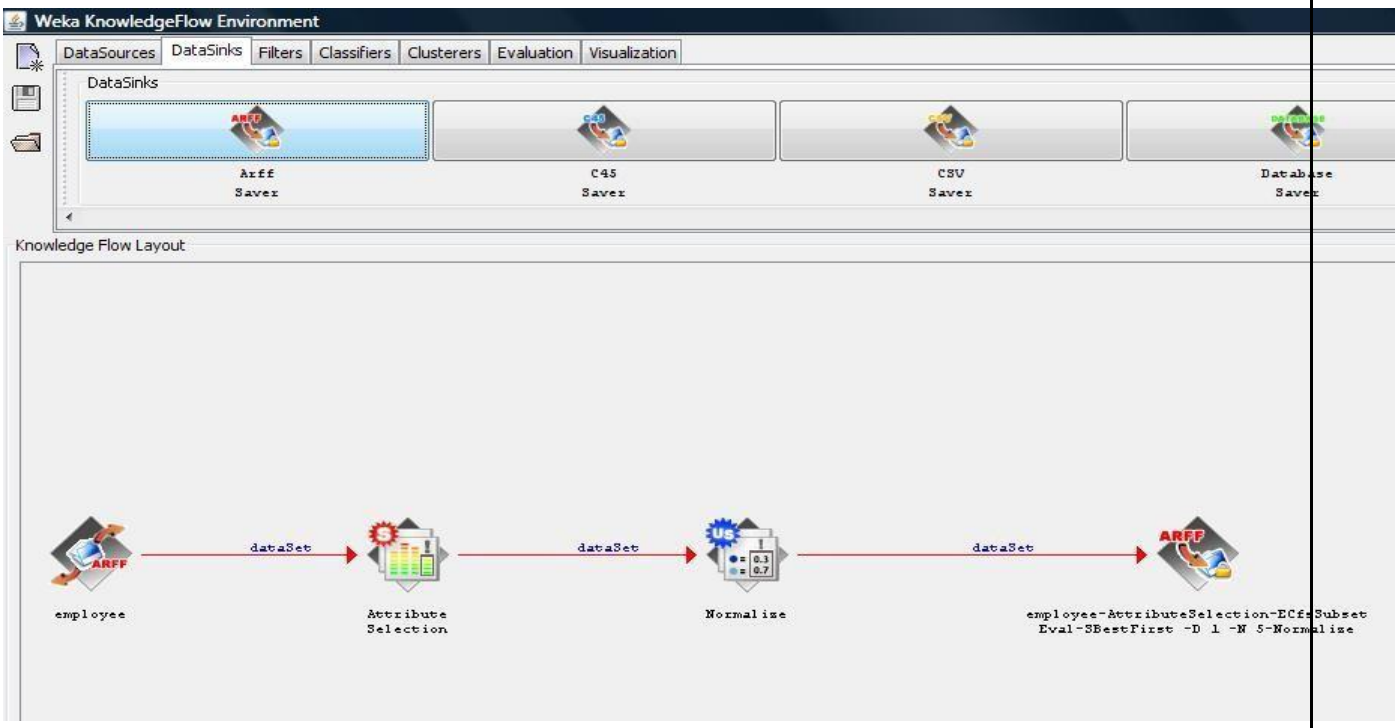
1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open the **Knowledge Flow**.
3) Select the **Data Source component** and **add Arff Loader** into the **knowledge layout canvas**.
4) Select the **Filters component** and **add Attribute Selection** and N**ormalize** into the knowledge layout canvas.
5) Select the **Data Sinks** component and **add Arff Saver** into the knowledge layout canvas.
6) Right click on **Arff Loader** and select **Configure option** then the new window will be opened and select **Employee.arff**
7) Right click on **Arff Loader** and select **Dataset option** then establish a link between **Arff Loader** and A**ttribute Selection**.
8) Right click on **Attribute Selection** and select **Dataset option** then establish a link between **Attribute Selection** and **Normalize**.
9) Right click on **Attribute Selection** and select **Configure option** and choose the best attribute for Employee data.
10) Right click on **Normalize** and select **Dataset option** then establish a link between **Normalize** and **Arff Saver**.
11) Right click on **Arff Saver** and select **Configure option** then new window will be opened and set the path, enter **.arff** in look in dialog box to save normalize data.
12) Right click on **Arff Loader** and click on **Start Loading option** then everything will be executed one by one.

13) Check whether output is created or not by selecting the preferred path.

14) Rename the data name as **a.arff**

15) Double click on **a.arff** then automatically the output will be opened in **MS-Excel**.





Result:

This program has been successfully executed.

EXPERIMENT NO:11

Aim: Finding Association Rules for Buying data.

Description:

In data mining, **association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases. It can be described as analyzing and presenting strong rules discovered in databases using different measures of interestingness. In market basket analysis association rules are used and they are also employed in many application areas including Web usage mining, intrusion detection and bioinformatics.

Creation of Buying Table:

Procedure:

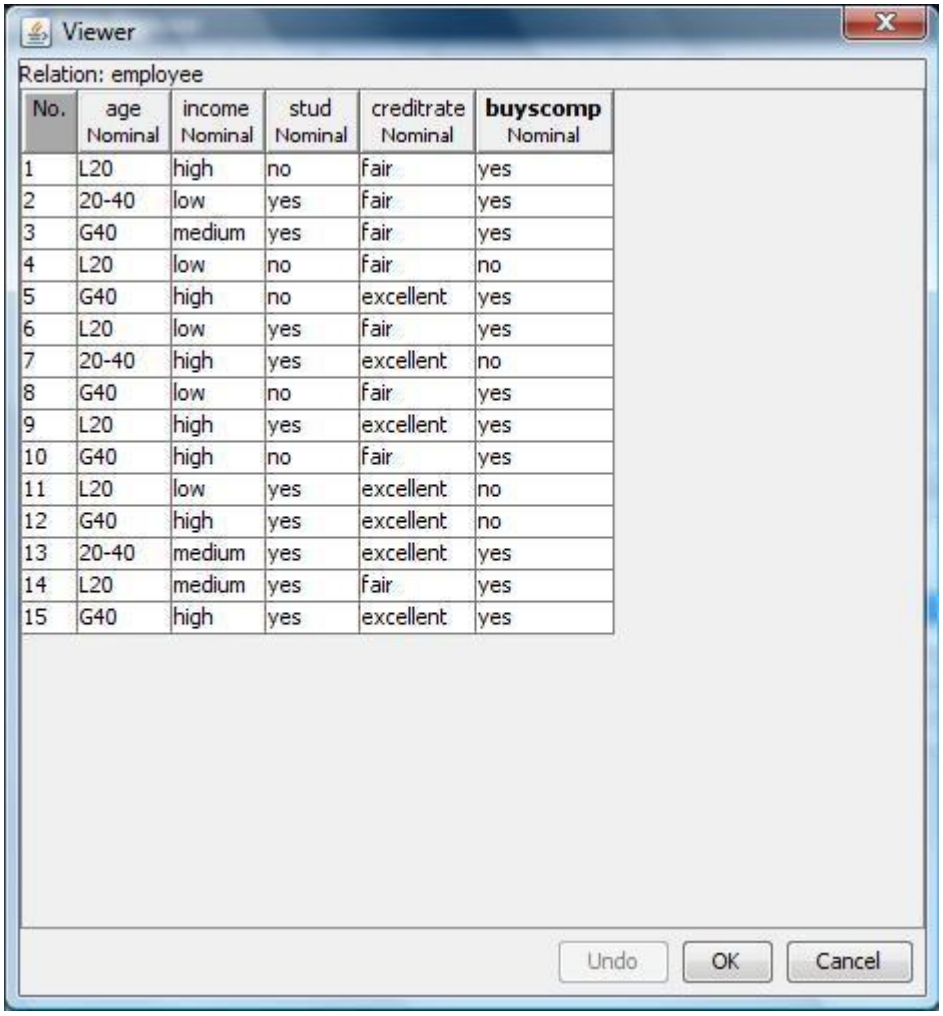
1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Buying Table.

```
@relation buying
@attribute age {L20,20-40,G40}
@attribute income {high,medium,low}
@attribute stud {yes,no}
@attribute creditrate {fair,excellent}
@attribute buyscomp {yes,no}
@data
L20,high,no,fair,yes
20-40,low,yes,fair,yes
G40,medium,yes,fair,yes
L20,low,no,fair,no
G40,high,no,excellent,yes
L20,low,yes,fair,yes
20-40,high,yes,excellent,no
G40,low,no,fair,yes
L20,high,yes,excellent,yes
G40,high,no,fair,yes
L20,low,yes,excellent,no
G40,high,yes,excellent,no
20-40,medium,yes,excellent,yes
L20,medium,yes,fair,yes
G40,high,yes,excellent,yes
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows buying table on weka.

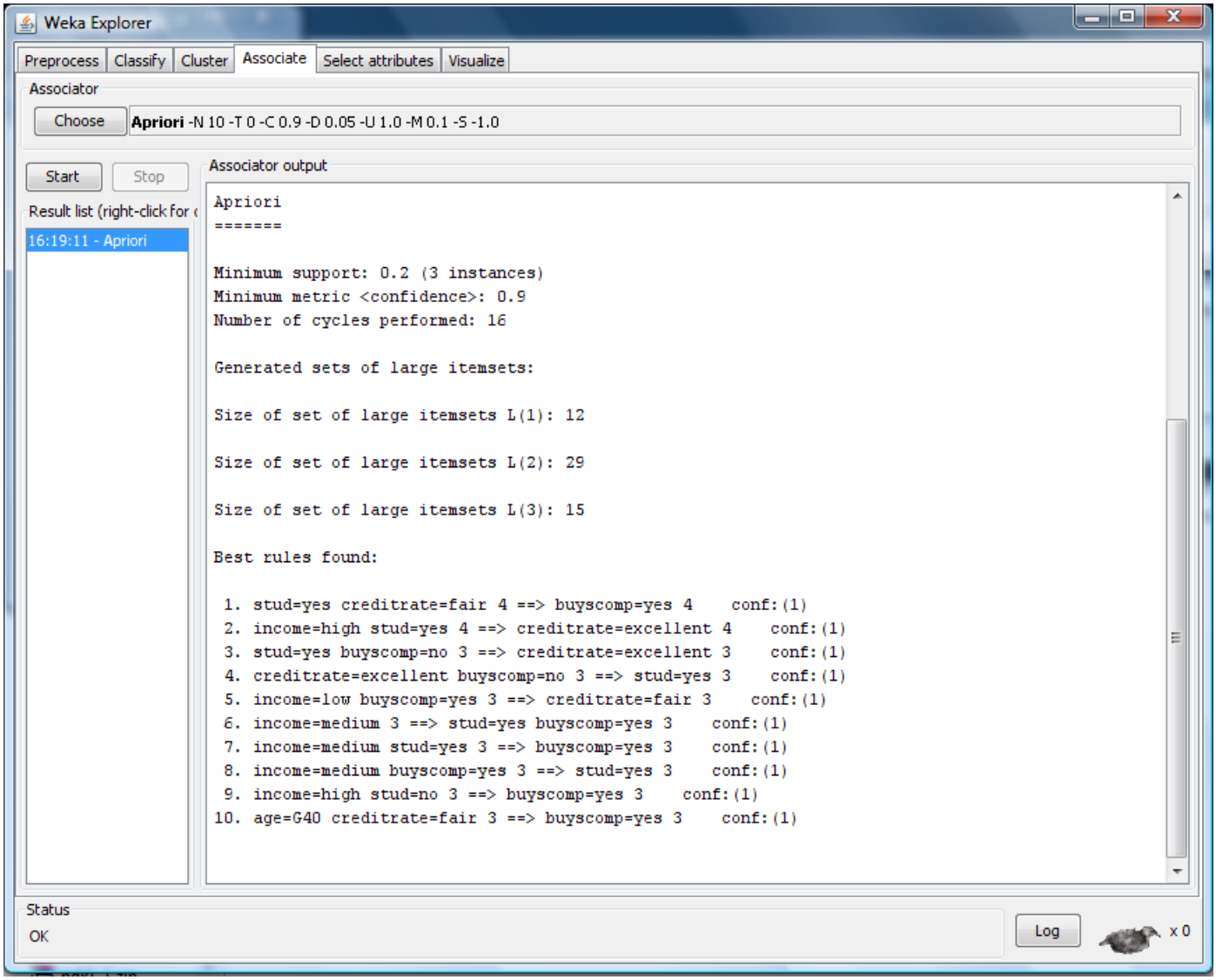
Output:

Training Data Set → Buying Table



Procedure for Association Rules:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **buying.arff**
4) Select **Associate option** on the top of the Menu bar.
5) Select **Choose button** and then click on **Apriori Algorithm**.
6) Click on **Start button** and output will be displayed on the **right side** of the window.



Result:

This program has been successfully executed.

EXPERIMENT NO:12

Aim: Finding Association Rules for Banking data.

Description:

In data mining, **association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases. It can be described as analyzing and presenting strong rules discovered in databases using different measures of interestingness. In market basket analysis association rules are used and they are also employed in many application areas including Web usage mining, intrusion detection and bioinformatics.

Creation of Banking Table:

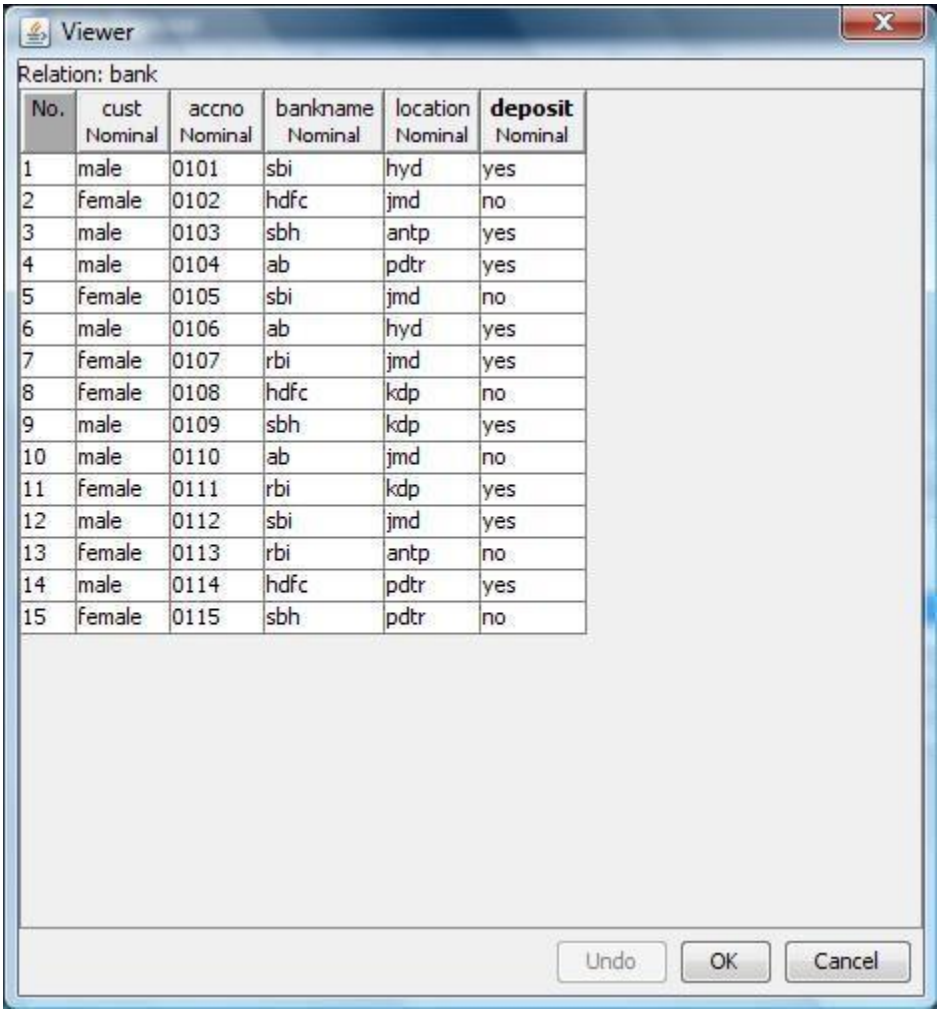
Procedure:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Banking Table.

