LAB MANUAL

DATA WAREHOUSE AND DATA MINING

EXPERIMENT NO: 1

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:

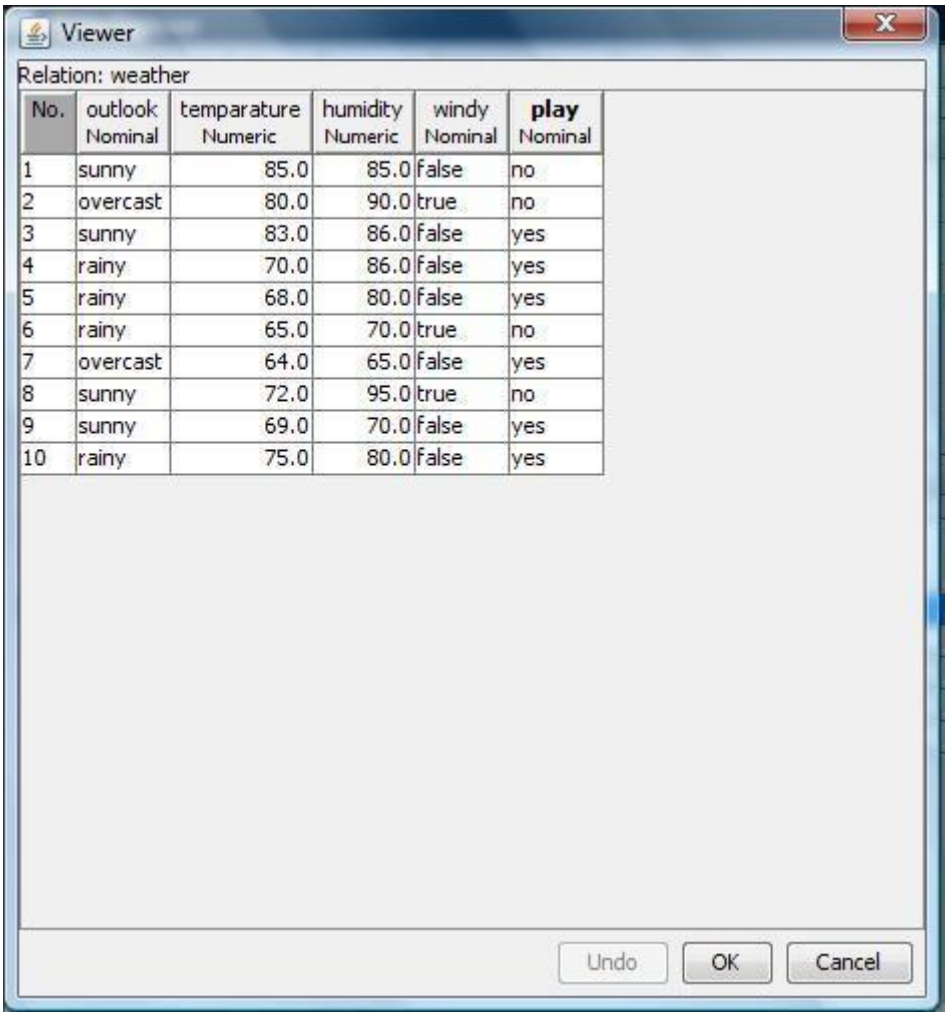
Steps:

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



Result:

This program has been successfully executed.

EXPERIMENT NO:2

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:

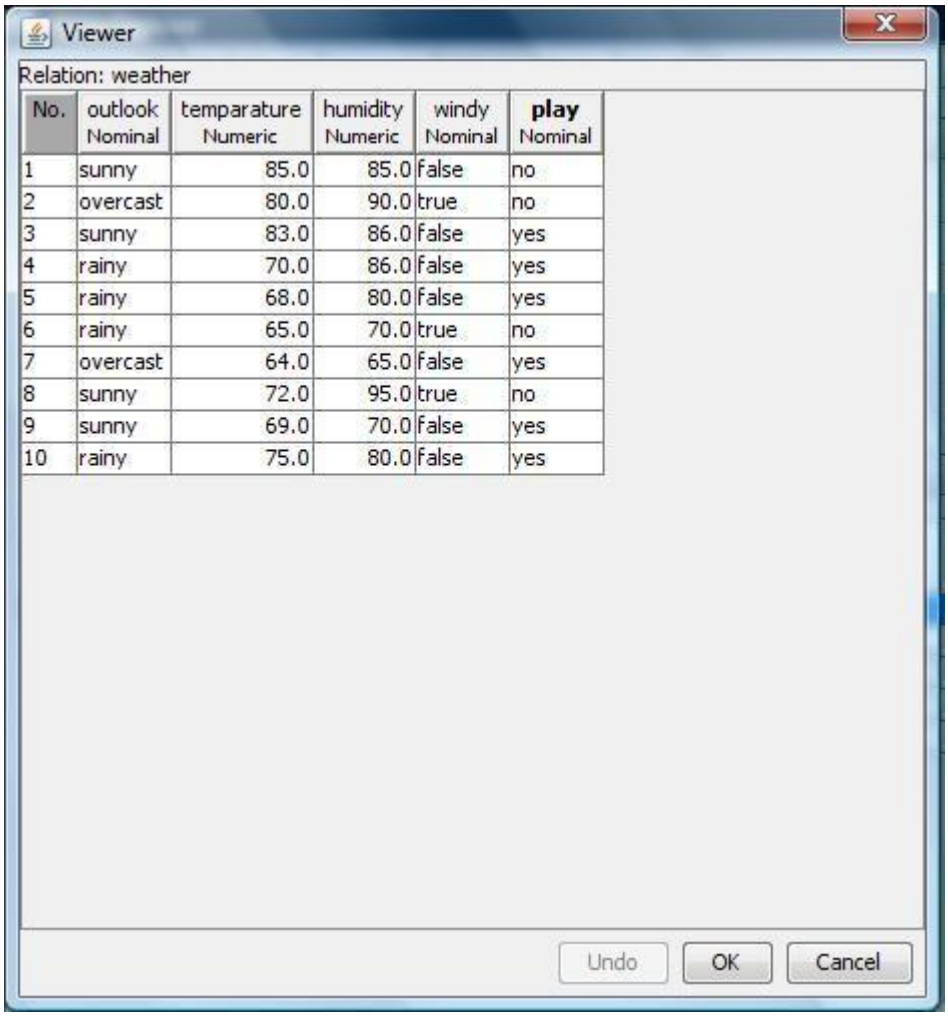
Steps:

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:3

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

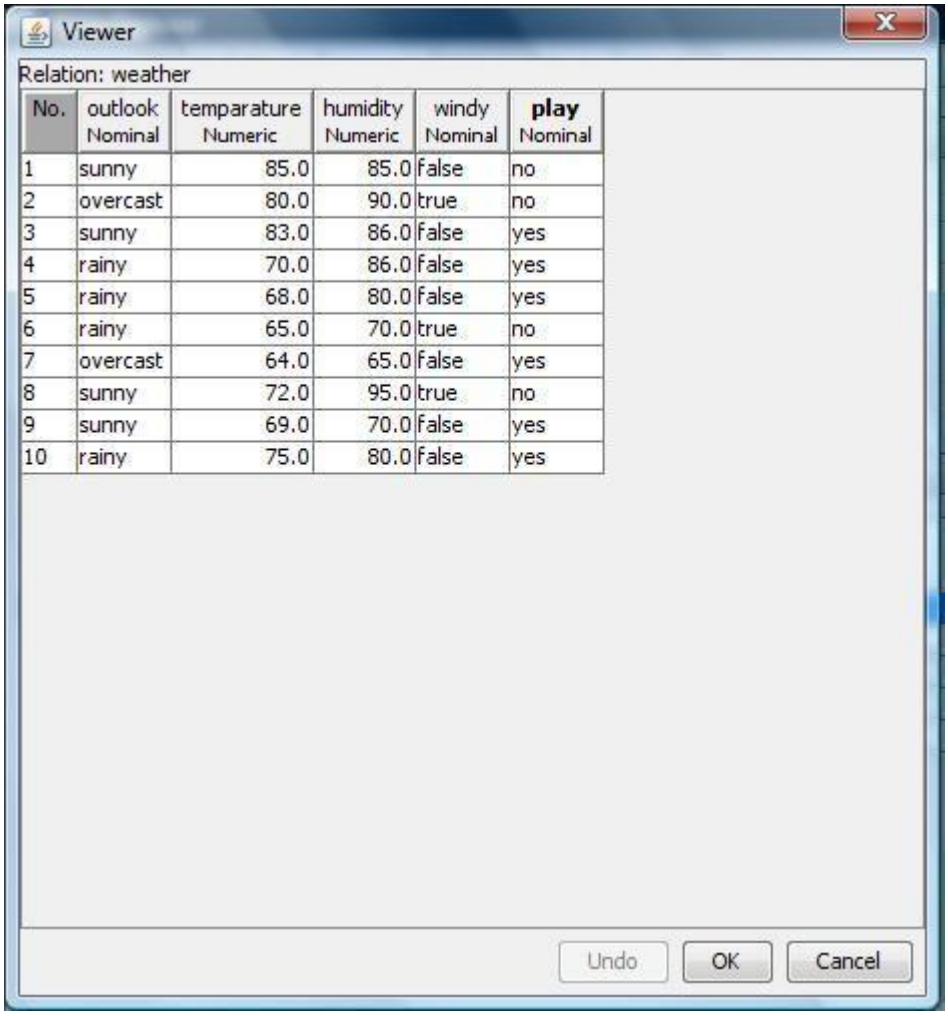
Creation of Weather Table:

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.

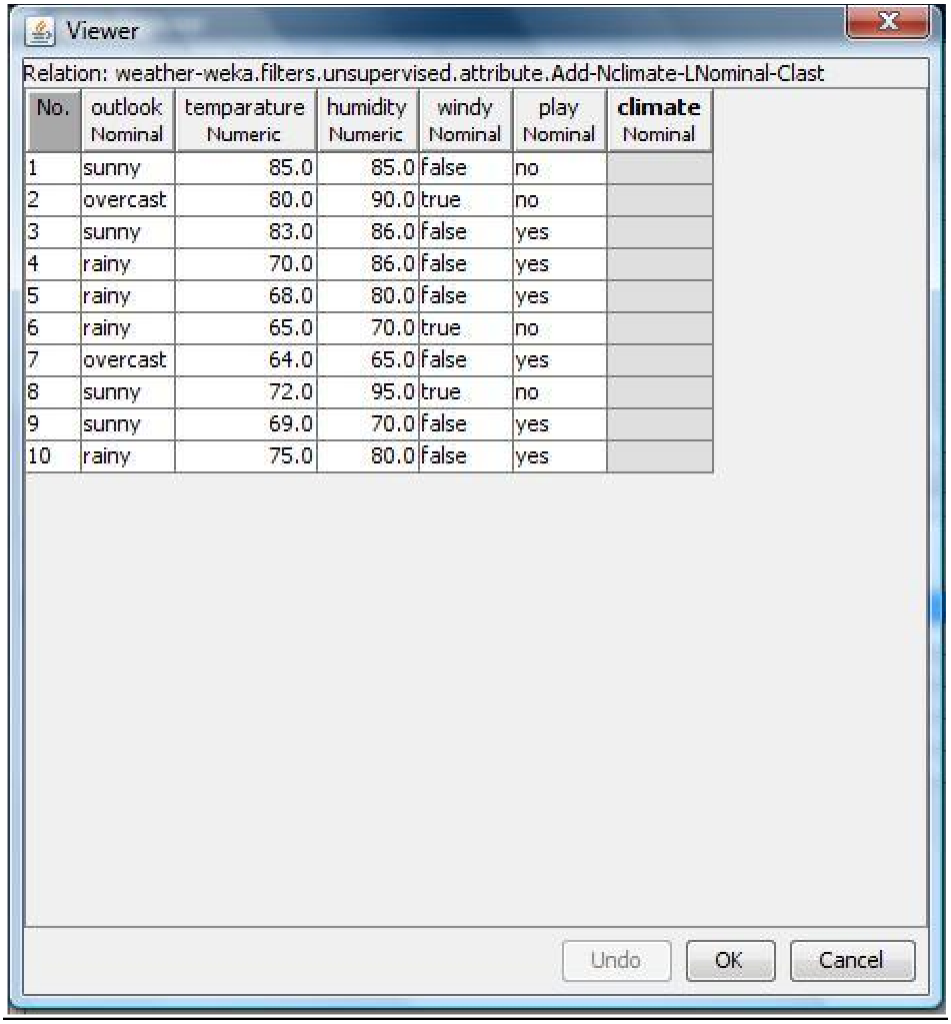


Add → Pre-Processing Technique:

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**.

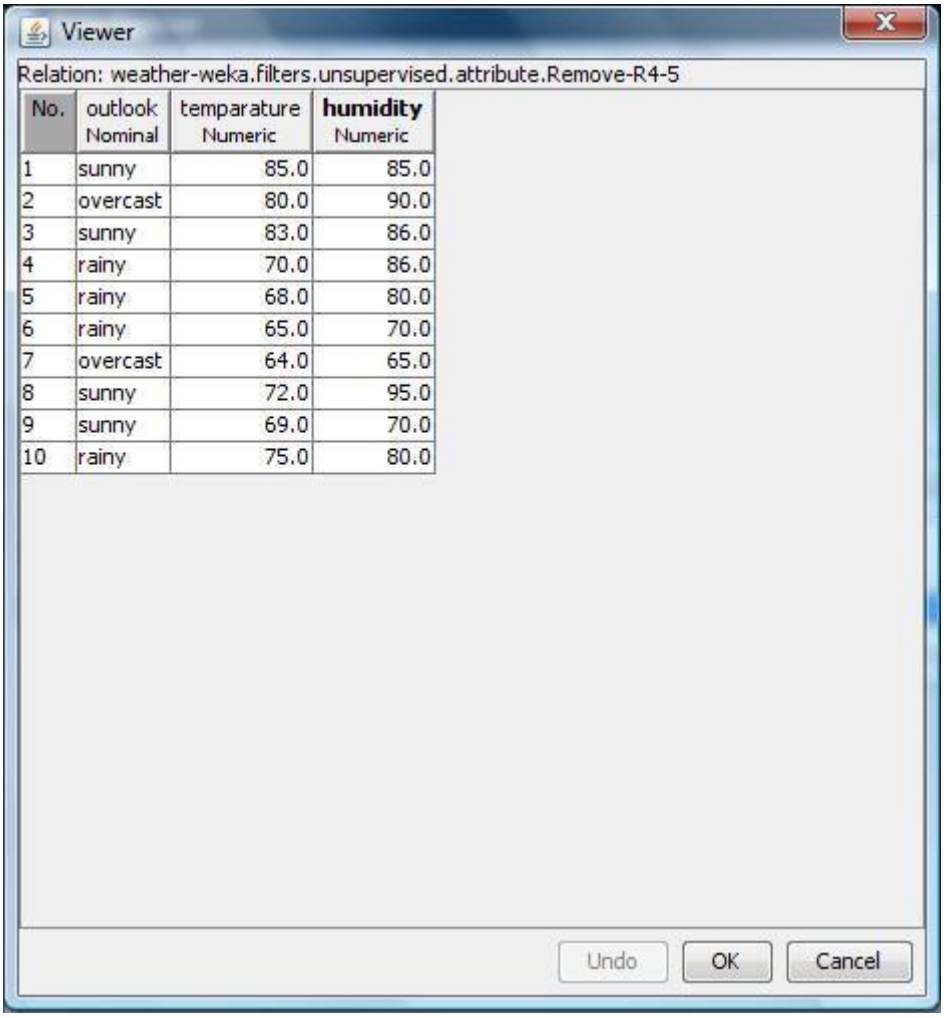
8) Select the attribute **Remove**.