```
@relation bank
@attribute cust {male,female}
@attribute accno
{0101,0102,0103,0104,0105,0106,0107,0108,0109,0110,0111,0112,0113,0114,0115}
@attribute bankname {sbi,hdfc,sbh,ab,rbi}
@attribute location {hyd,jmd,antp,pdtr,kdp}
@attribute deposit {yes,no}
@data
male,0101,sbi,hyd,yes
female,0102,hdfc,jmd,no
male,0103,sbh,antp,yes
male,0104,ab,pdtr,yes
female,0105,sbi,jmd,no
male,0106,ab,hyd,yes
female,0107,rbi,jmd,yes
female,0108,hdfc,kdp,no
male,0109,sbh,kdp,yes
male,0110,ab,jmd,no
female,0111,rbi,kdp,yes
male,0112,sbi,jmd,yes
female,0113,rbi,antp,no
male,0114,hdfc,pdtr,yes
female,0115,sbh,pdtr,no
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows banking table on weka.

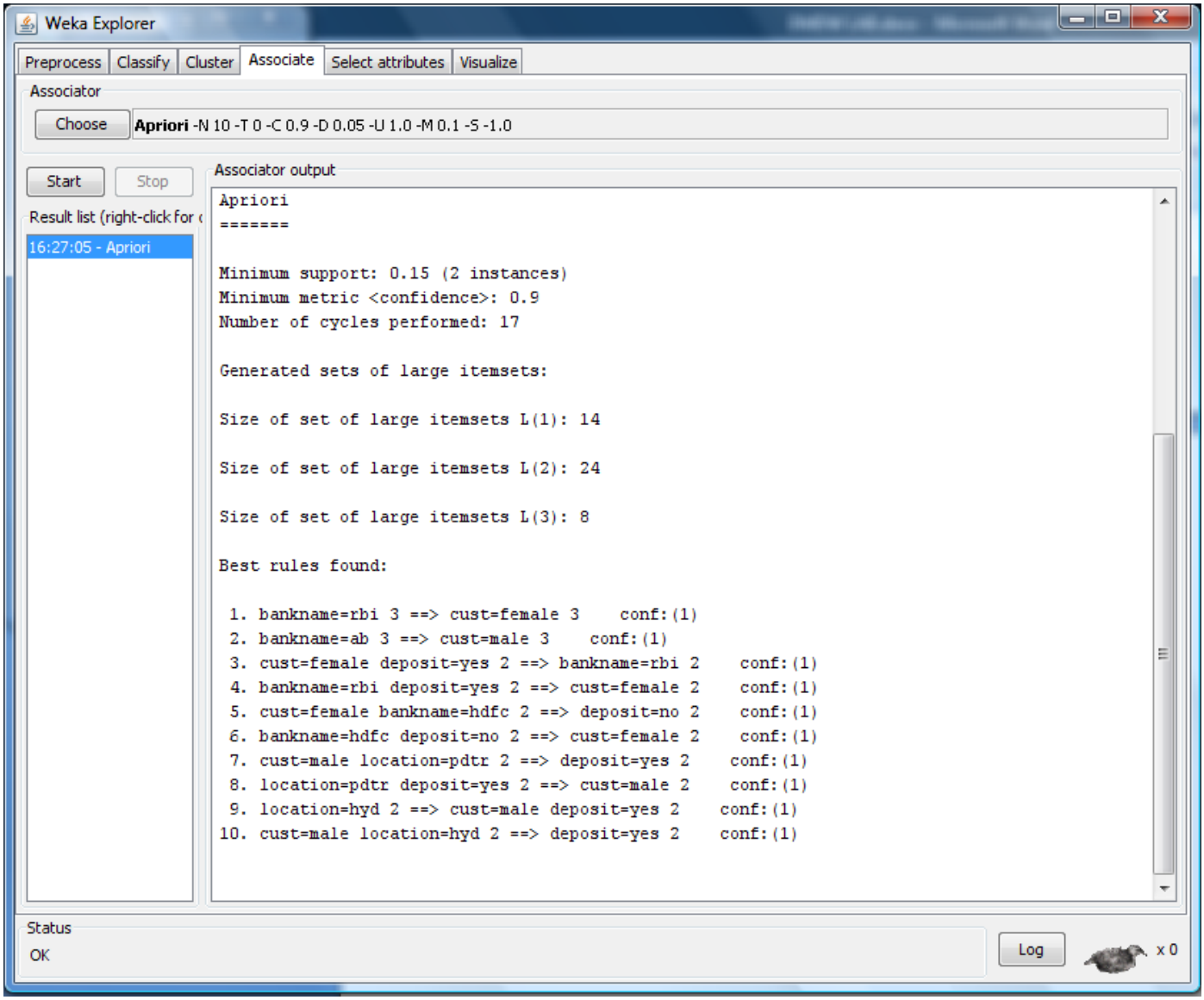
Training Data Set → Banking Table



Procedure for Association Rules:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **bank.arff**
4) Select **Associate option** on the top of the Menu bar.
5) Select **Choose button** and then click on **Apriori Algorithm**.
6) Click on **Start button** and output will be displayed on the **right side** of the window.

Output:



Result:

This program has been successfully executed.

EXPERIMENT NO:13

Aim: Finding Association Rules for Employee data.

Description:

In data mining, **association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases. It can be described as analyzing and presenting strong rules discovered in databases using different measures of interestingness. In market basket analysis association rules are used and they are also employed in many application areas including Web usage mining, intrusion detection and bioinformatics.

Creation of Banking Table:

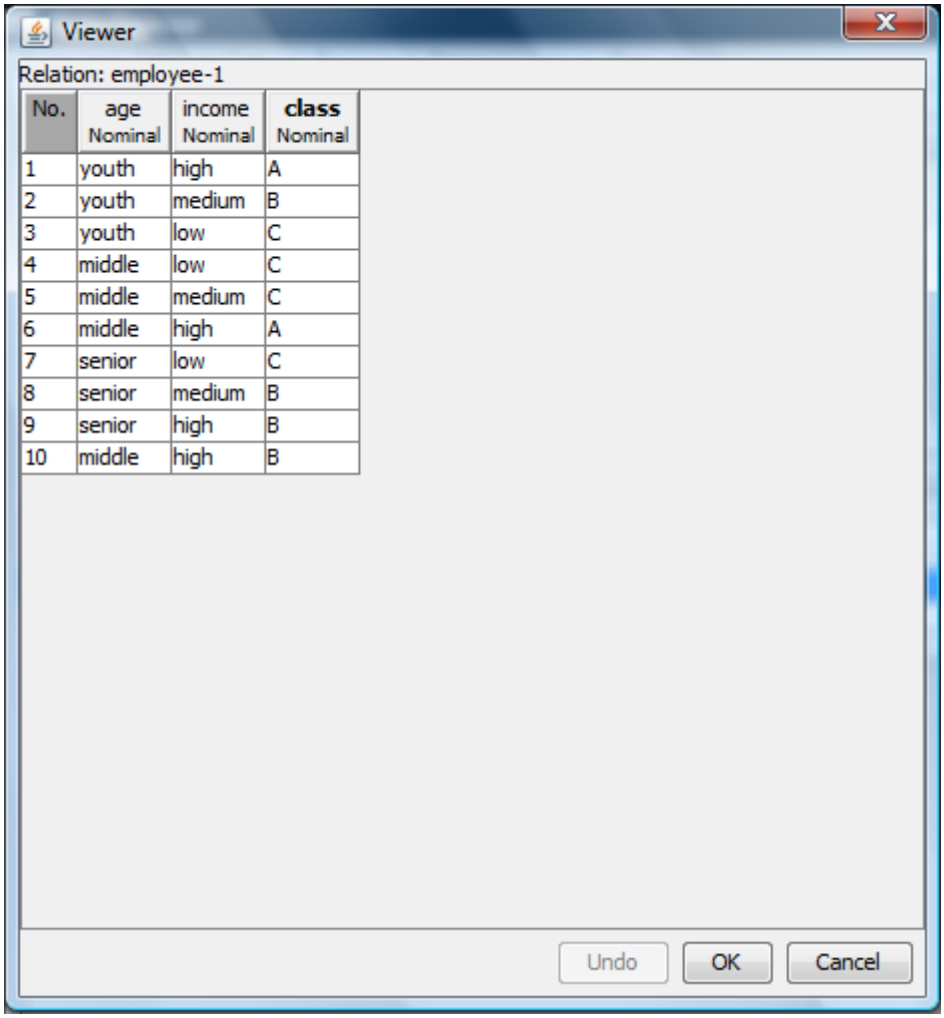
Procedure:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Employee Table.

```
@relation employee-1
@attribute age {youth, middle, senior}
@attribute income {high, medium, low}
@attribute class {A, B, C}

@data
youth, high, A
youth,medium,B
youth, low, C
middle, low, C
middle, medium, C
middle, high, A
senior, low, C
senior, medium, B
senior, high, B
middle, high, B
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows employee table on weka.

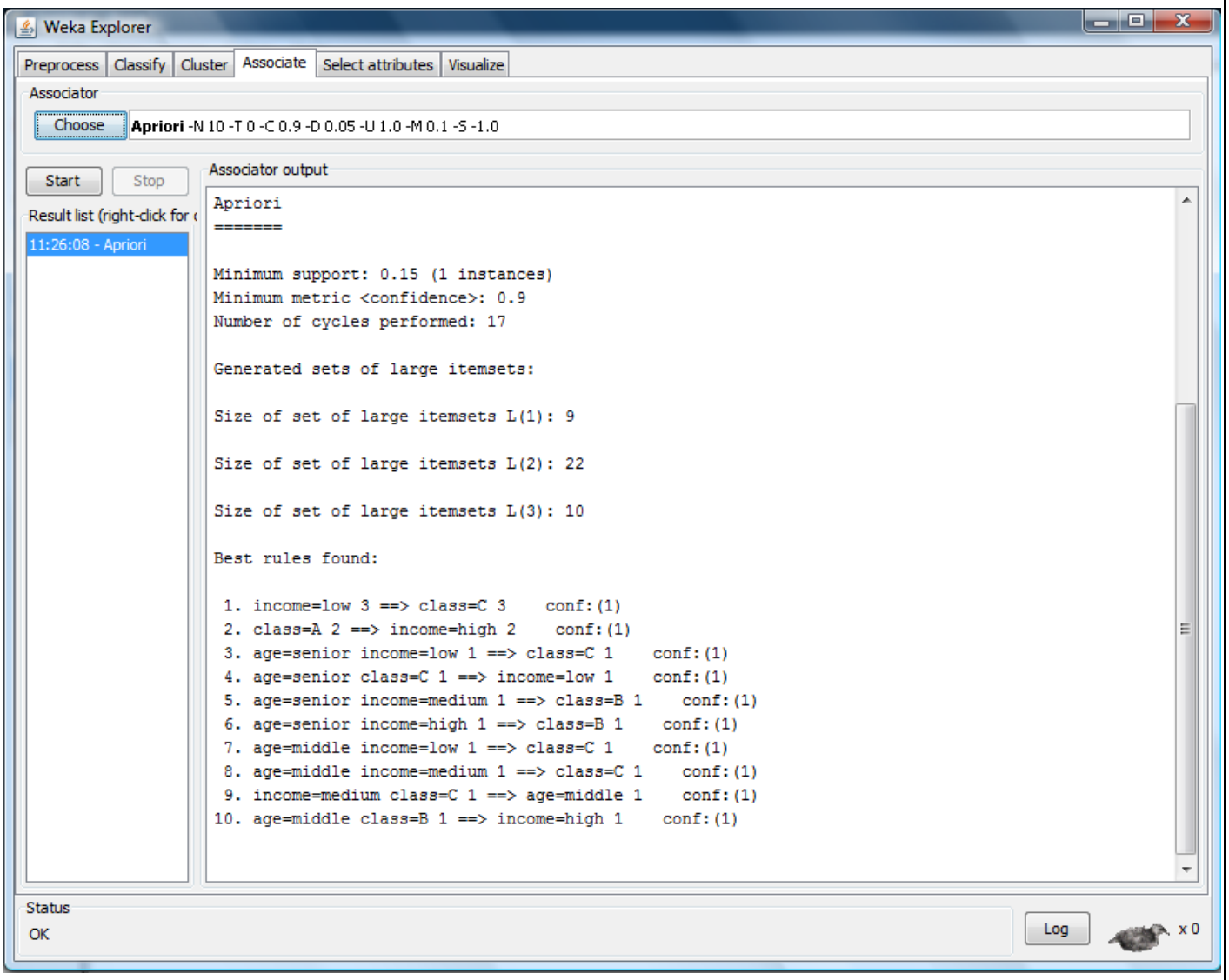
Training Data Set → Employee Table



Procedure for Association Rules:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **employee-1.arff**
4) Select **Associate option** on the top of the Menu bar.
5) Select **Choose button** and then click on **Apriori Algorithm**.
6) Click on **Start button** and output will be displayed on the **right side** of the window.

Output:



Result:

This program has been successfully executed.

EXPERIMENT NO:14

Aim:

To Construct Decision Tree for Weather data and classify it.

Description:

Classification & Prediction:

Classification is the process for finding a model that describes the data values and concepts for the purpose of Prediction.

Decision Tree:

A decision Tree is a classification scheme to generate a tree consisting of root node, internal nodes and external nodes.

Root nodes representing the attributes. Internal nodes are also the attributes. External nodes are the classes and each branch represents the values of the attributes

Decision Tree also contains set of rules for a given data set; there are two subsets in Decision Tree. One is a Training data set and second one is a Testing data set. Training data set is previously classified data. Testing data set is newly generated data.

Creation of Weather Table:

Procedure:

1) Open Start → Programs → Accessories → Notepad

2) Type the following training data set with the help of Notepad for Weather Table.

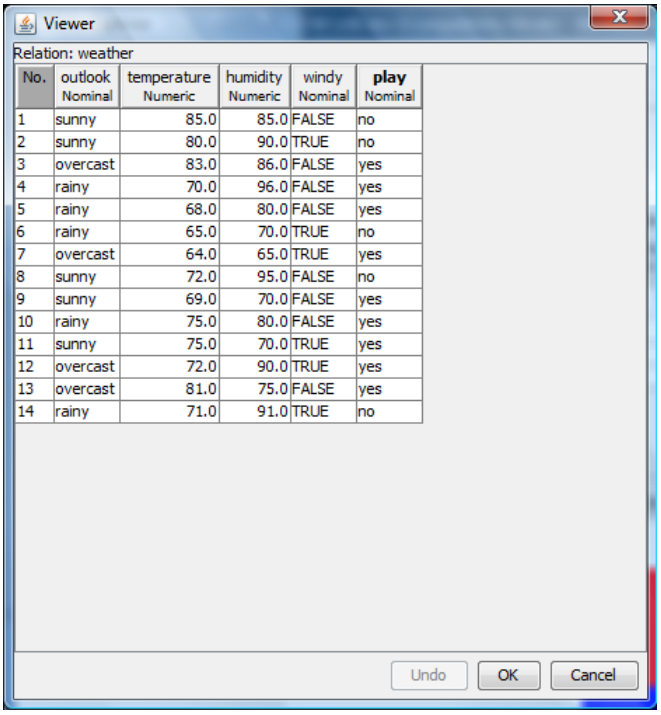
```
@relation weather
@attribute outlook {sunny, rainy, overcast}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@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
```

3) After that the file is saved with **.arff** file format.

4) Minimize the arff file and then open Start → Programs → weka-3-4.

5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows weather table on weka.

Training Data Set → Weather Table

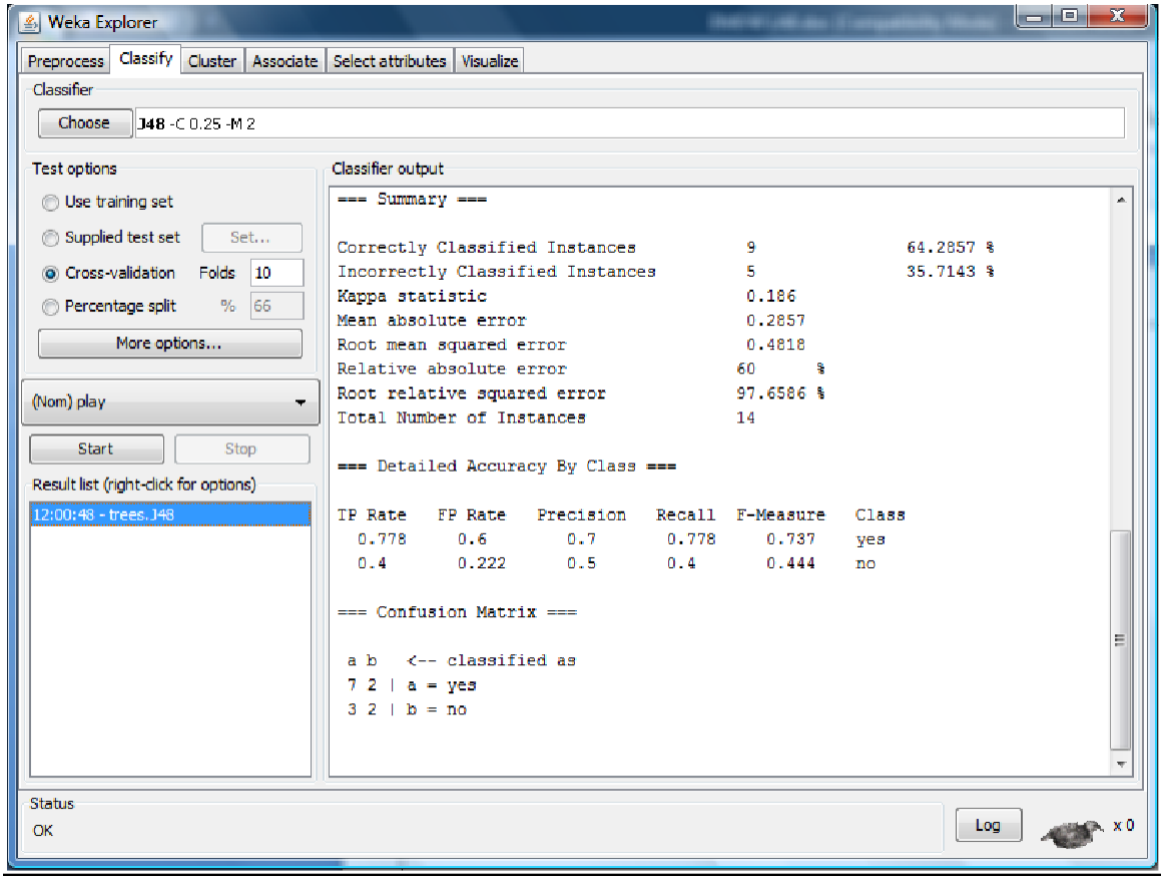


No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	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

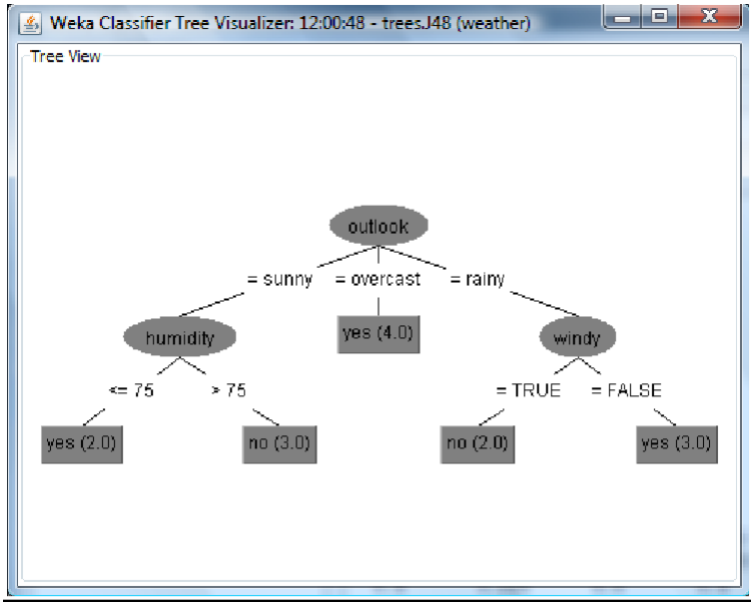
Procedure for Decision Trees:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **weather.arff**
4) Select **Classifier option** on the top of the Menu bar.
5) Select **Choose button** and click on **Tree option**.
6) Click on **J48.**
7) Click on **Start button** and output will be displayed on the **right side** of the window.
8) Select the **result list** and **right click** on result list and select **Visualize Tree option**.
9) Then **Decision Tree** will be displayed on **new window**.

Output:



Decision Tree:



Result: This program has been successfully executed.

EXPERIMENT NO:11

Aim:

To Construct Decision Tree for Customer data and classify it.

Description:

Classification & Prediction:

Classification is the process for finding a model that describes the data values and concepts for the purpose of Prediction.

Decision Tree:

A decision Tree is a classification scheme to generate a tree consisting of root node, internal nodes and external nodes.

Root nodes representing the attributes. Internal nodes are also the attributes. External nodes are the classes and each branch represents the values of the attributes

Decision Tree also contains set of rules for a given data set; there are two subsets in Decision Tree. One is a Training data set and second one is a Testing data set. Training data set is previously classified data. Testing data set is newly generated data.

Creation of Customer Table:

Procedure:

1) Open Start → Programs → Accessories → Notepad

2) Type the following training data set with the help of Notepad for Customer Table.