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:

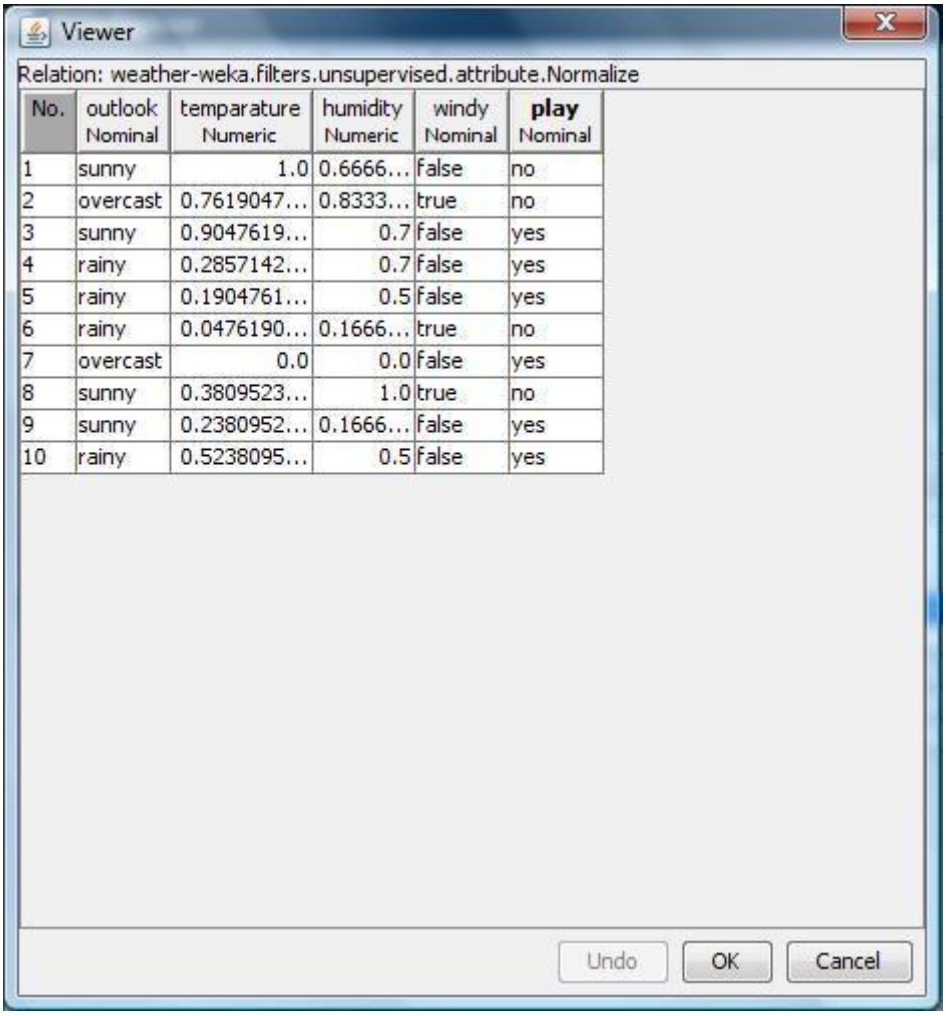


Normalize → Pre-Processing Technique:

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:4

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:

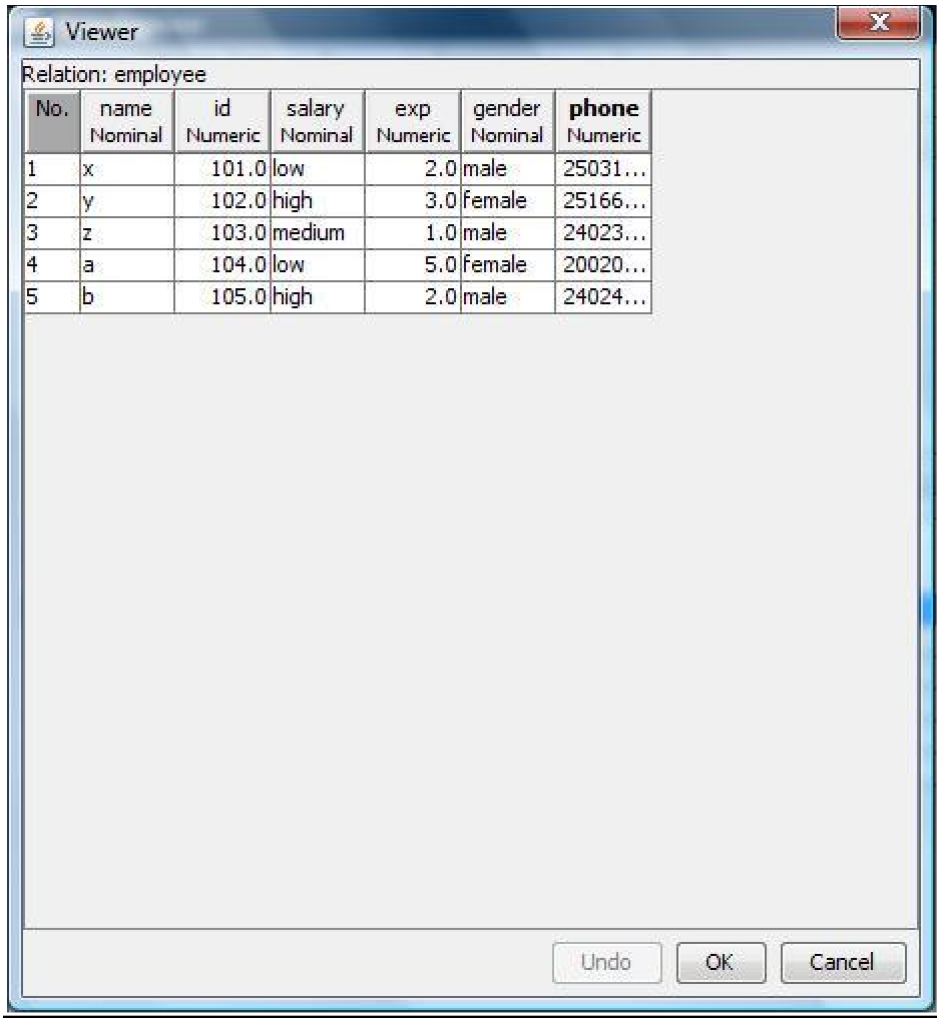
Procedure:

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**.

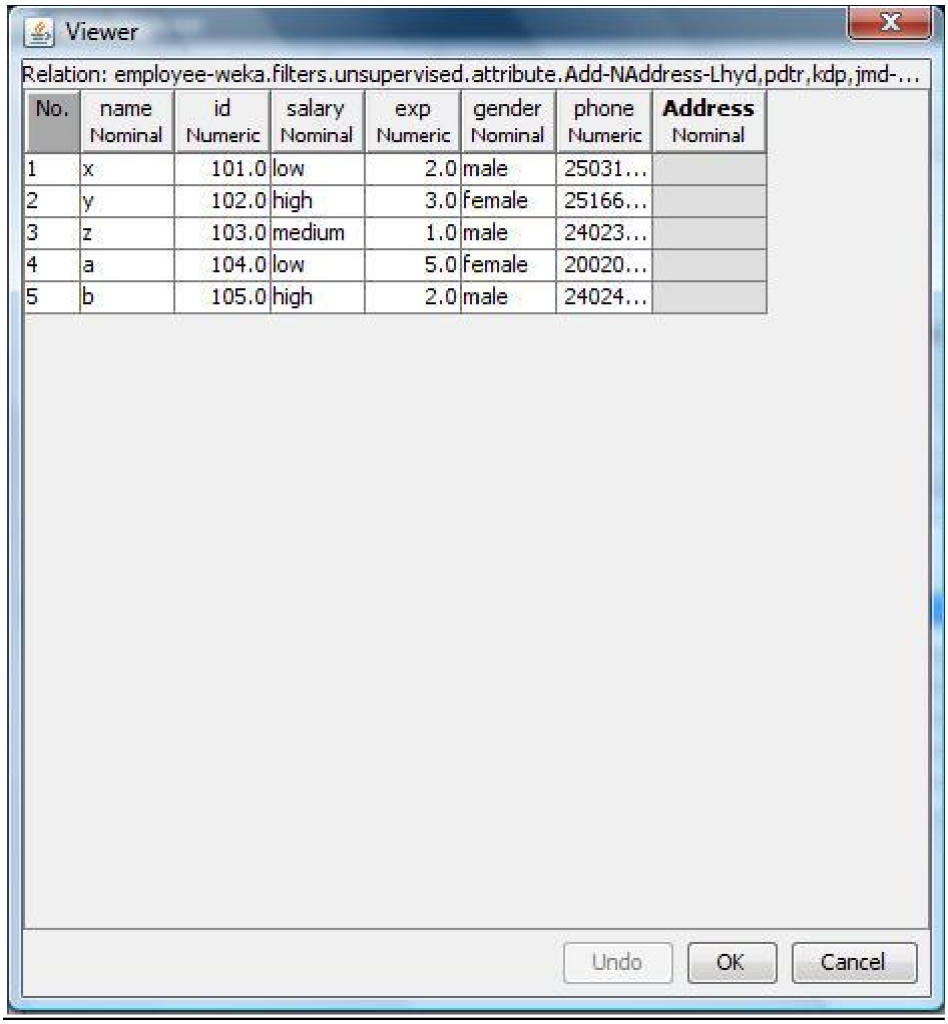
11) Click on **OK**.

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:

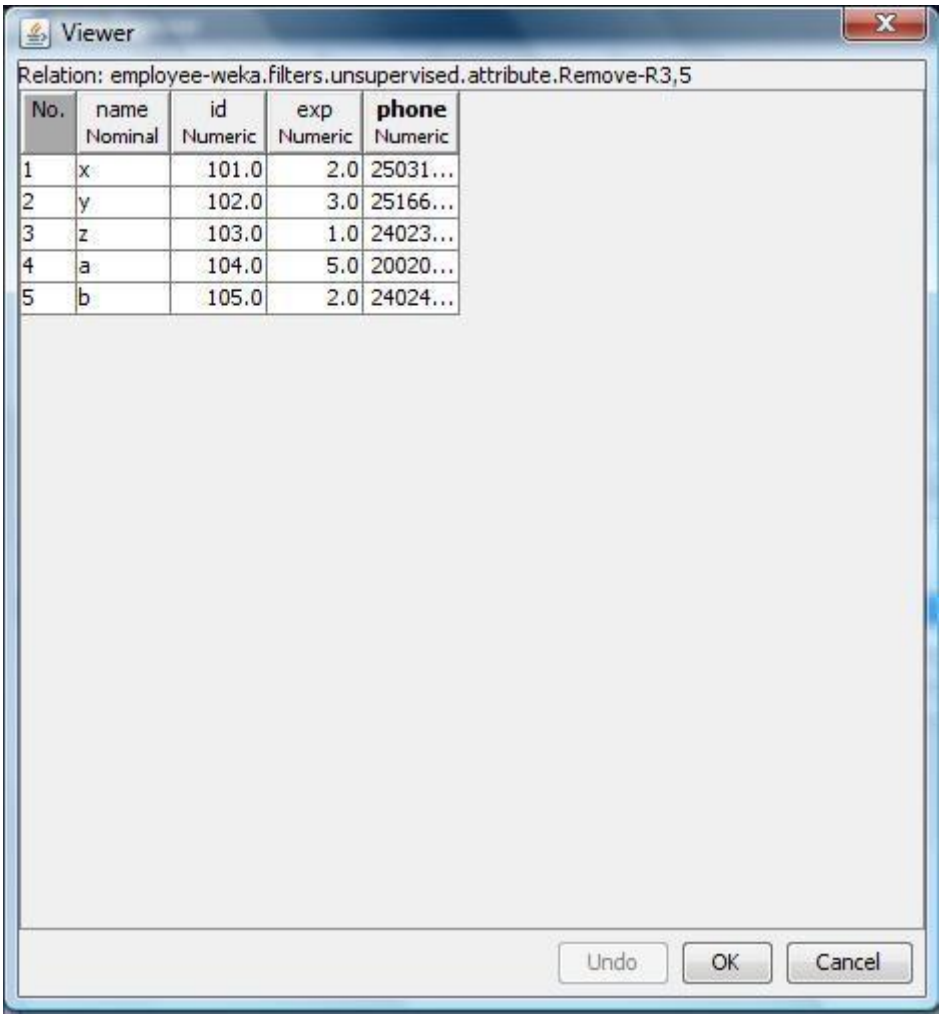


Remove → Pre-Processing Technique:

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:

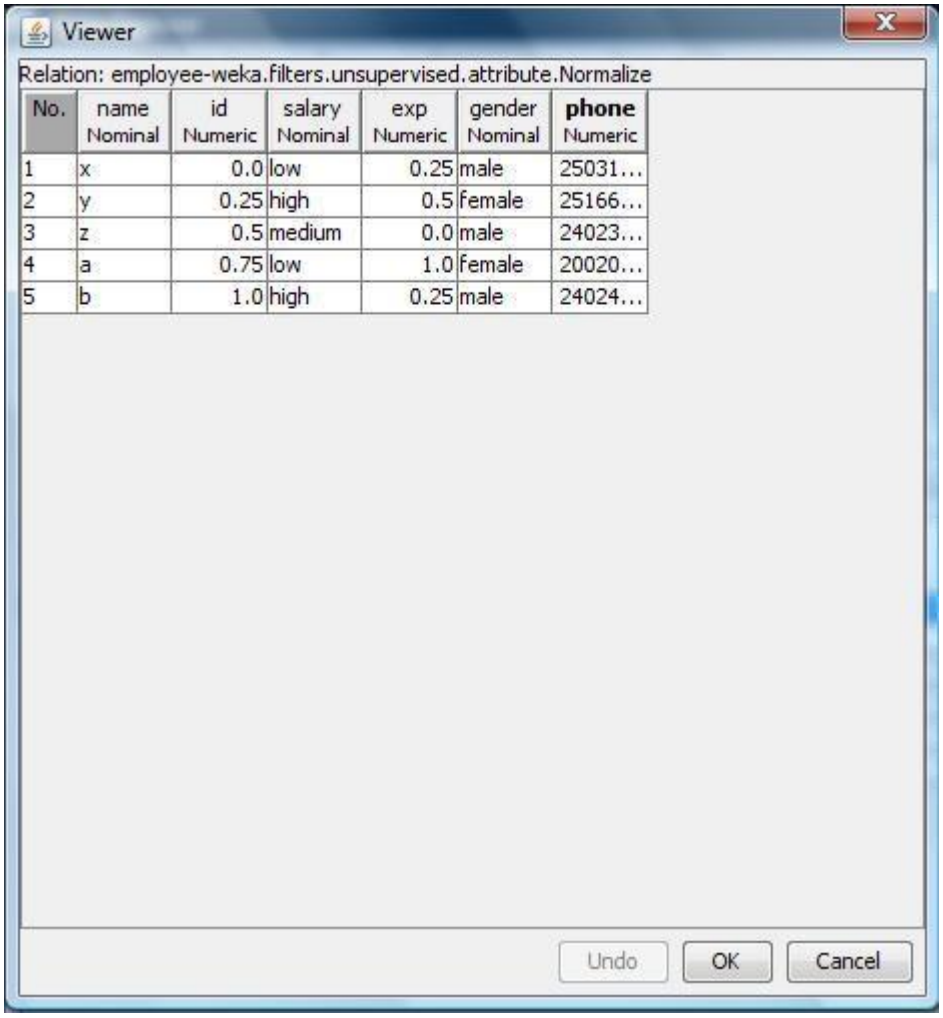


Normalize → Pre-Processing Technique:

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: 5

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:

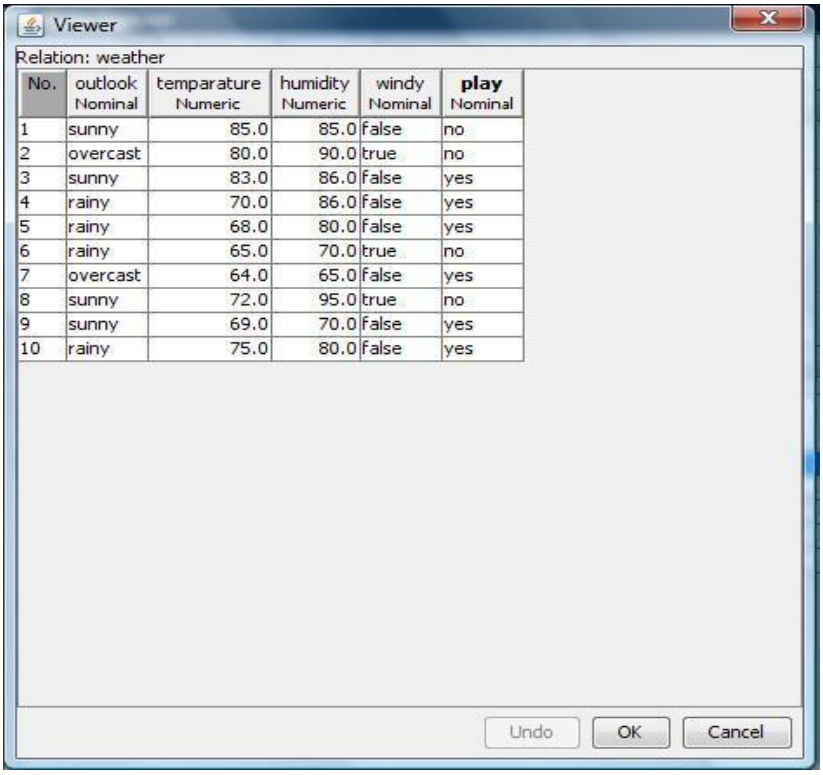
1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Weather Table.

```
@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.

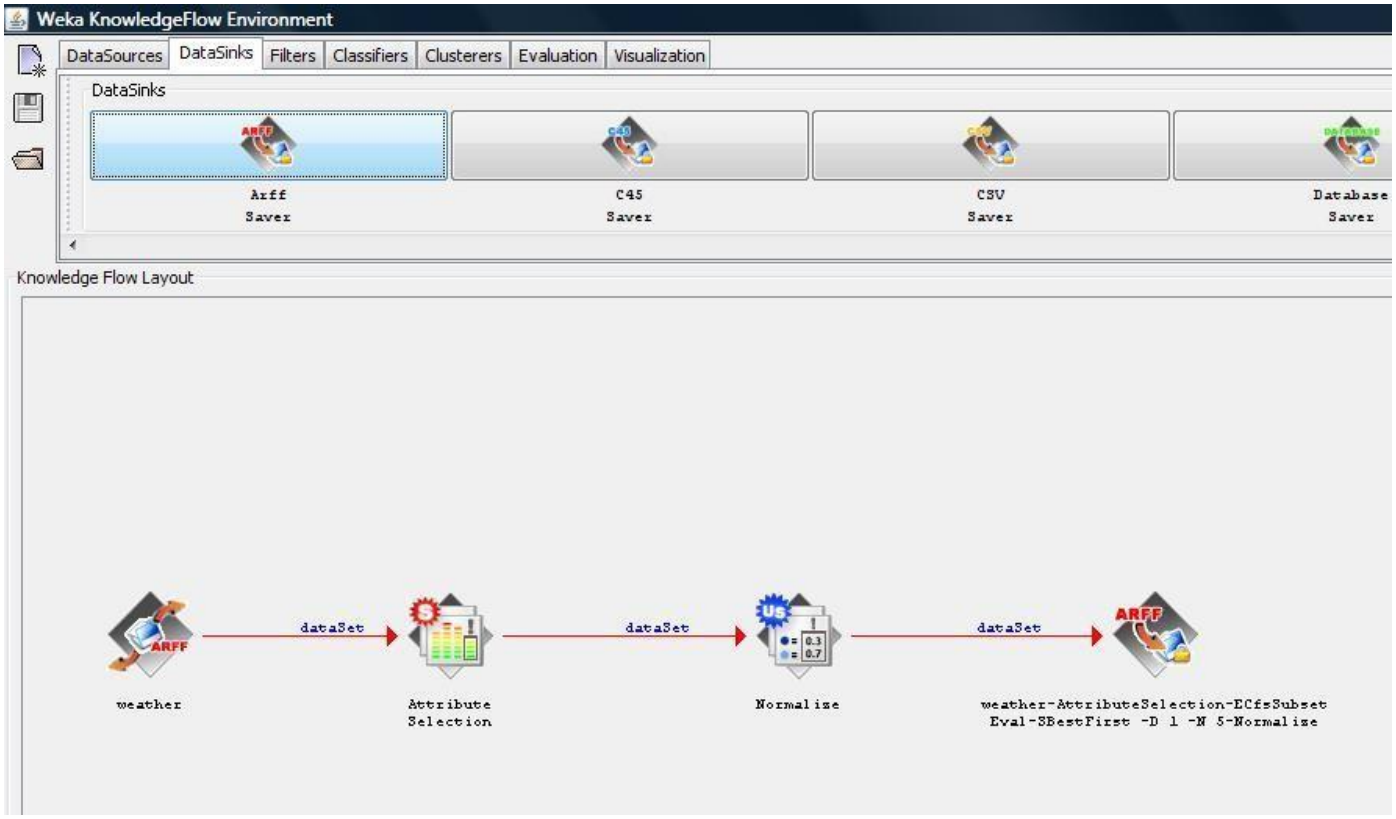
Output:

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:6

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.

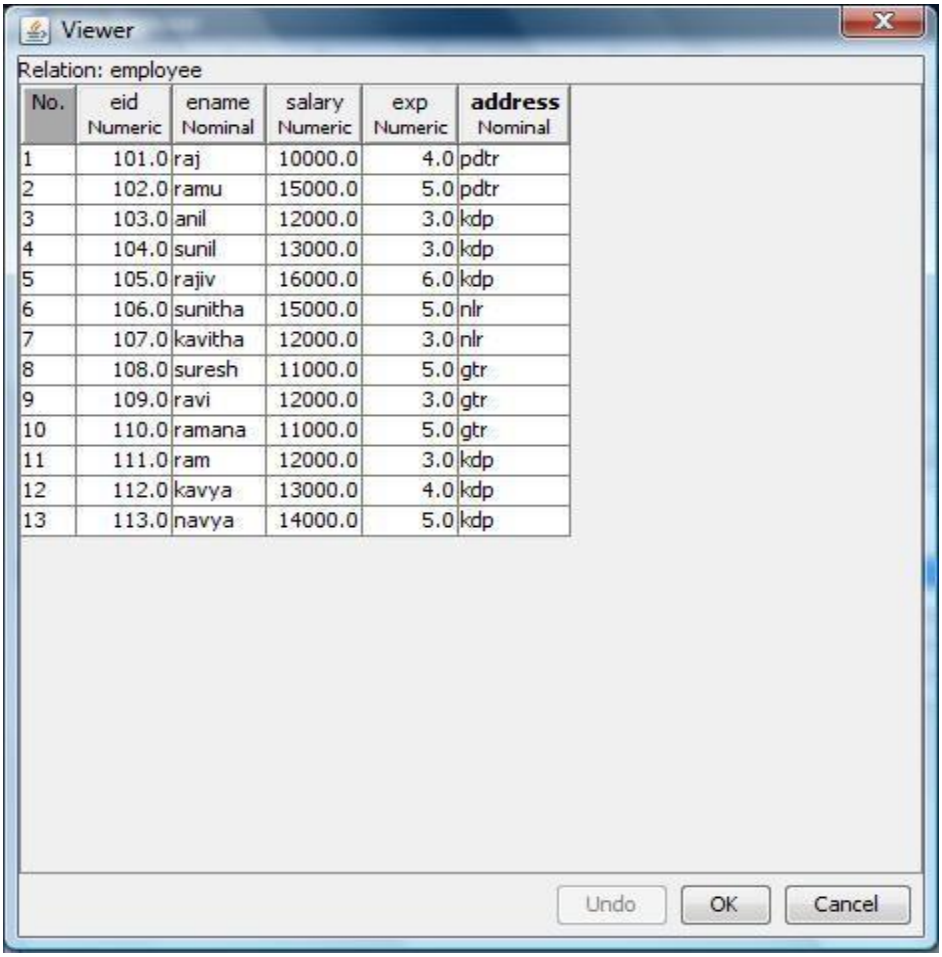
6) In that dialog box there are four modes, click on **explorer**.

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



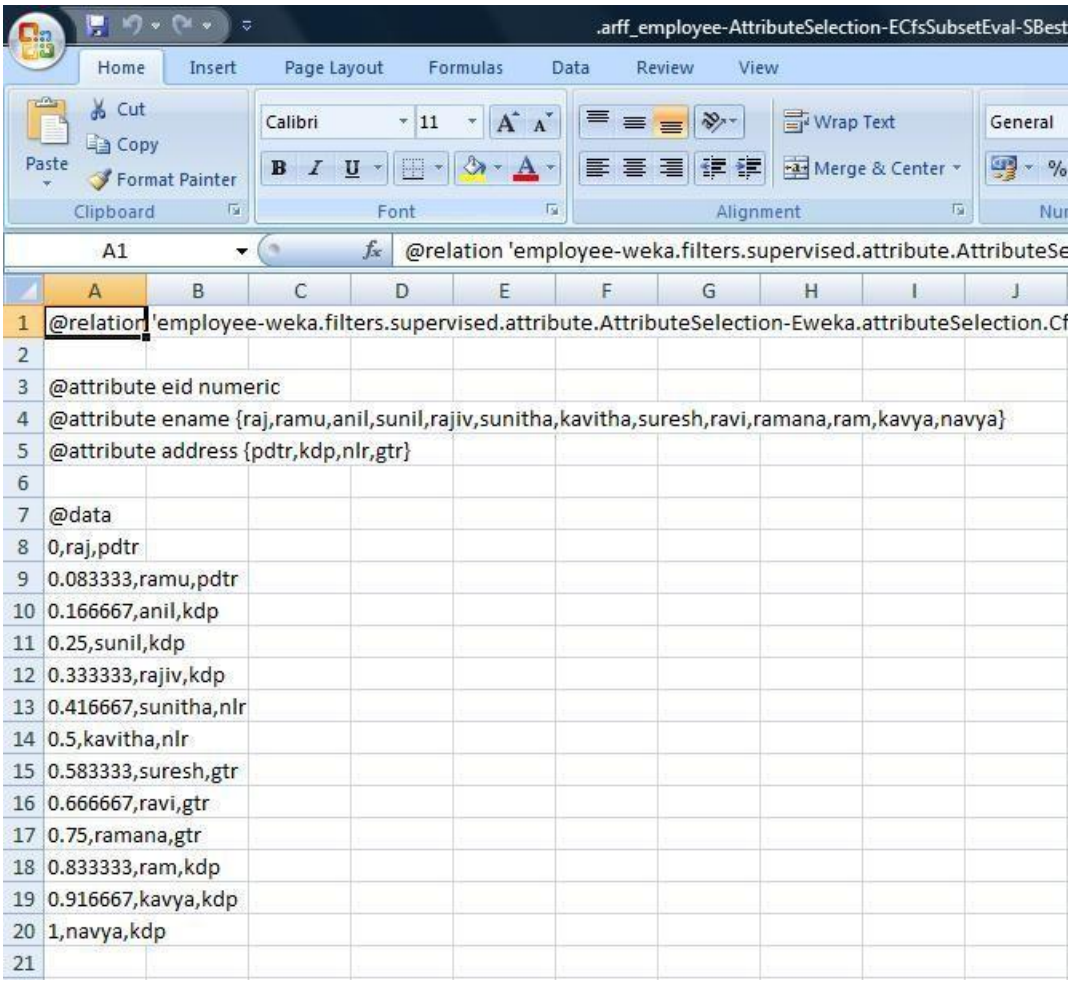
Procedure for Knowledge Flow:

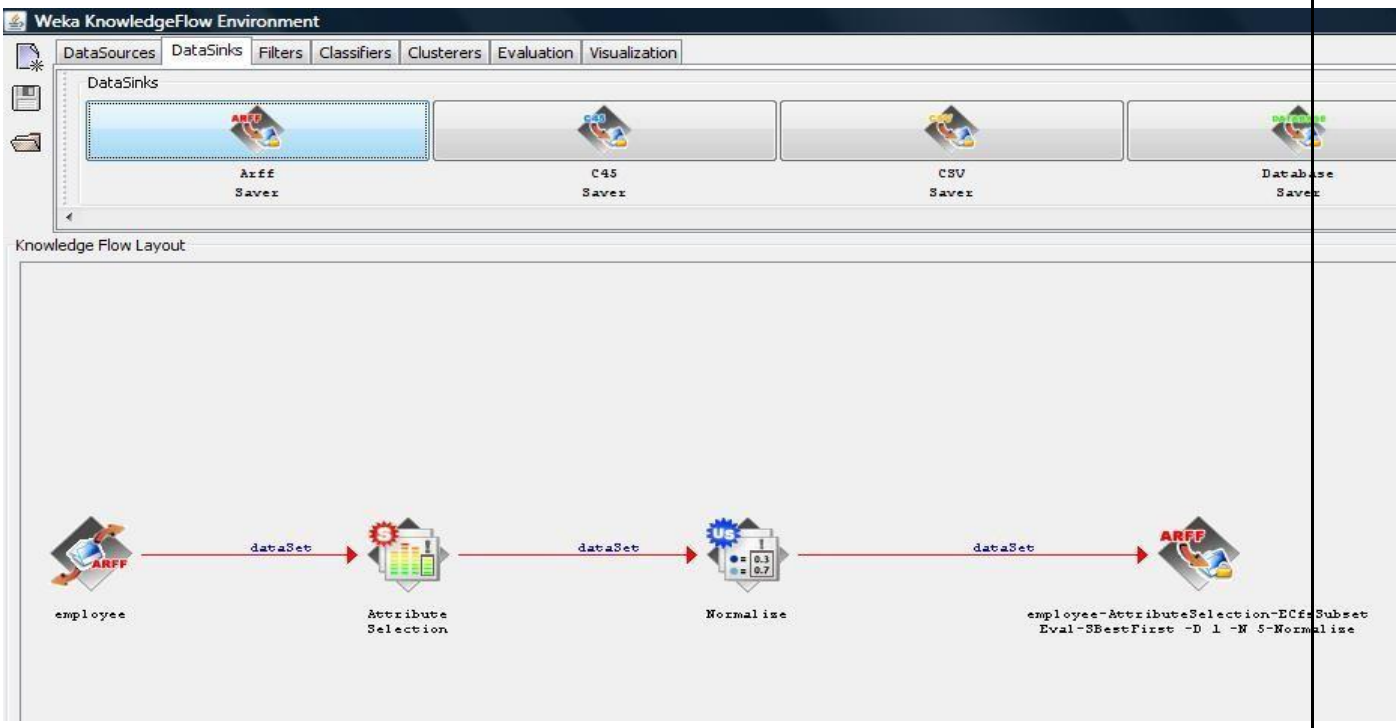
1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open the **Knowledge Flow**.
3) Select the **Data Source component** and **add Arff Loader** into the **knowledge layout canvas**.
4) Select the **Filters component** and **add Attribute Selection** and N**ormalize** into the knowledge layout canvas.
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:7

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:

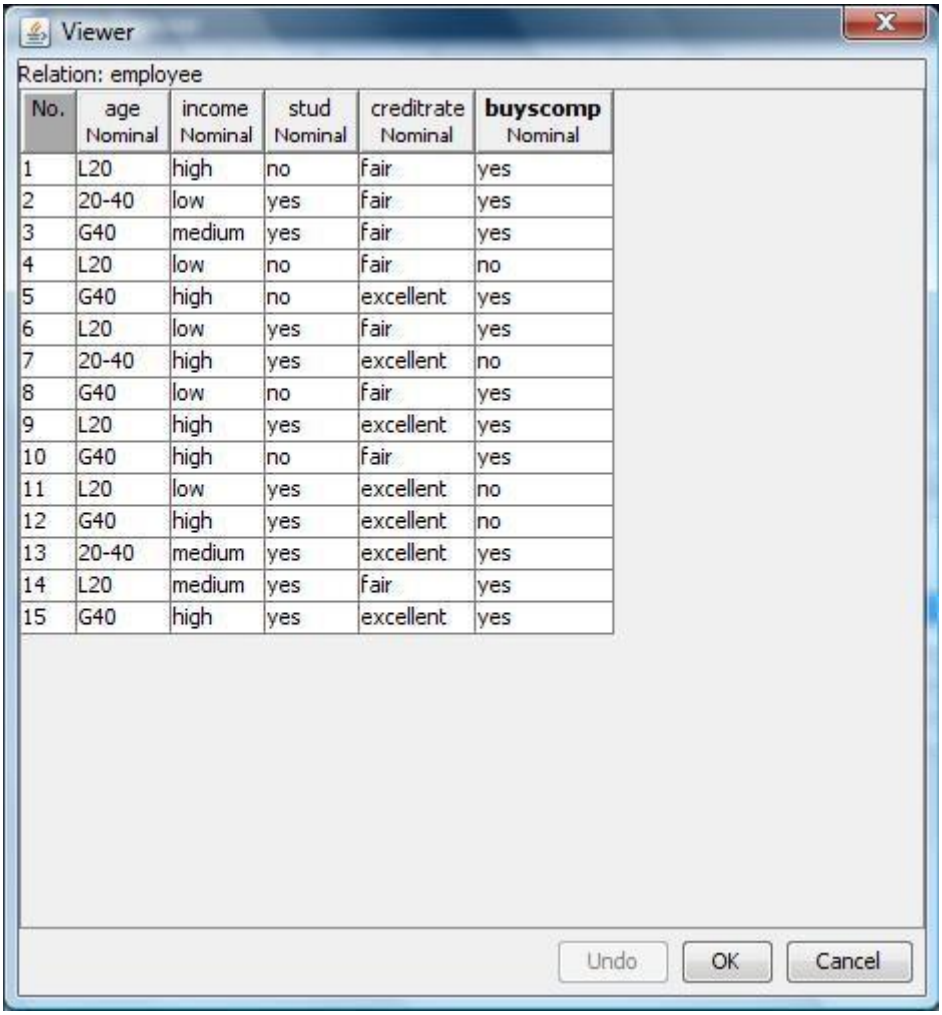
1) Open Start → Programs → Accessories → Notepad
2) Type the following training data set with the help of Notepad for Buying Table.

```
@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.

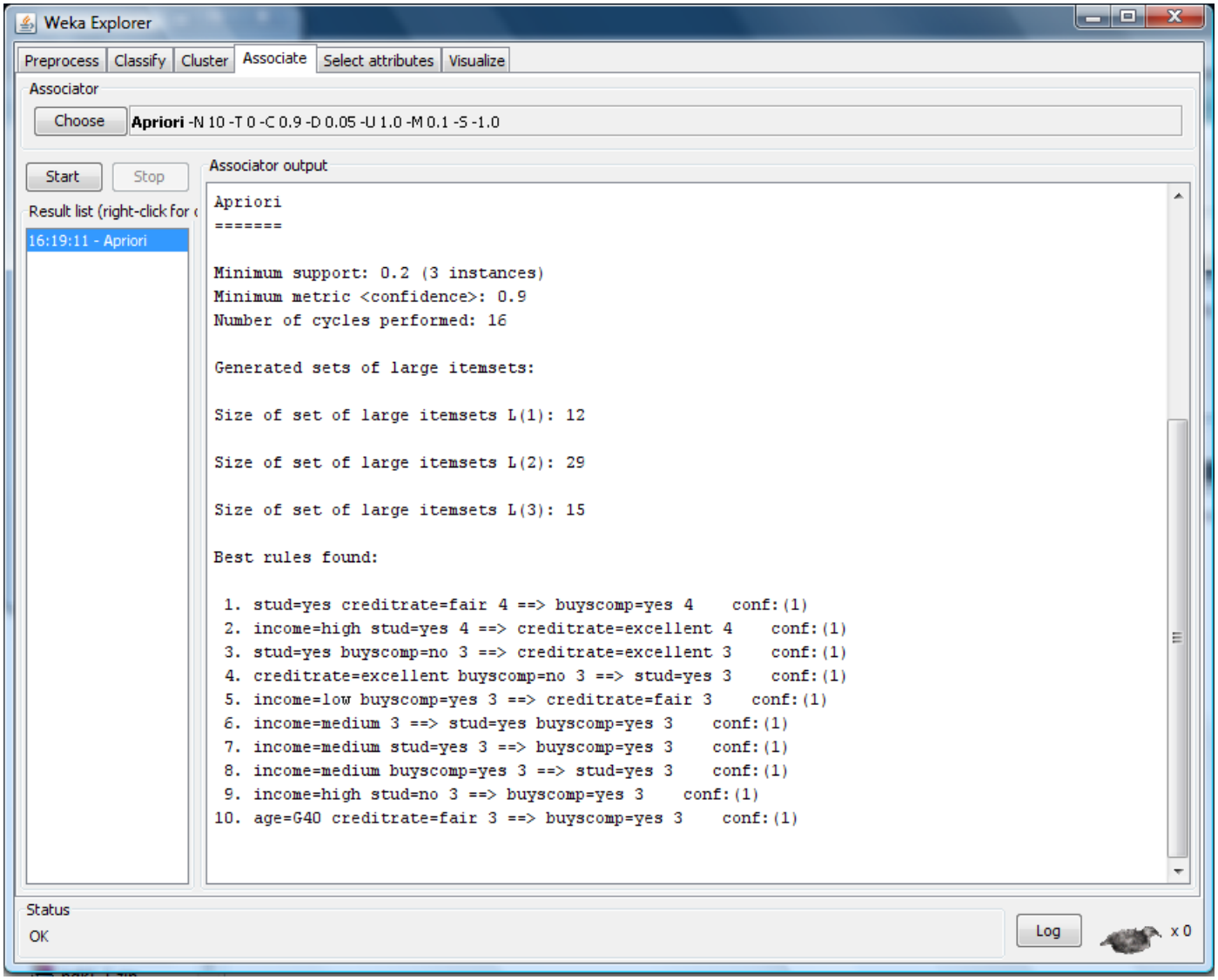
Output:

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:8

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:

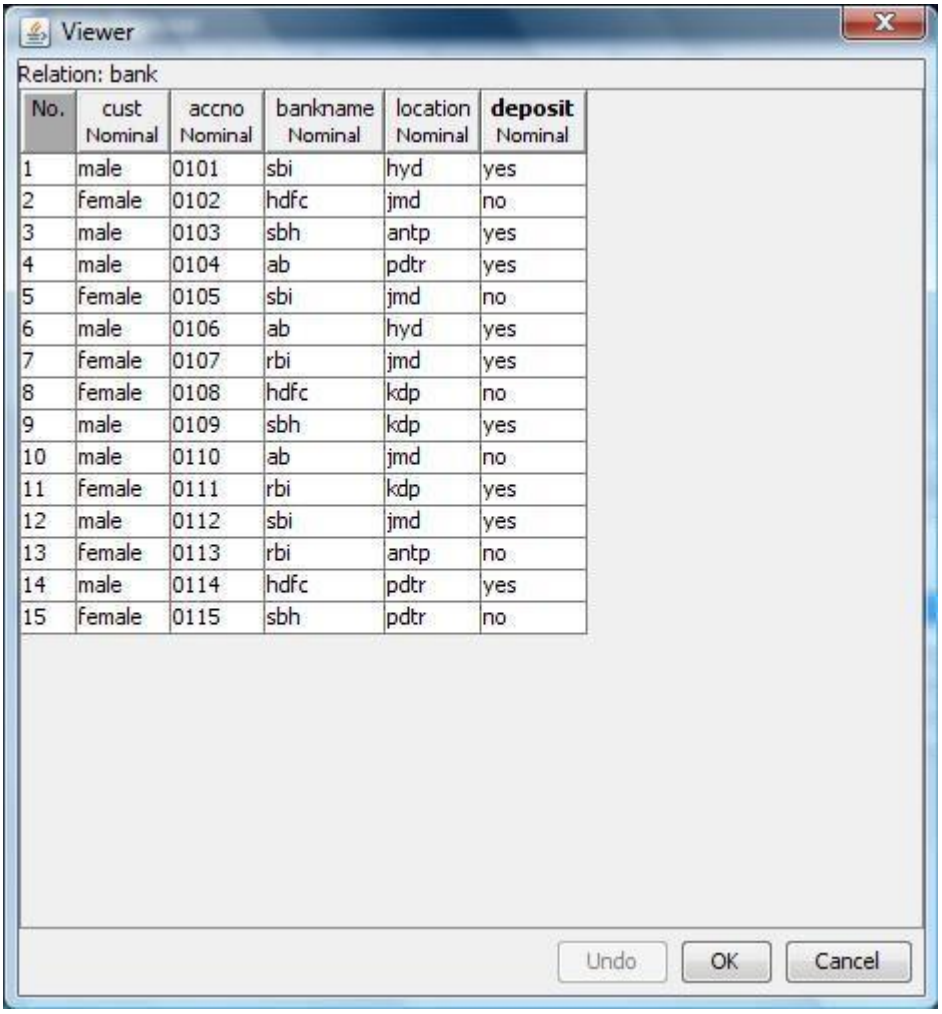
Procedure:

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.