```
@relation customer
@attribute name {x,y,z,u,v,l,w,q,r,n}
@attribute age {youth,middle,senior}
@attribute income {high,medium,low}
@attribute class {A,B}

@data
x,youth,high,A
y,youth,low,B
z,middle,high,A
u,middle,low,B
v,senior,high,A
l,senior,low,B
w,youth,high,A
q,youth,low,B
r,middle,high,A
n,senior,high,A
```

3) After that the file is saved with **.arff** file format.

4) Minimize the arff file and then open Start → Programs → weka-3-4.

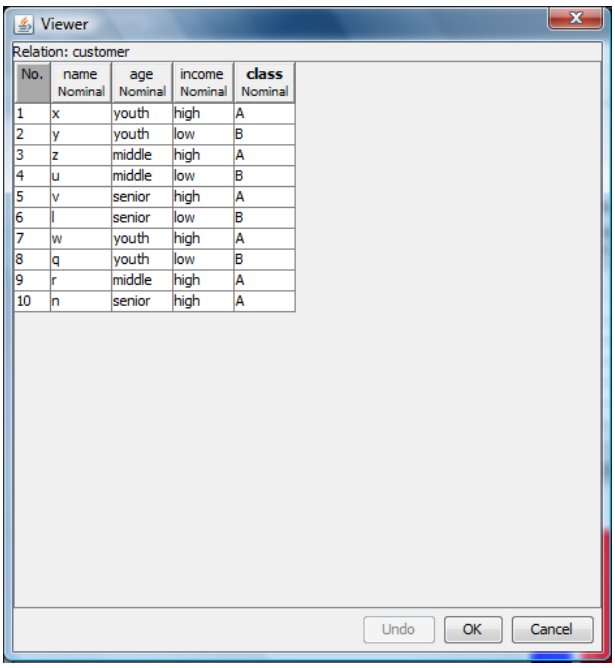
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

6) In that dialog box there are four modes, click on **explorer**.

7) Explorer shows many options. In that click on **'open file'** and select the arff file

8) Click on **edit button** which shows customer table on weka.

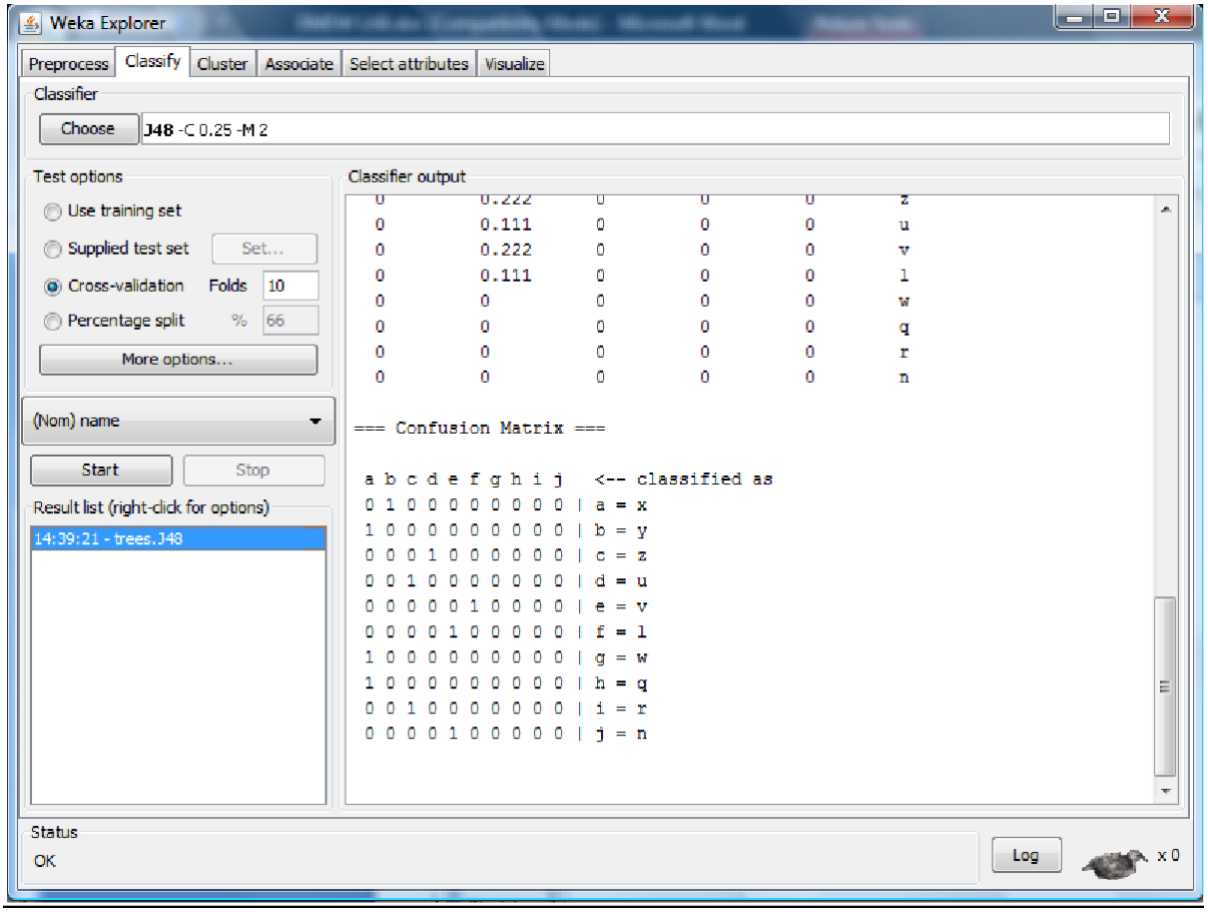
Training Data Set → Customer Table



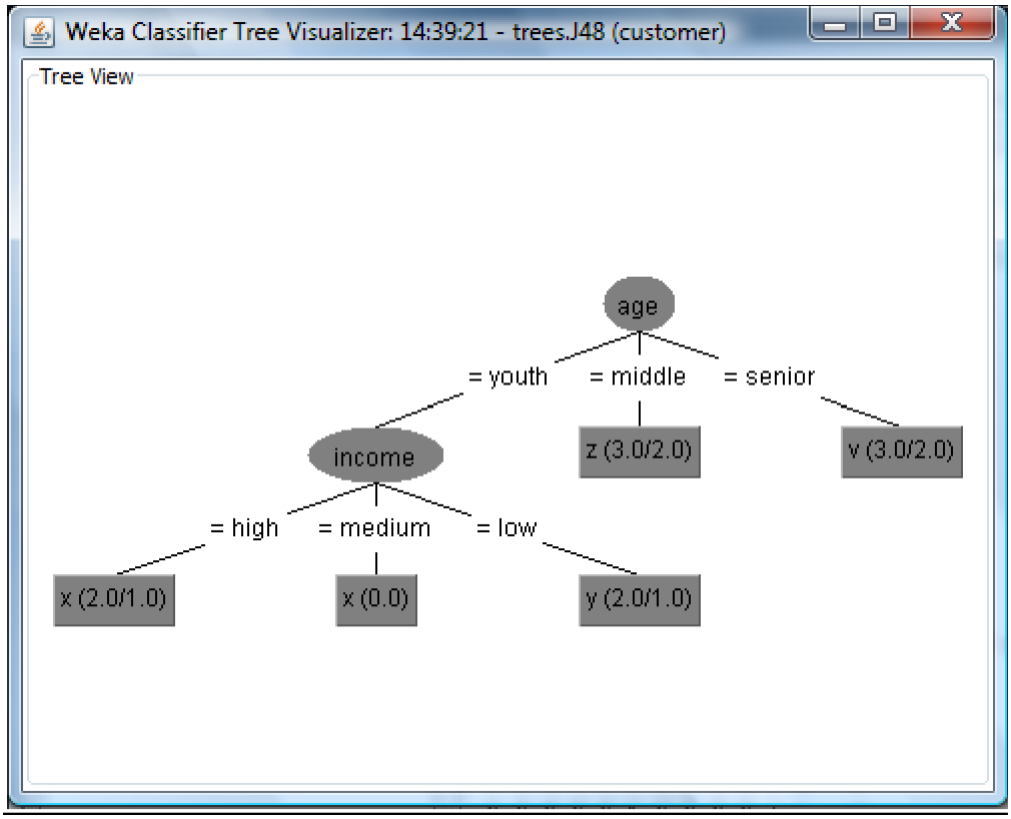
Procedure for Decision Trees:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **customer.arff**
4) Select **Classifier option** on the top of the Menu bar.
5) Select **Choose button** and click on **Tree option**.
6) Click on **J48.**
7) Click on **Start button** and output will be displayed on the **right side** of the window.