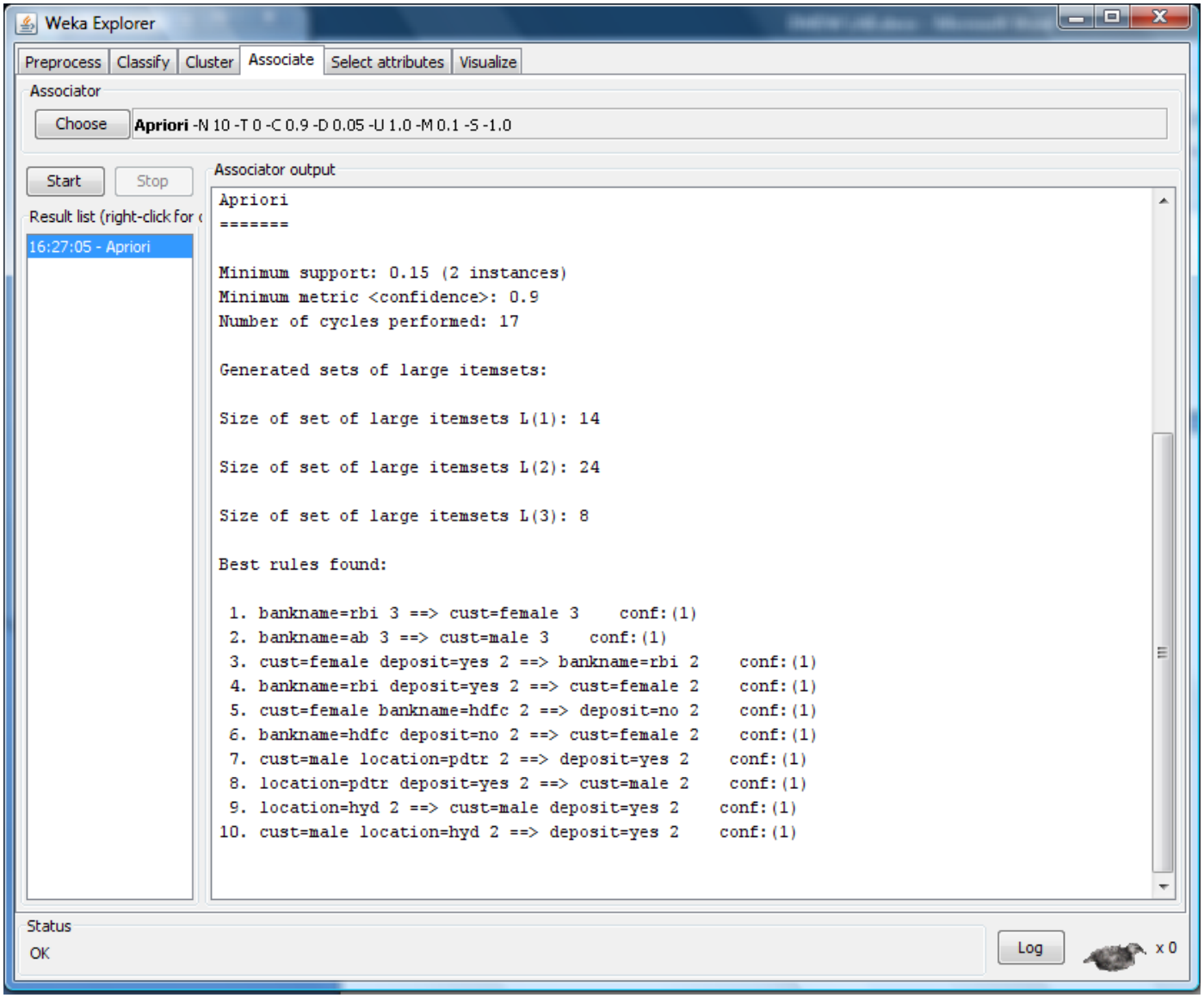
Training Data Set → Banking Table



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:9

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.

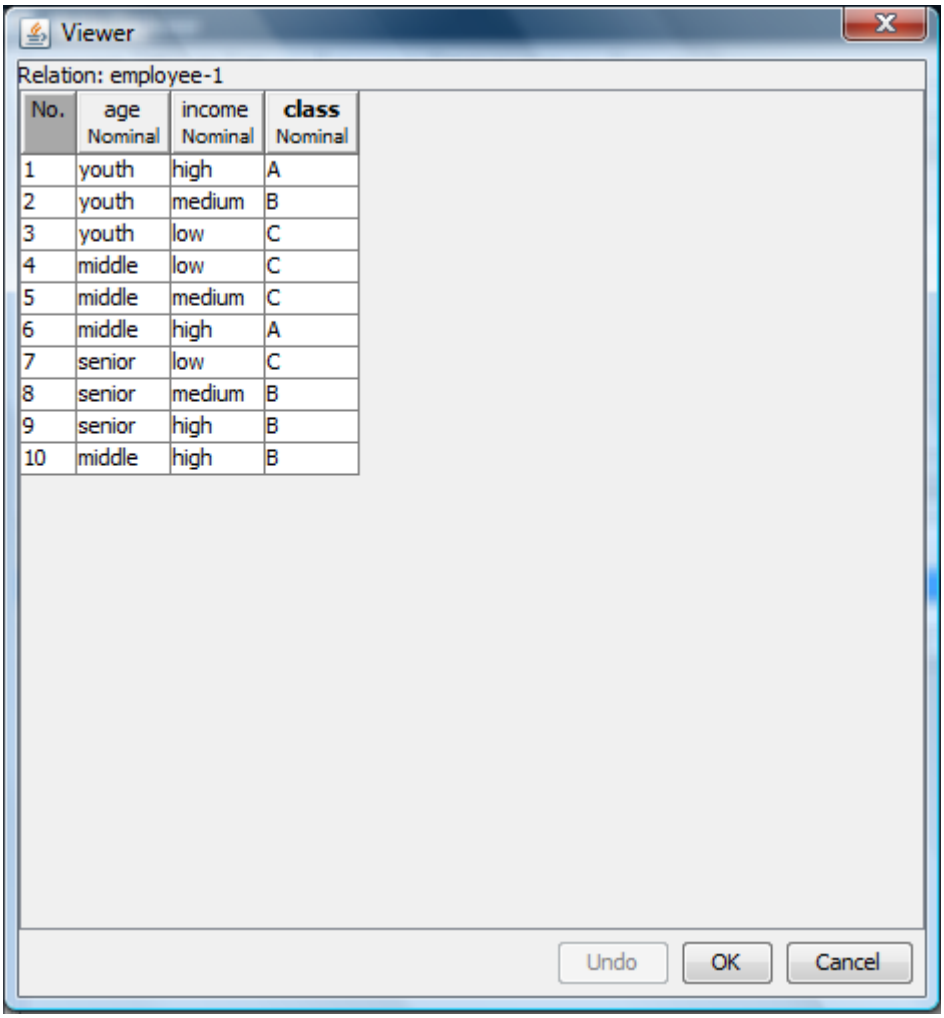
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

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.

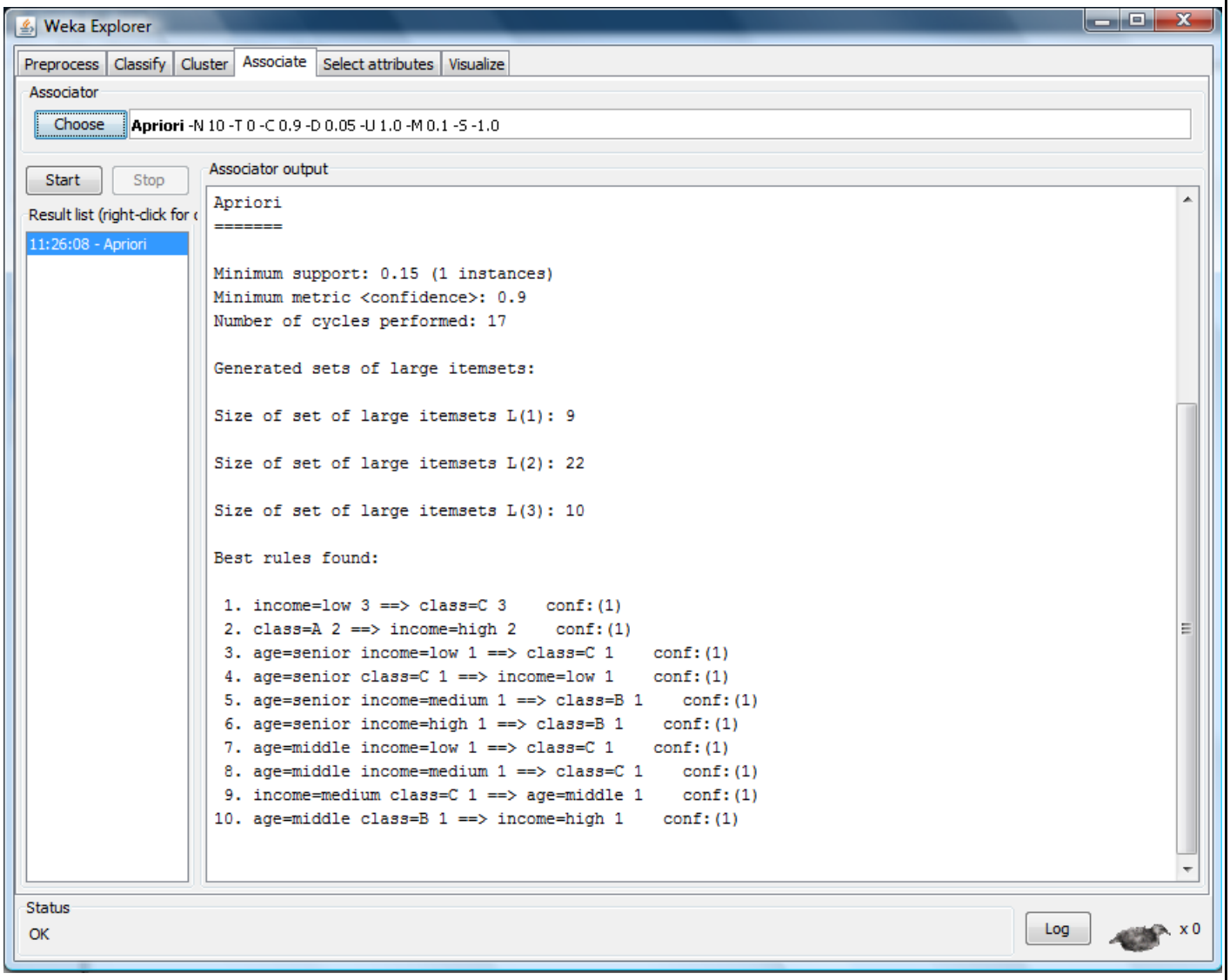
Training Data Set → Employee Table



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:10

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.