8) Select the **result list** and **right click** on result list and select **Visualize Tree option**.
9) Then **Decision Tree** will be displayed on **new window**.

Output:



Decision Tree:



Result: This program has been successfully executed.

EXPERIMENT NO:15

Aim:

To Construct Decision Tree for Location data and classify it.

Description:

Classification & Prediction:

Classification is the process for finding a model that describes the data values and concepts for the purpose of Prediction.

Decision Tree:

A decision Tree is a classification scheme to generate a tree consisting of root node, internal nodes and external nodes.

Root nodes representing the attributes. Internal nodes are also the attributes. External nodes are the classes and each branch represents the values of the attributes

Decision Tree also contains set of rules for a given data set; there are two subsets in Decision Tree. One is a Training data set and second one is a Testing data set. Training data set is previously classified data. Testing data set is newly generated data.

Creation of Weather Table:

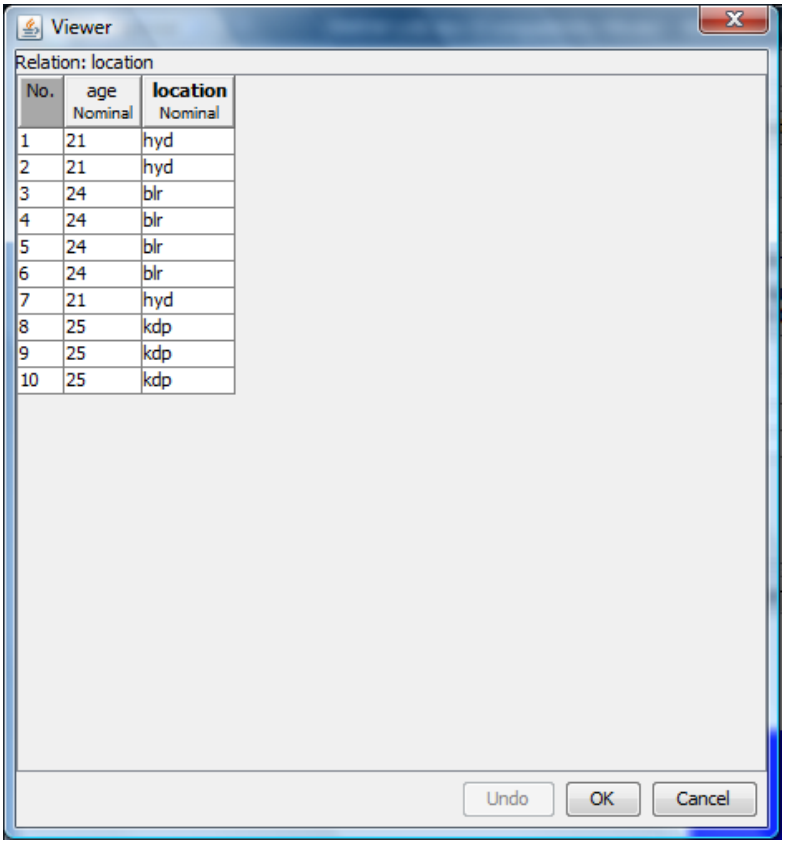
Procedure:

1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Location Table.