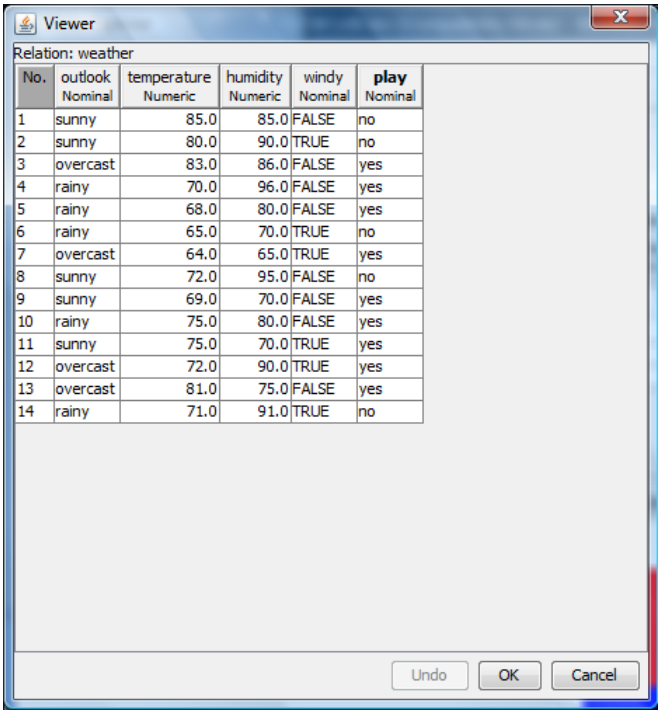
```
@relation weather
@attribute outlook {sunny, rainy, overcast}
@attribute temperature numeric
@attribute humidity numeric
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}

@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.

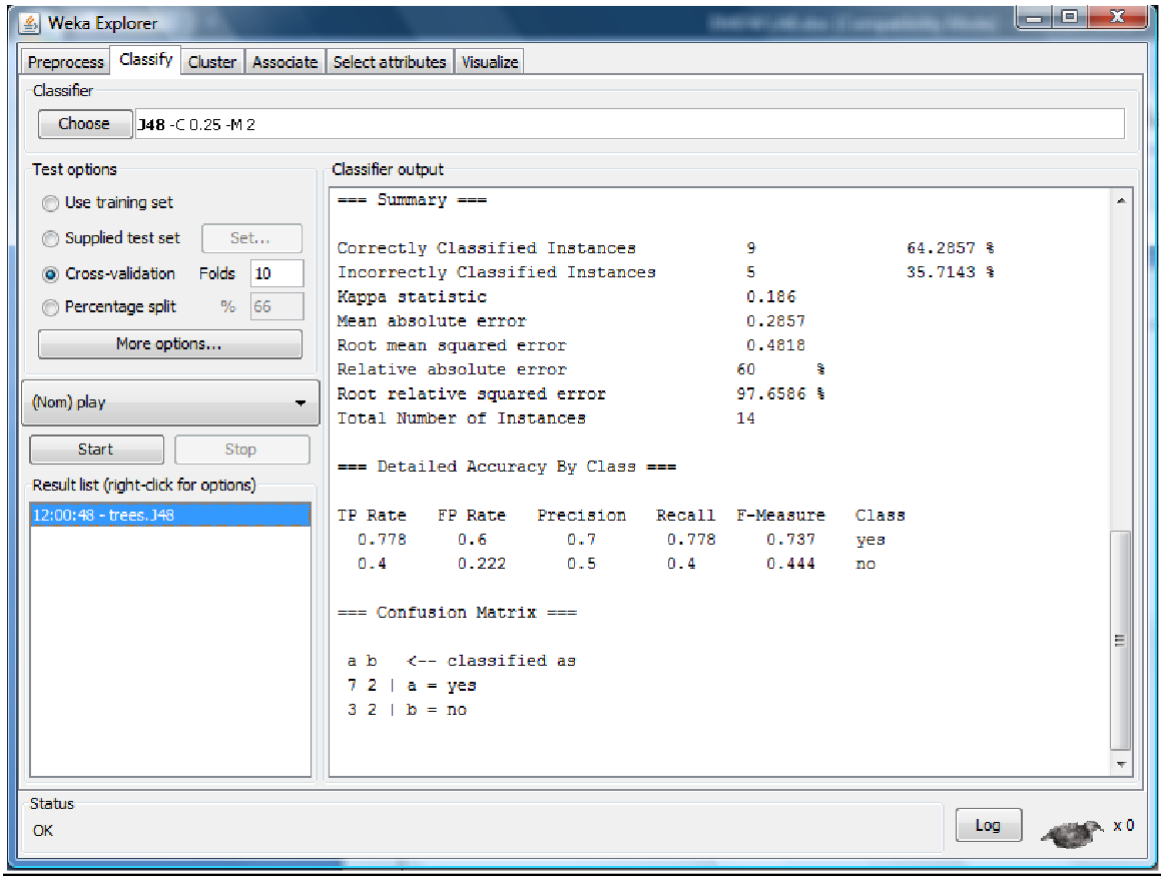
Training Data Set → Weather Table



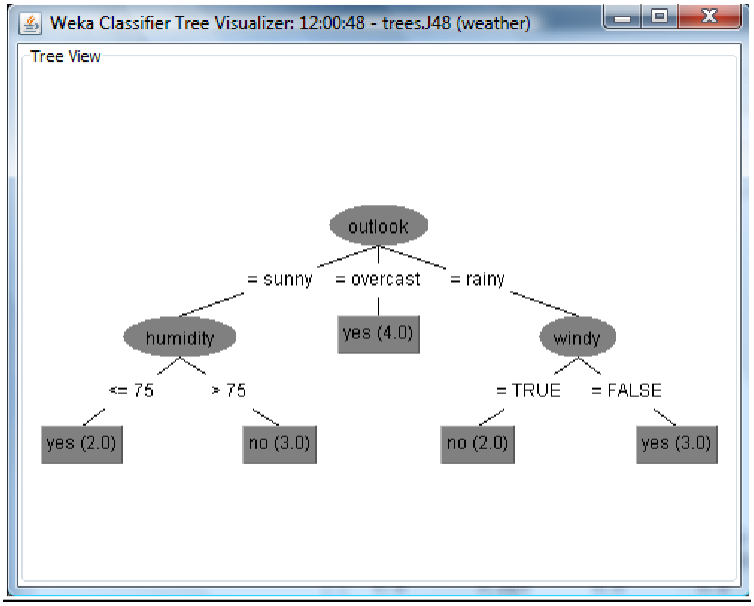
Procedure for Decision Trees:

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.

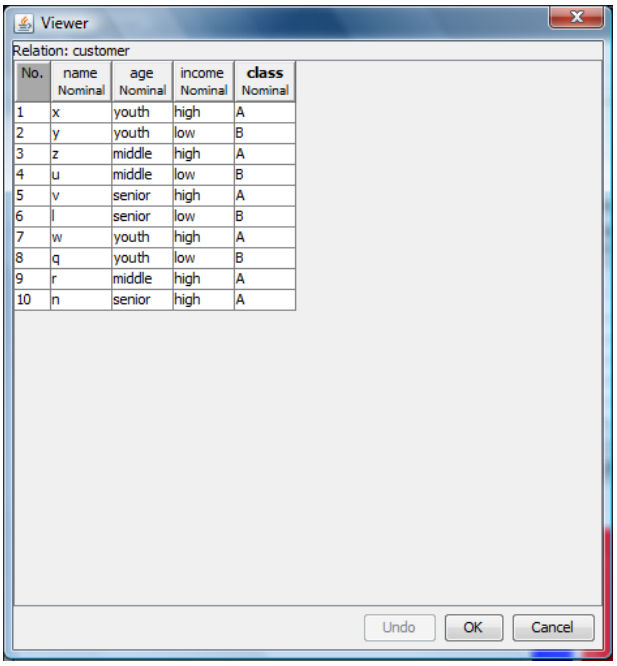
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

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.

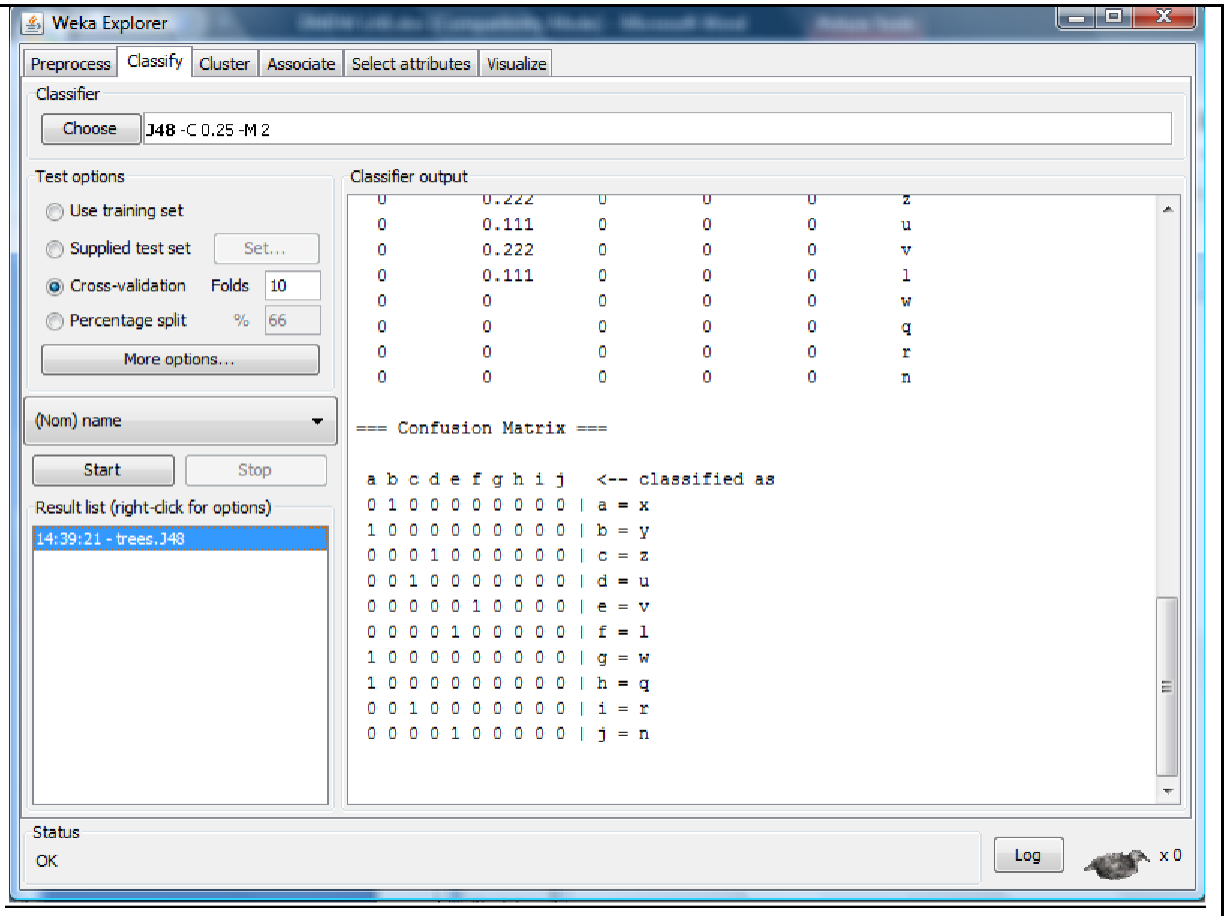
Training Data Set → Customer Table



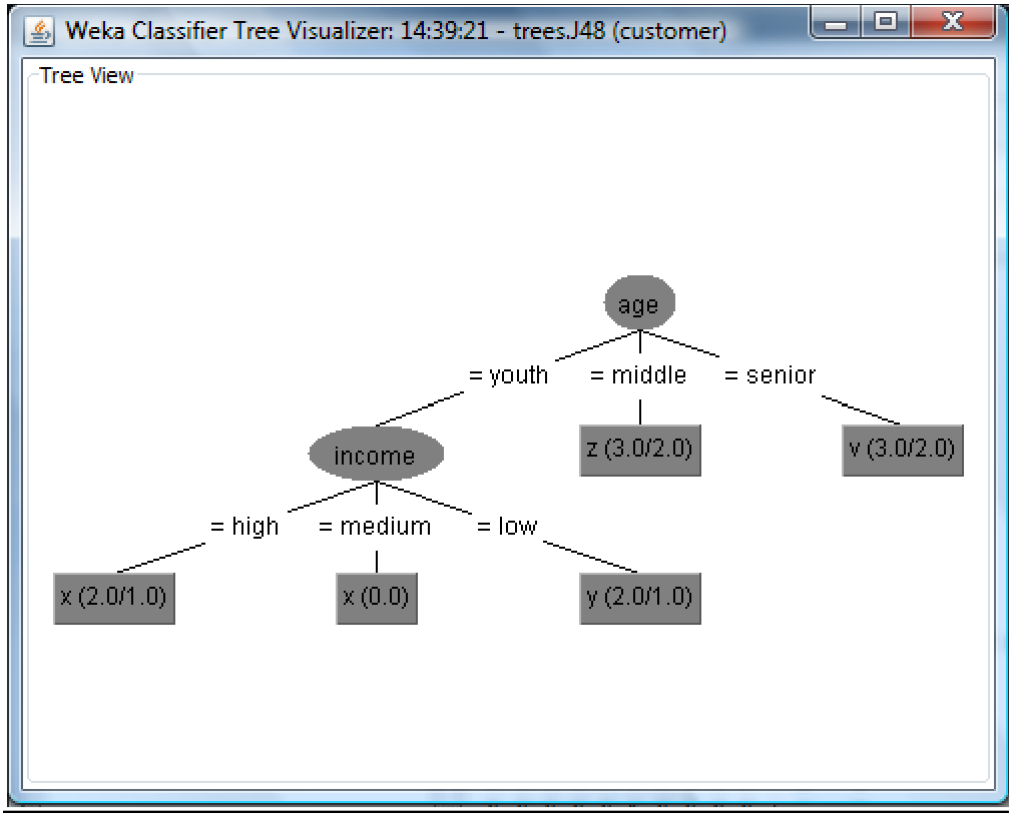
Procedure for Decision Trees:

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:12

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:

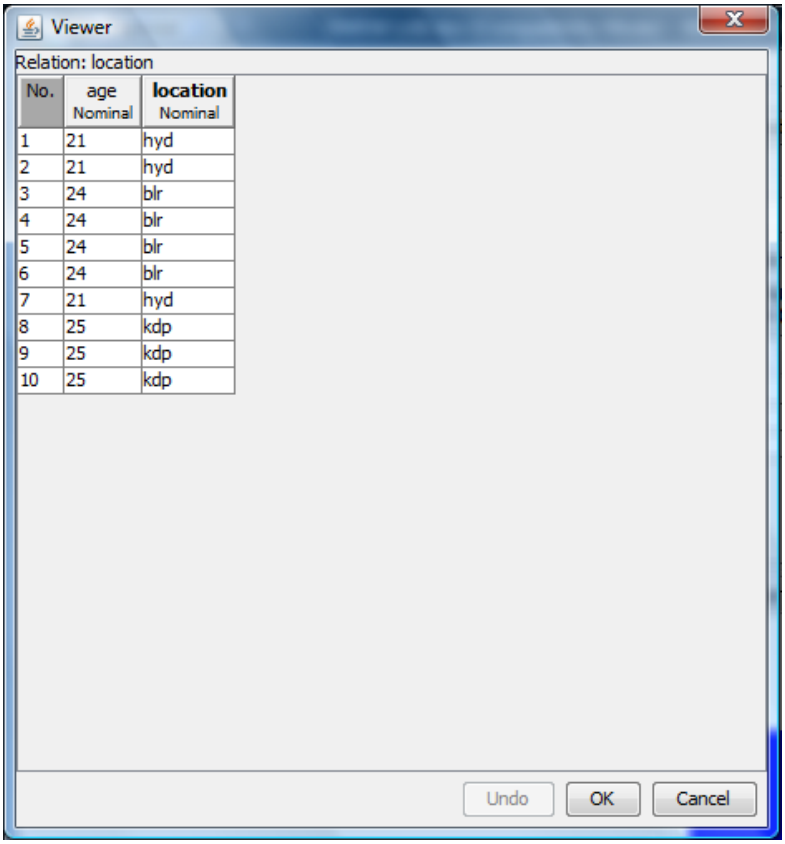
Procedure:

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.

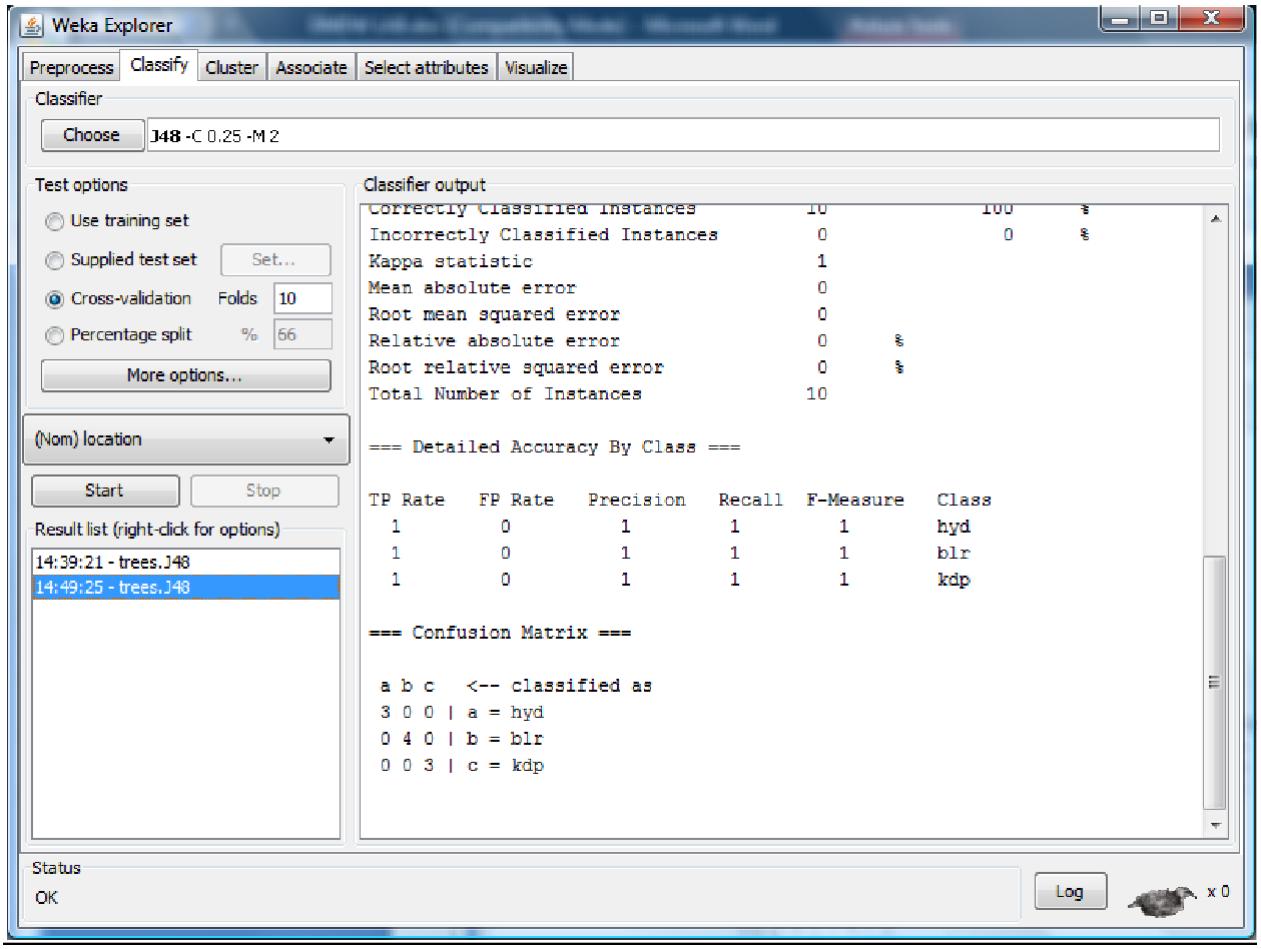
Training Data Set → Location Table



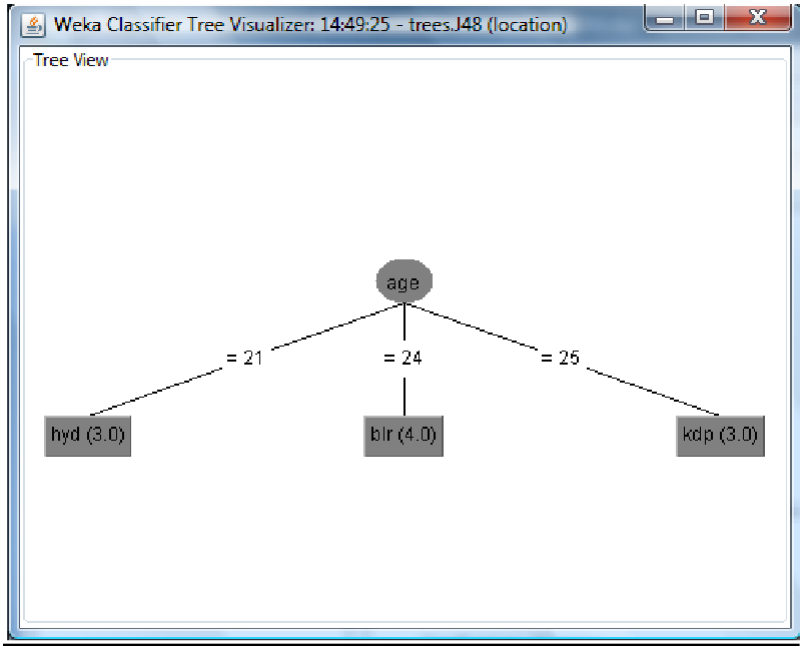
Procedure for Decision Trees:

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.

EXPERIMENT NO:13

Aim:

Write a procedure for Visualization for Weather Table.

Description:

This program calculates and has comparisons on the data set selection of attributes and methods of manipulations have been chosen. The Visualization can be shown in a 2-D representation of the information.

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

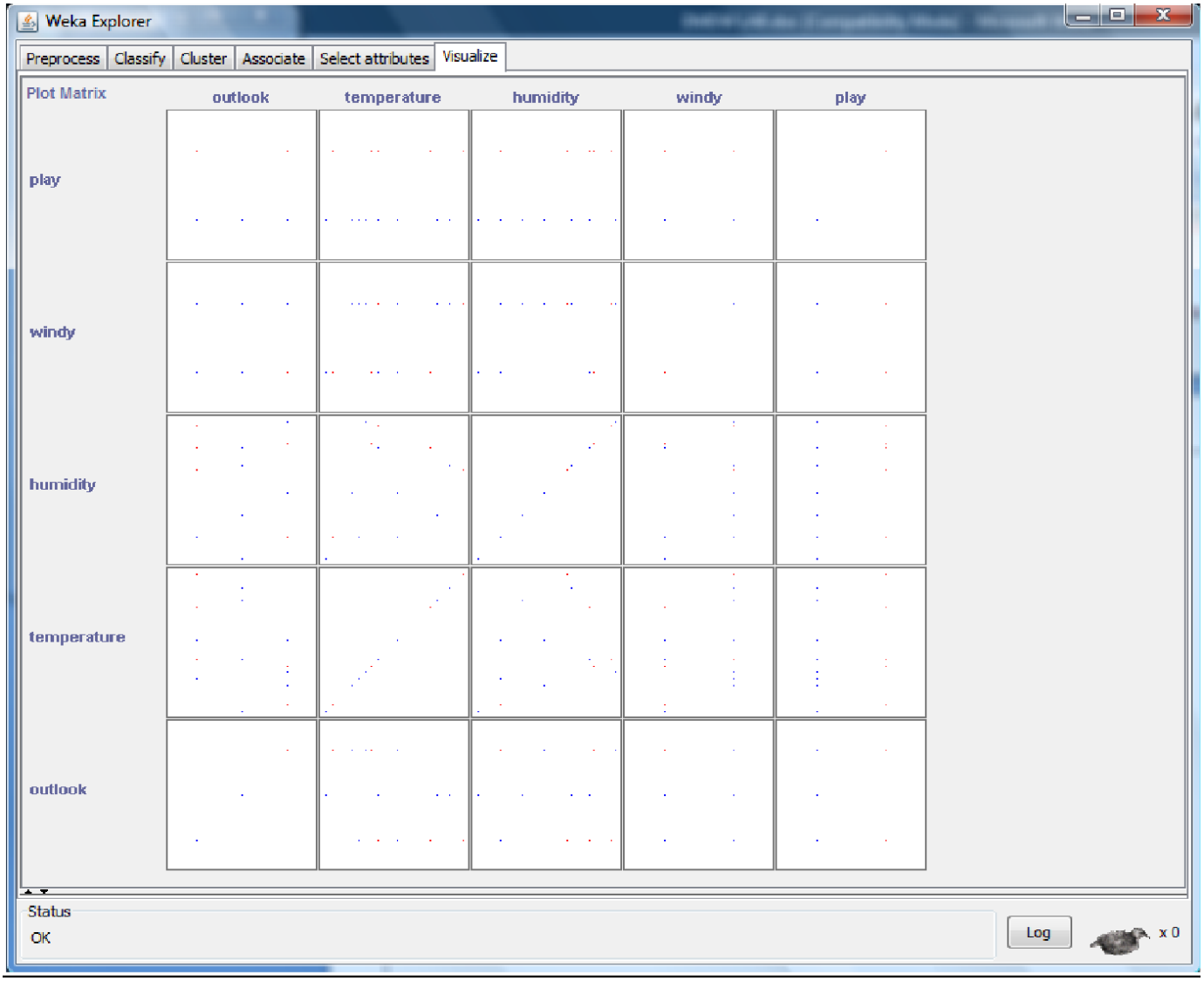
Viewer

Relation: weather

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

Undo OK Cancel

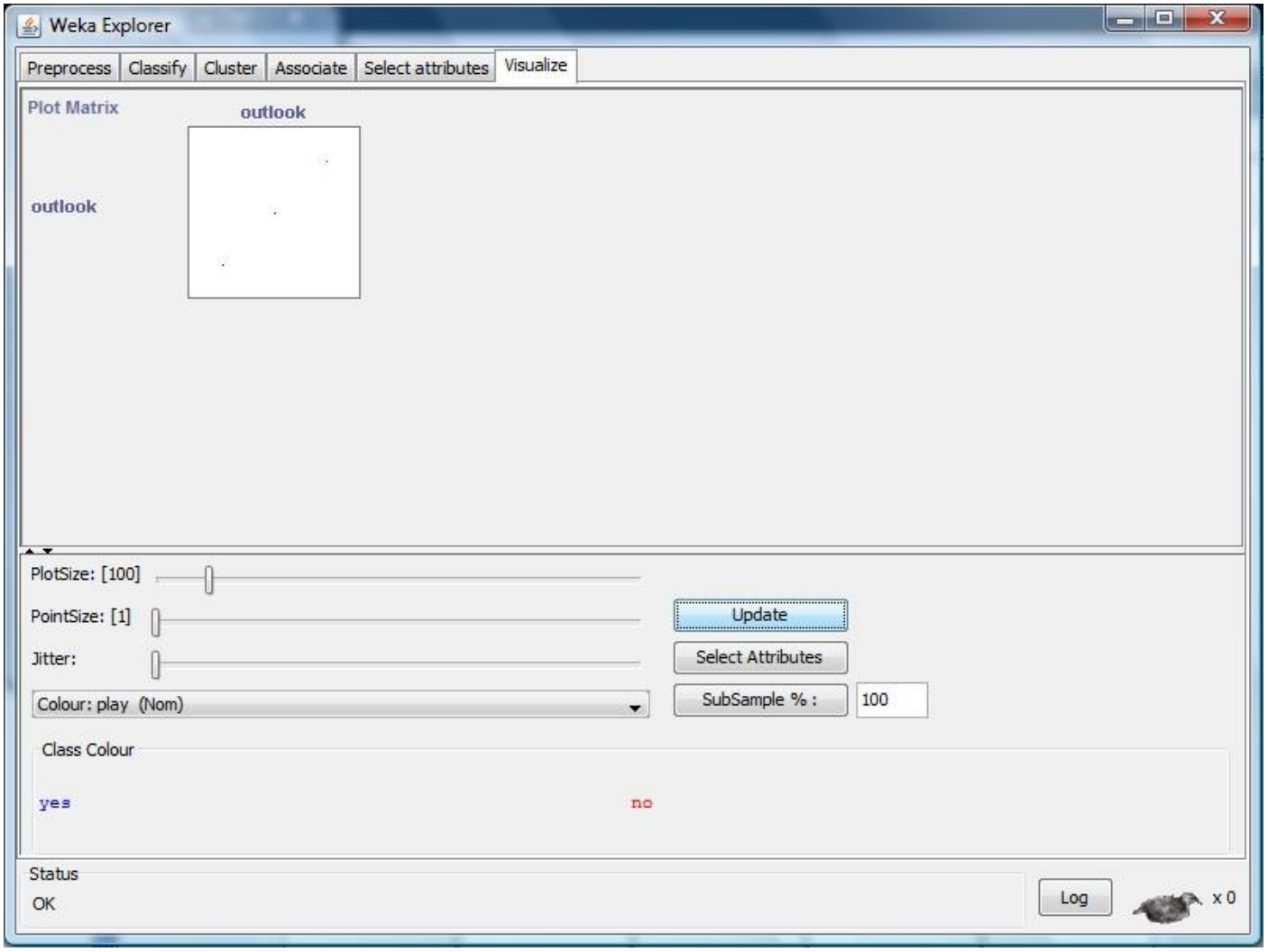
Plot Matrix:



Procedure:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open the explorer and click on **Preprocess**, then a new window will appear. In that window select **weather.arff** file then the data will be displayed.
3) After that click on the **Visualize tab** on the top of the Menu bar.
4) When we select **Visualize tab** then **Plot Matrix** is displayed on the screen.

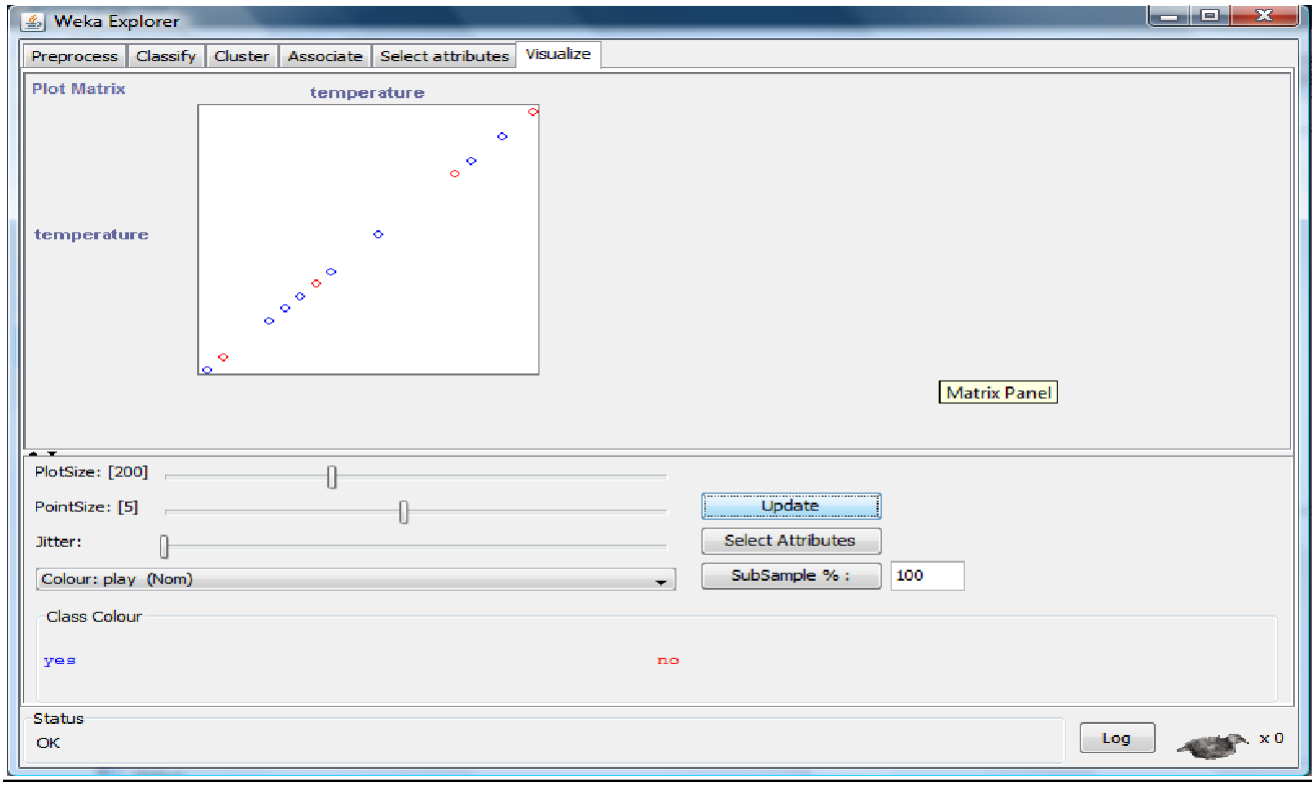
Output:



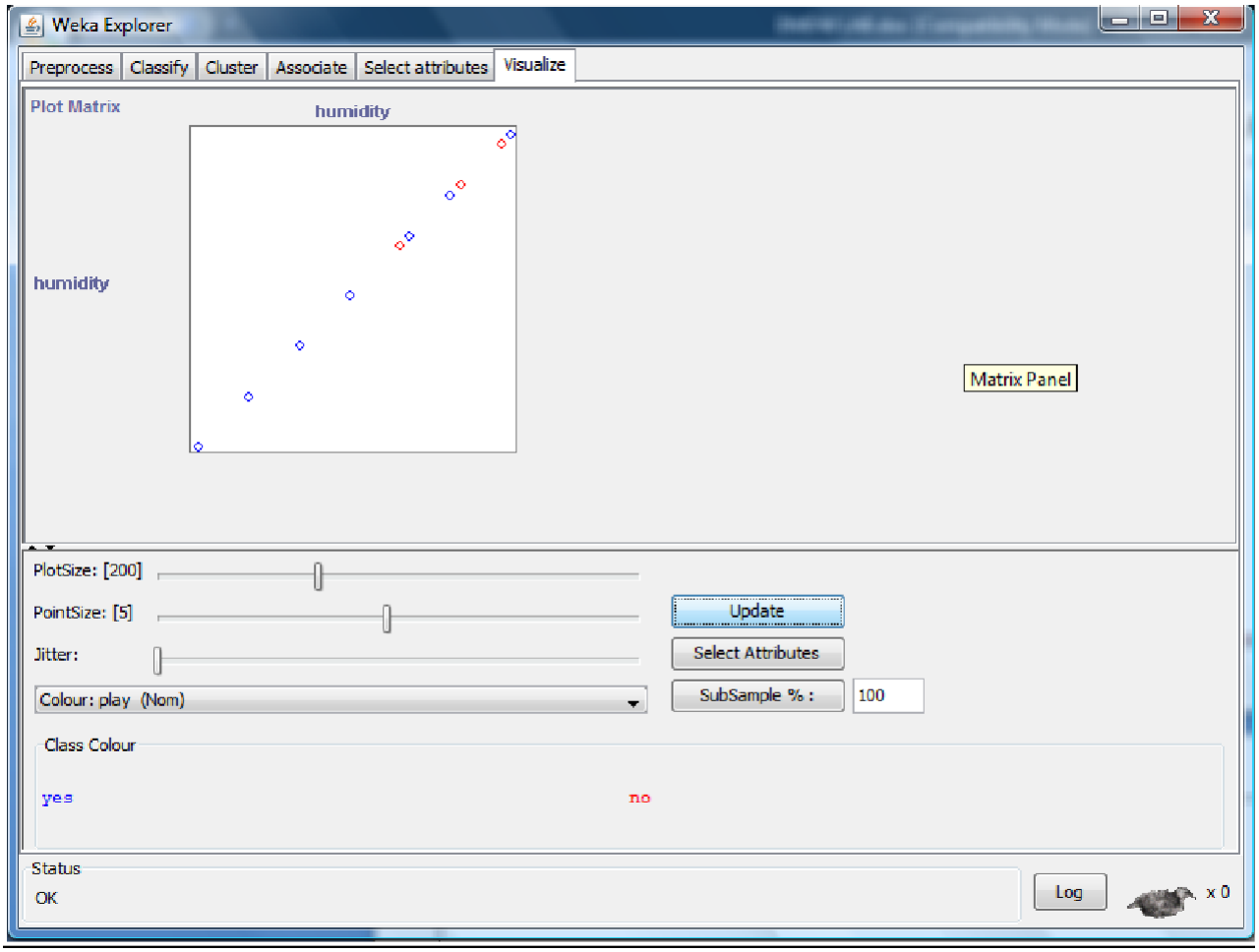
5) After that we select the **Select Attribute button**, then select **Outlook attribute** and clock OK.
6) Click on the **Update button** to display the output.
7) After that select the **Select Attribute button** and select **Temperature attribute** and then click OK.
8) **Increase** the **Plot Size** and **Point Size**.
9) Click on the **Update button** to display the output.
10) After that we select the **Select Attribute button**, then select **Humidity attribute** and clock OK.
11) Click on the **Update button** to display the output.
12) After that select the **Select Attribute button** and select **Windy attribute** and then click OK.
13) **Increase** the **Jitter Size**.
14) Click on the **Update button** to display the output.
15) After that we select the **Select Attribute button**, then select **Play attribute** and clock OK.

16) Click on the **Update button** to display the output.

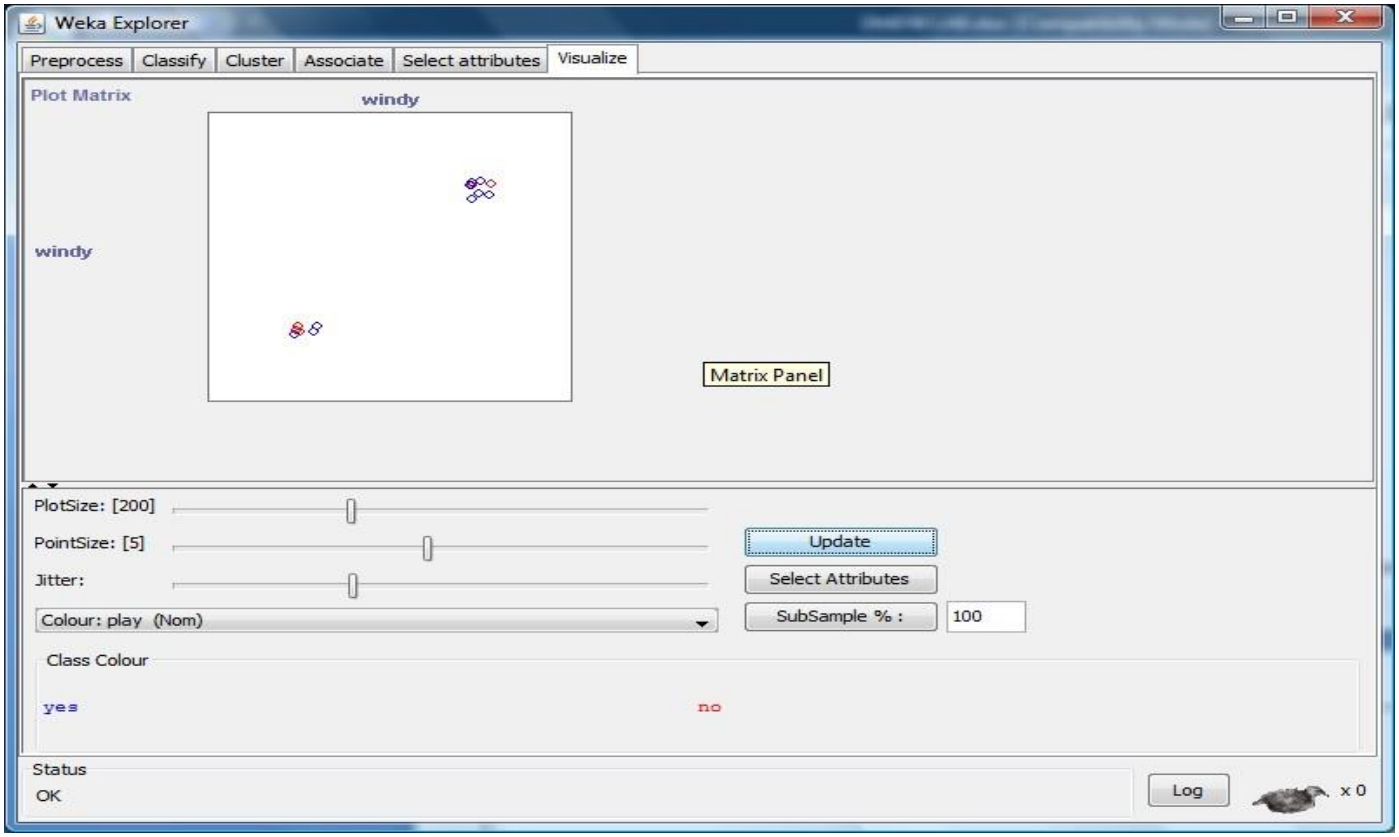
Output:



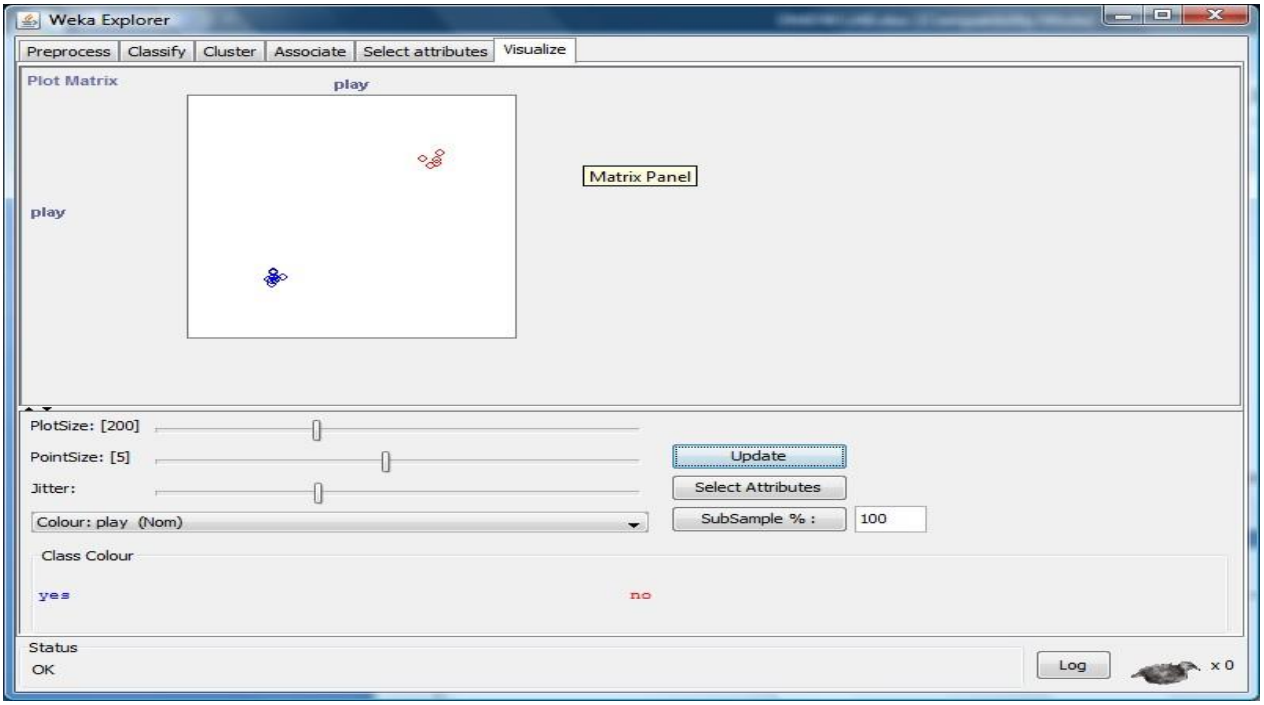
Output:



Output:



Output:



Result:

This program has been successfully executed.

EXPERIMENT NO:14

Aim:

Write a procedure for Visualization of Banking Table.

Description:

This program calculates and has comparisons on the data set selection of attributes and methods of manipulations have been chosen. The Visualization can be shown in a 2-D representation of the information.

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.

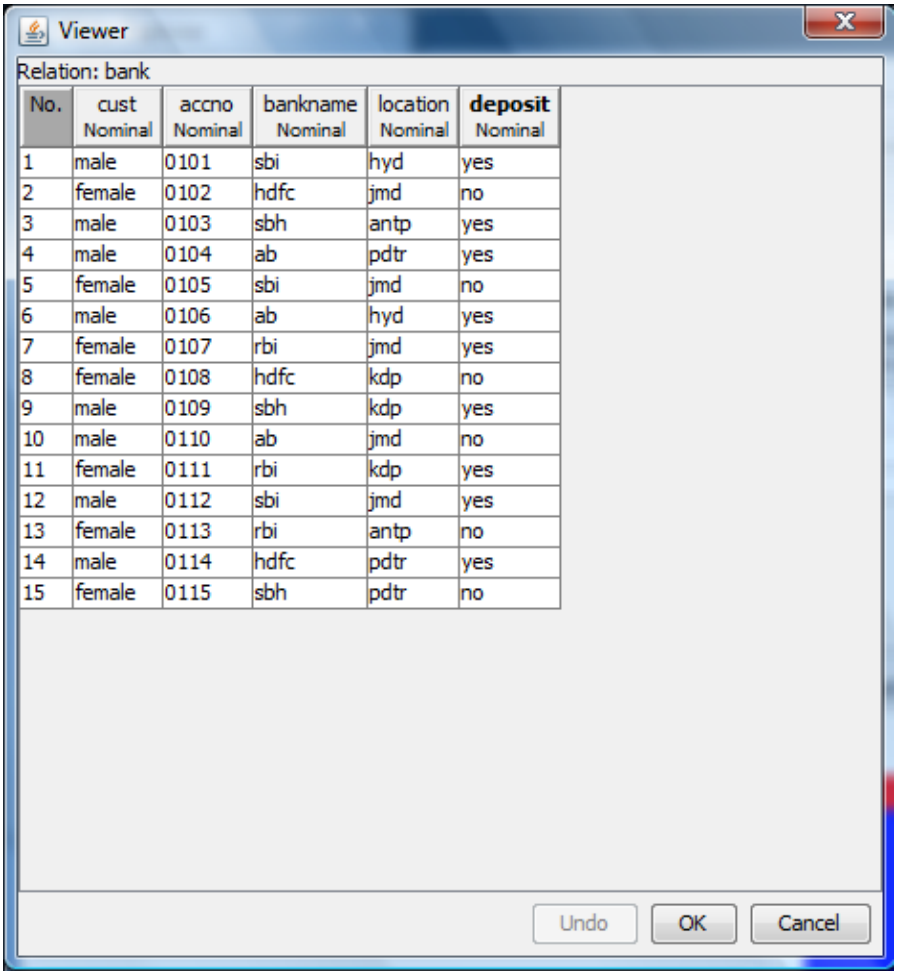
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

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.

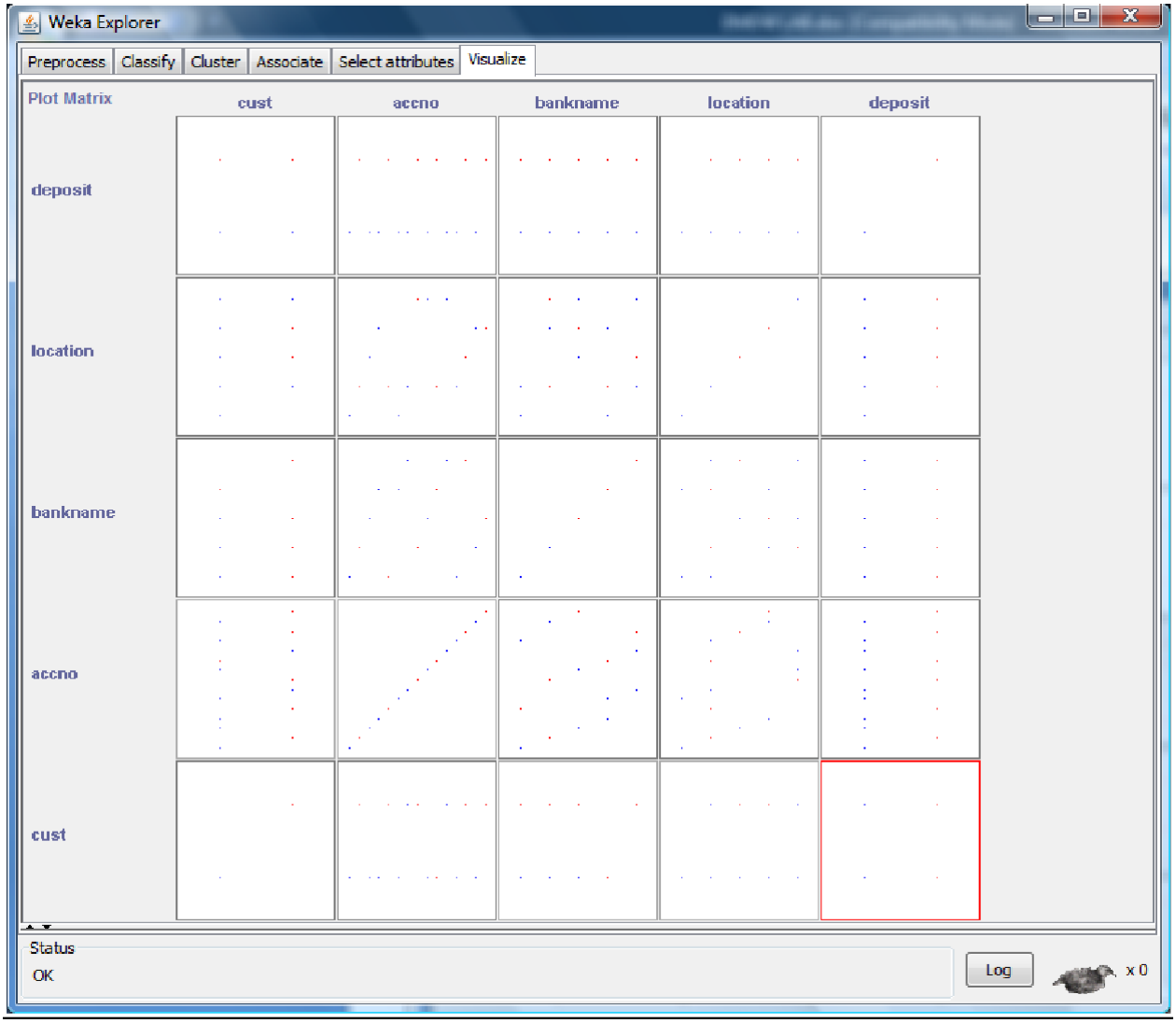
Training Data Set → Banking Table



Procedure:

1) Open Start → Programs → Weka-3-4 → Weka-3-4
2) Open the explorer and click on **Preprocess**, then a new window will appear. In that window select **bank.arff** file then the data will be displayed.
3) After that click on the **Visualize tab** on the top of the Menu bar.
4) When we select **Visualize tab** then **Plot Matrix** is displayed on the screen.