```
@relation location
@attribute age {21,24,25}
@attribute location {hyd,blr,kdp}

@data
21,hyd
21,hyd
24,blr
24,blr
24,blr
24,blr
21,hyd
25,kdp
25,kdp
25,kdp
```

3) After that the file is saved with **.arff** file format.
4) Minimize the arff file and then open Start → Programs → weka-3-4.
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.
6) In that dialog box there are four modes, click on **explorer**.
7) Explorer shows many options. In that click on **'open file'** and select the arff file
8) Click on **edit button** which shows location table on weka.

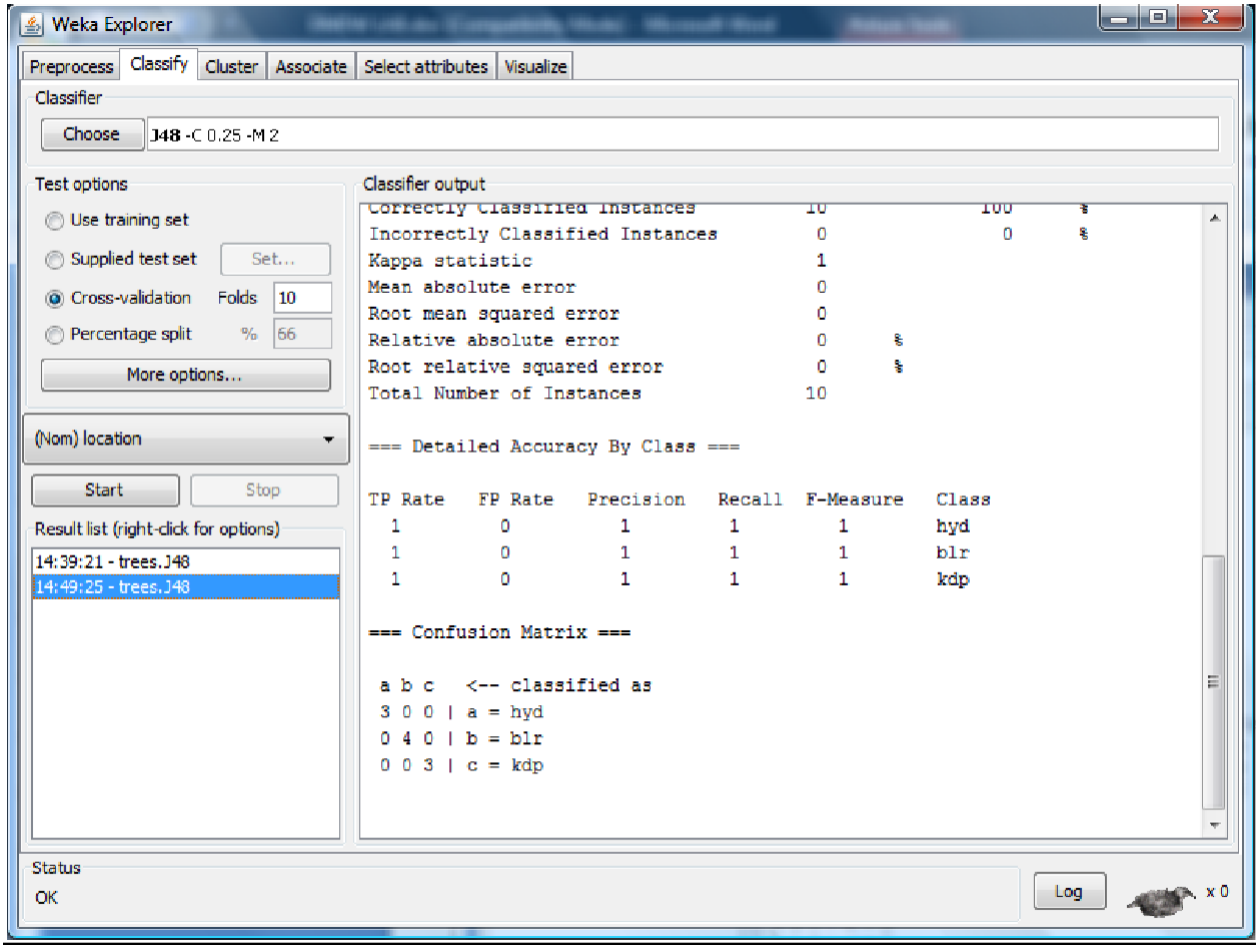
Training Data Set → Location Table



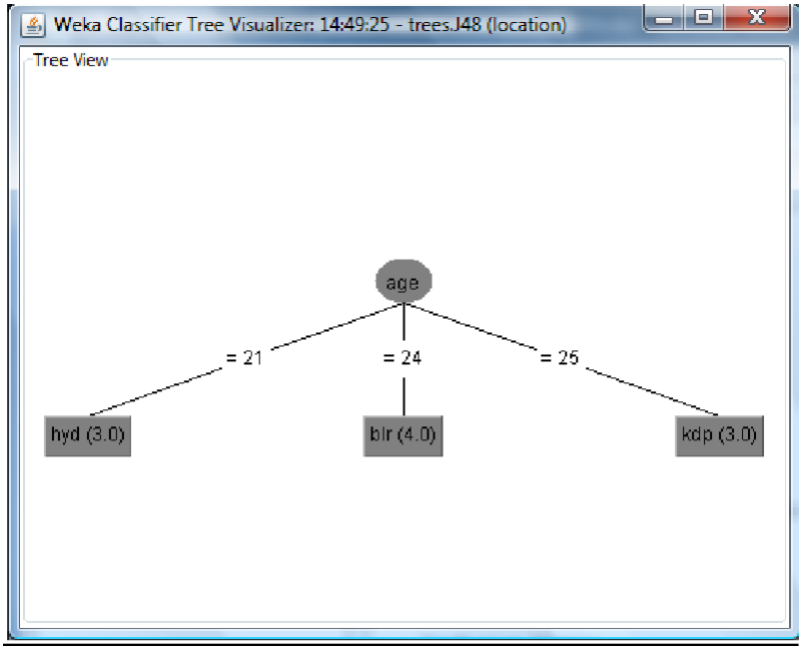
Procedure for Decision Trees:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open **explorer**.
3) Click on **open file** and select **location.arff**
4) Select **Classifier option** on the top of the Menu bar.
5) Select **Choose button** and click on **Tree option**.
6) Click on **J48.**
7) Click on **Start button** and output will be displayed on the **right side** of the window.
8) Select the **result list** and **right click** on result list and select **Visualize Tree option**.
9) Then **Decision Tree** will be displayed on **new window**.

Output:



Decision Tree:



Result:

This program has been successfully executed.