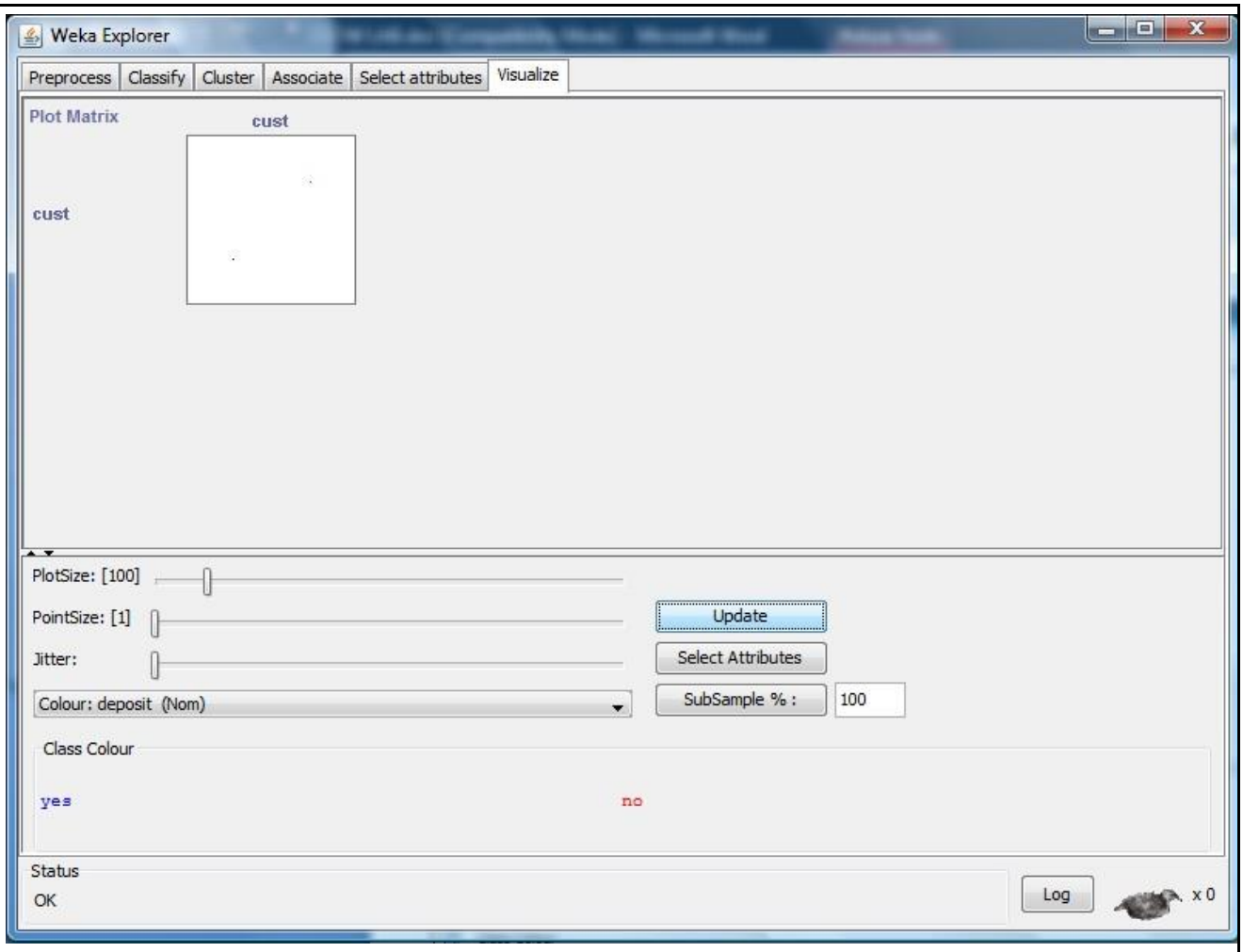
2-D Plot Matrix:



5) After that we select the **Select Attribute button**, then select **Cust attribute** and clock OK.

6) Click on the **Update button** to display the output.

7) **Output:**

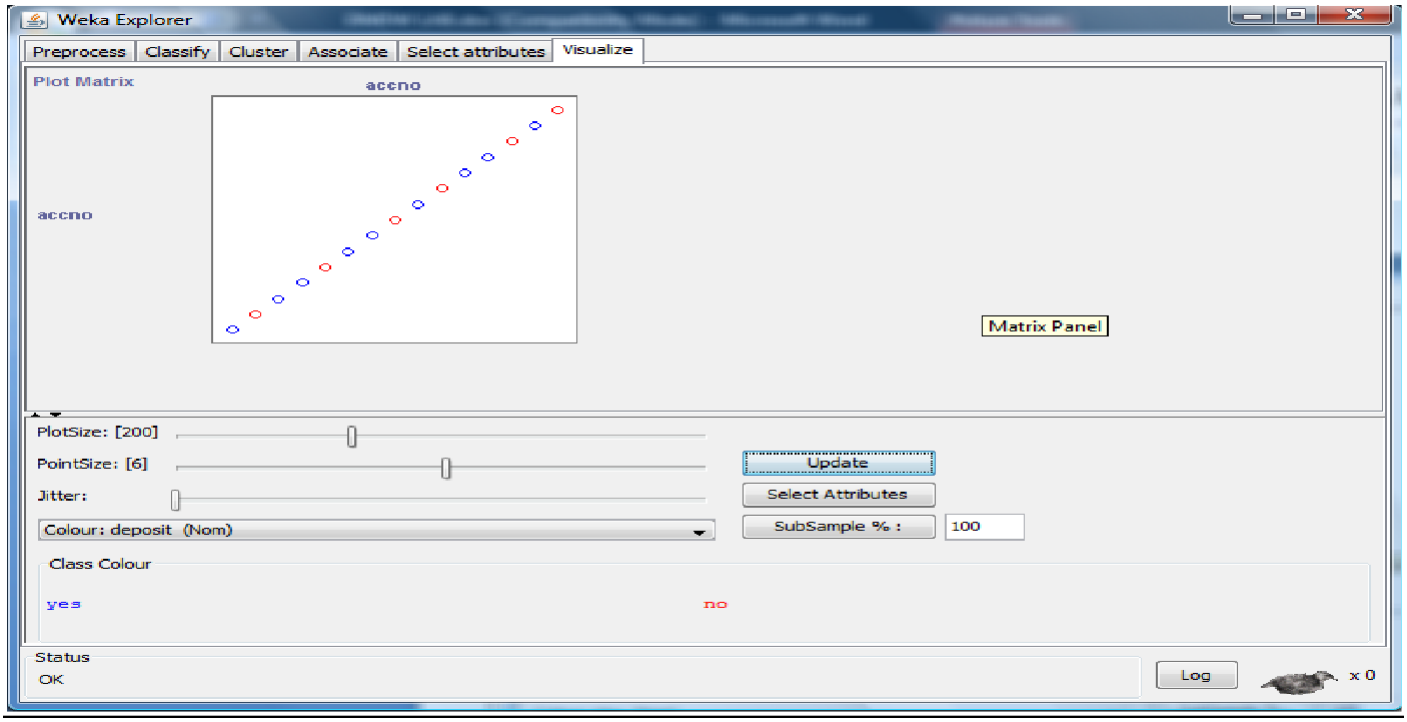


8) After that select the **Select Attribute button** and select **Accno attribute** and then click OK.

9) **Increase** the **Plot Size** and **Point Size**.

10) Click on the **Update button** to display the output.

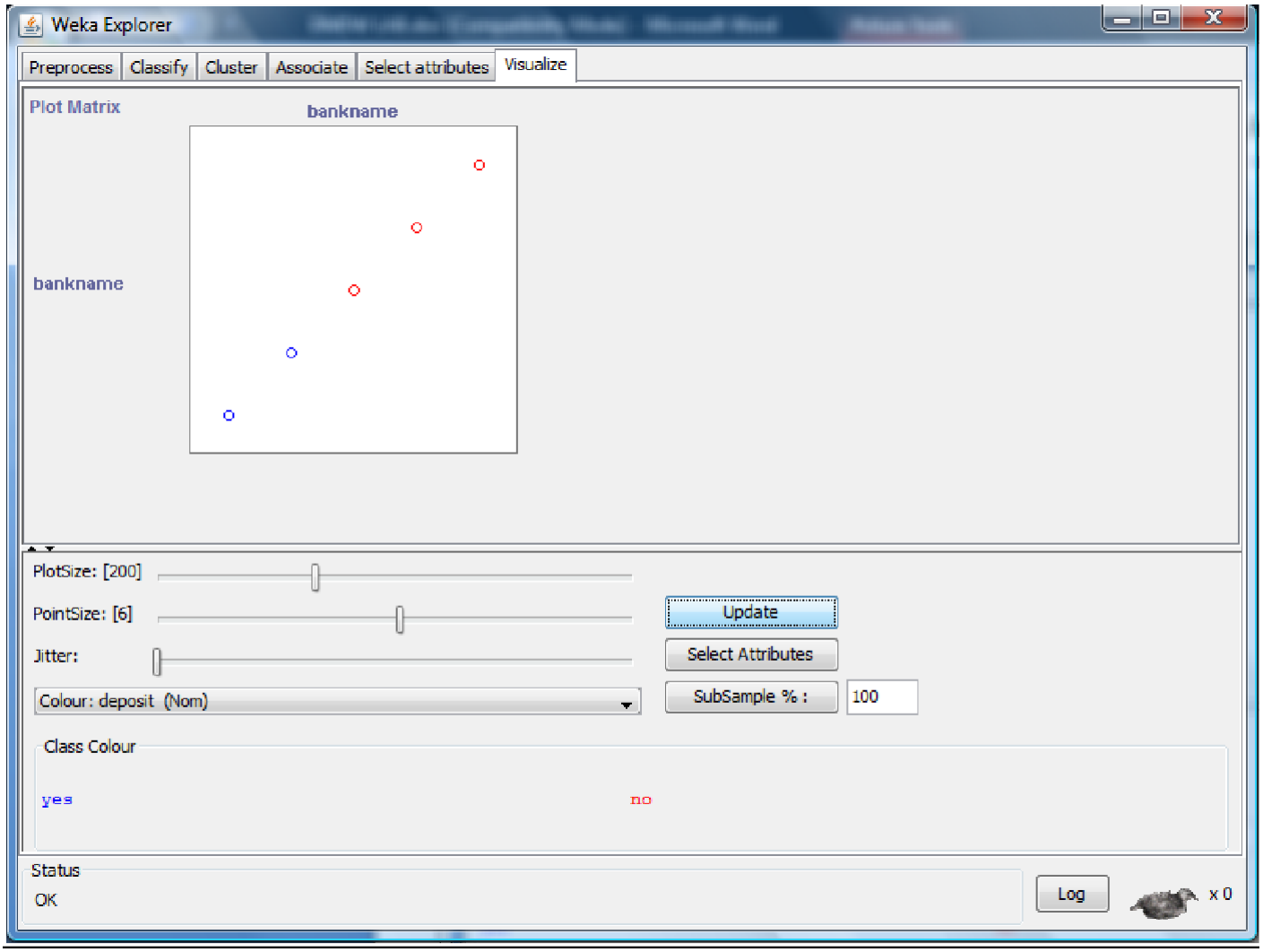
Output:



11) After that we select the **Select Attribute button**, then select **Bankname attribute** and clock OK.

12) Click on the **Update button** to display the output.

Output:

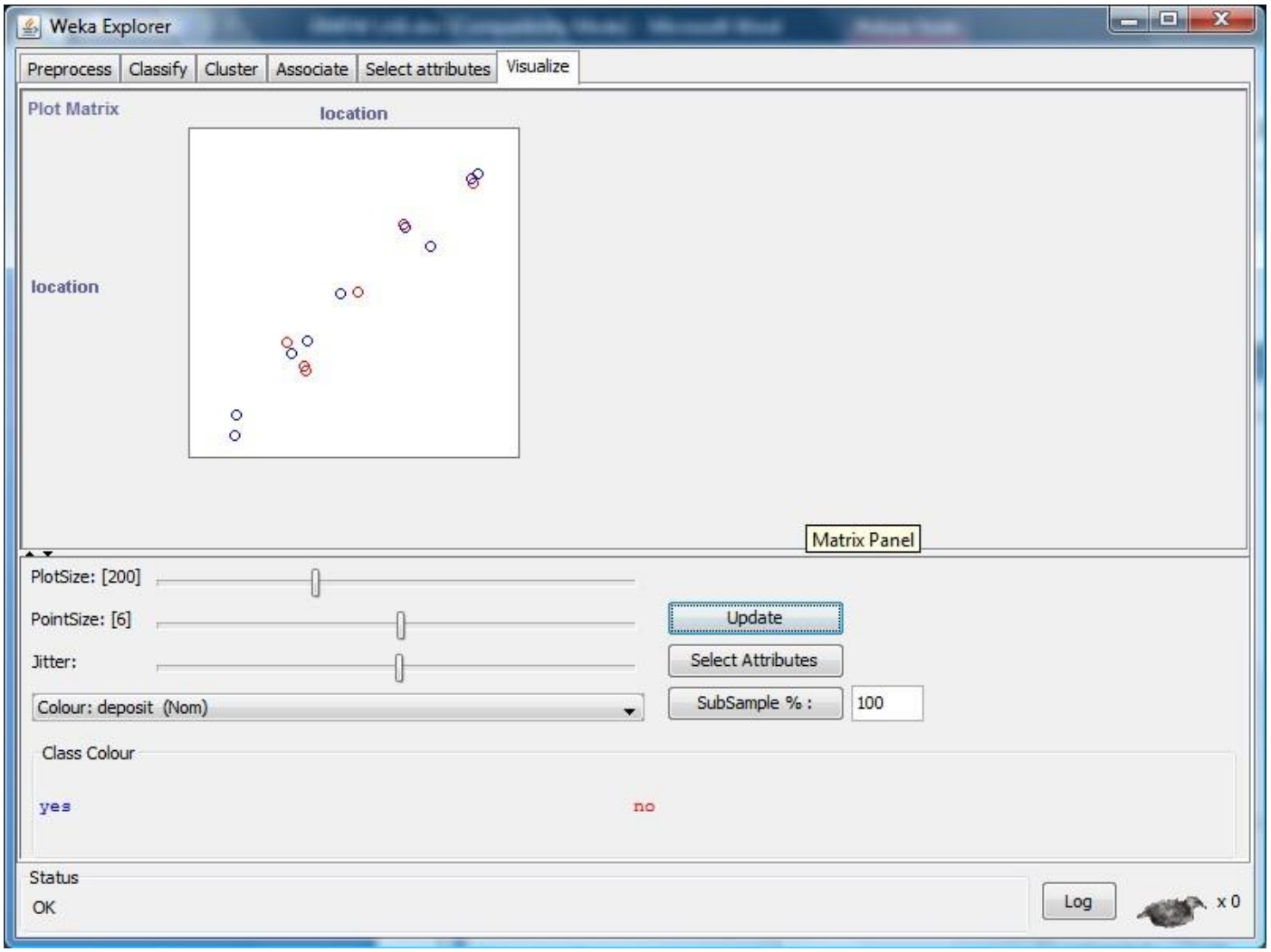


13) After that select the **Select Attribute button** and select **location attribute** and then click OK.

14) **Increase** the **Jitter Size**.

15) Click on the **Update button** to display the output.

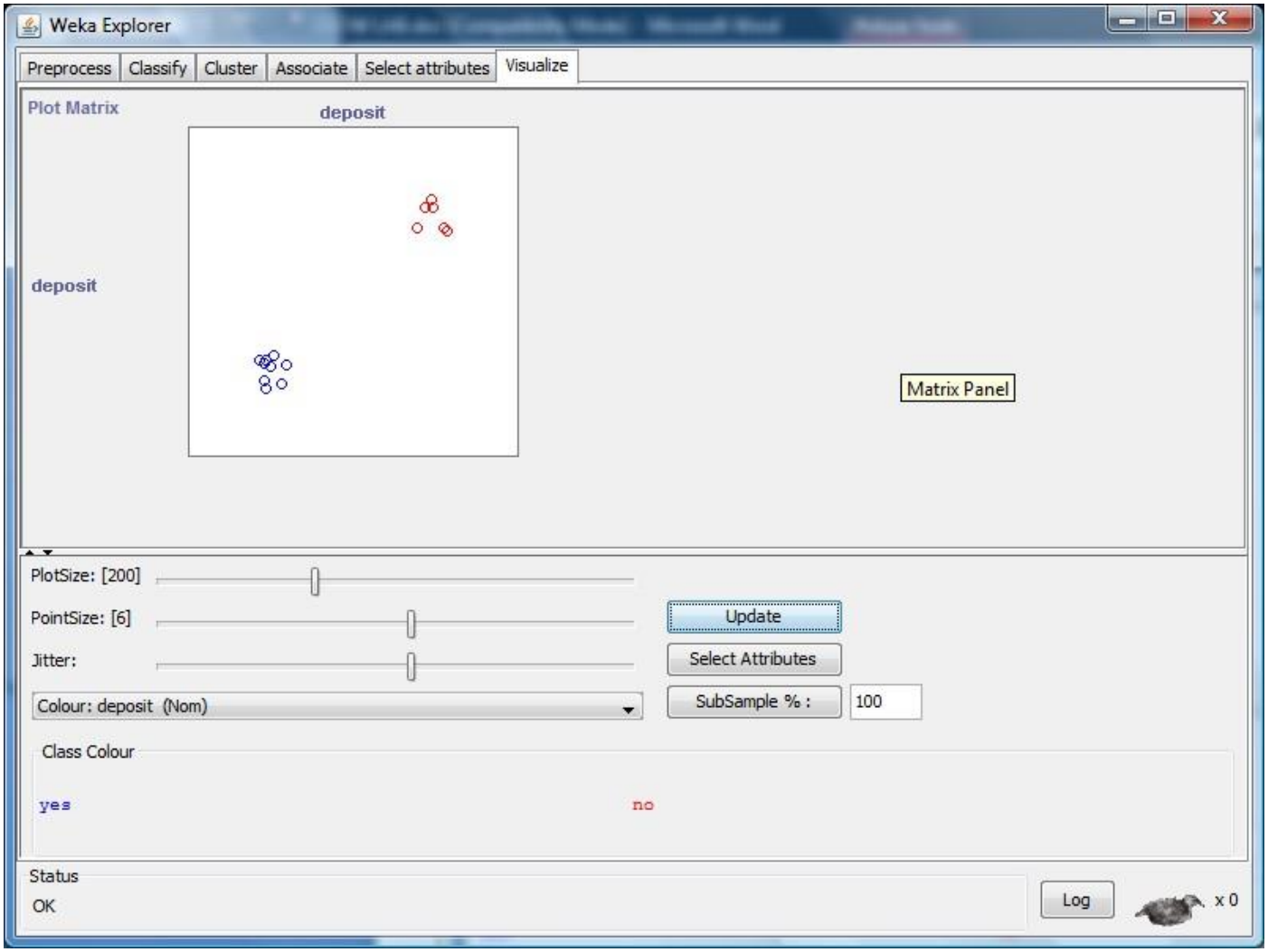
Output:



16) After that we select the **Select Attribute button**, then select **Deposit attribute** and clock OK.

17) Click on the **Update button** to display the output.

Output:



Result:

This program has been successfully executed.

EXPERIMENT NO:15

Aim:

Write a procedure for cross-validation using J48 Algorithm for weather table.

Description:

Cross-validation, sometimes called **rotation estimation**, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice. One round of cross-validation involves partitioning a sample of data into complementary subsets, performing the analysis on one subset (called the *training set*), and validating the analysis on the other subset (called the *validation set* or *testing set*).

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.

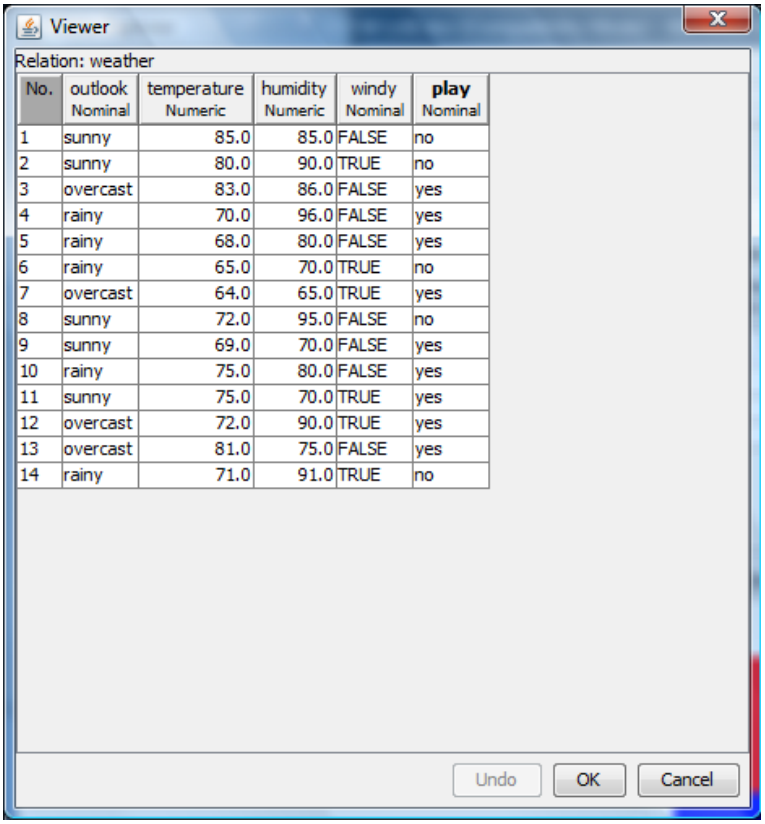
5) Click on **weka-3-4**, then Weka dialog box is displayed on the screen.

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

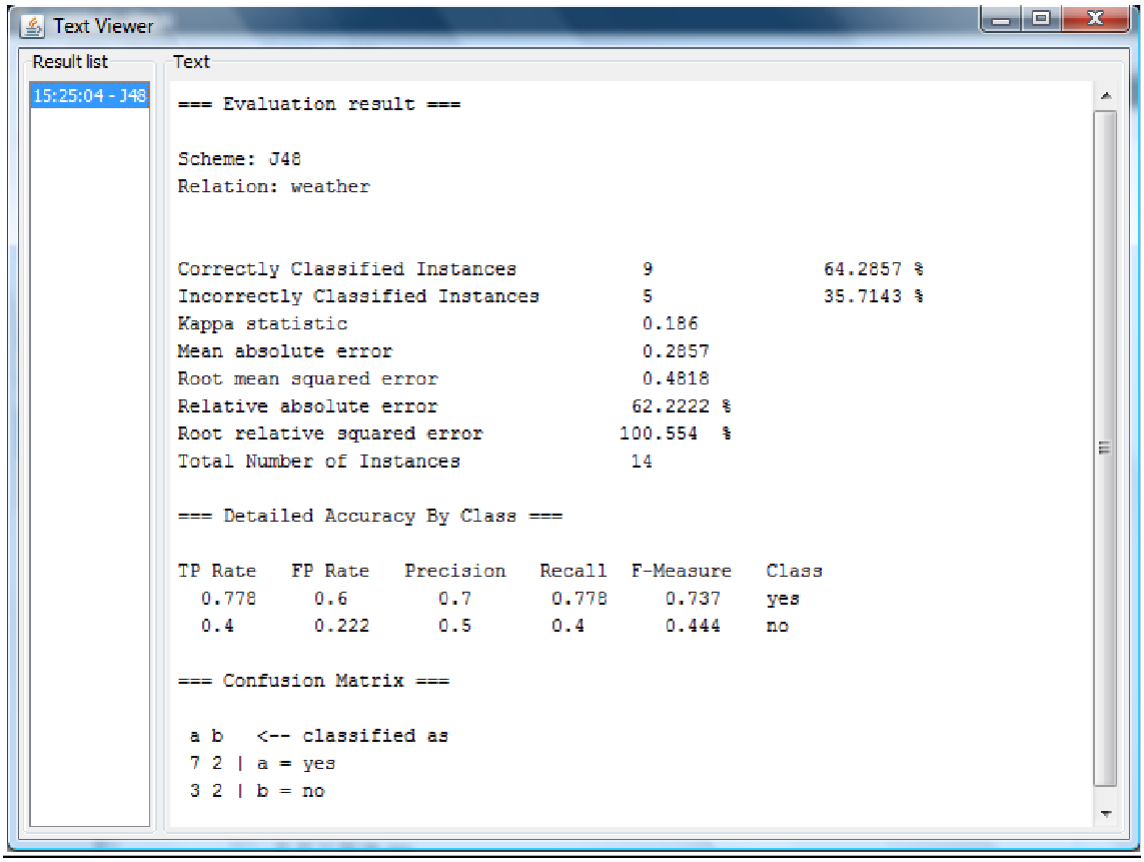


Procedure:

1) **Start** -> **Programs** -> **Weka 3.4**
2) Open **Knowledge Flow**.
3) Select **Data Source** tab & choose **Arff Loader**.
4) Place **Arff Loader** component on the **layout area** by clicking on that component.
5) Specify an Arff file to load by **right clicking on Arff Loader** icon, and then a pop-up menu will appear. In that select **Configure** & browse to the location of **weather.arff**
6) Click on the **Evaluation** tab & choose **Class Assigner** & place it on the layout.
7) Now **connect** the **Arff Loader** to the **Class Assigner** by right clicking on Arff Loader, and then select **Data Set** option, now a link will be established.
8) Right click on **Class Assigner** & choose **Configure** option, and then a new window will appear & specify a class to our data.
9) Select **Evaluation** tab & select **Cross-Validation Fold Maker** & place it on the layout.
10) Now **connect** the **Class Assigner** to the **Cross-Validation Fold Maker**.
11) Select **Classifiers** tab & select **J48** component & place it on the layout.
12) Now **connect Cross-Validation Fold Maker** to **J48 twice**; **first** choose **Training Data Set** option and **then Test Data Set** option.
13) Select **Evaluation Tab** & select **Classifier Performance Evaluator** component & place it on the layout.
14) Connect **J48** to **Classifier Performance Evaluator** component by right clicking on J48 & selecting **Batch Classifier**.
15) Select **Visualization** tab & select **Text Viewer** component & place it on the layout.
16) Connect **Text Viewer** to **Classifier Performance Evaluator** by right clicking on Text Viewer & by selecting **Text** option.
17) Start the flow of execution by selecting **Start Loading** from **Arff Loader**.

18) For viewing **result**, **right click** on **Text Viewer** & select the **Show Results**, and then the result will be displayed on the new window.

Output:



Result:

The program has been successfully executed.

EXPERIMENT NO:16

Aim: Write a procedure for Clustering Buying data using Cobweb Algorithm.

Description:

Cluster analysis or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Creation of Buying Table:

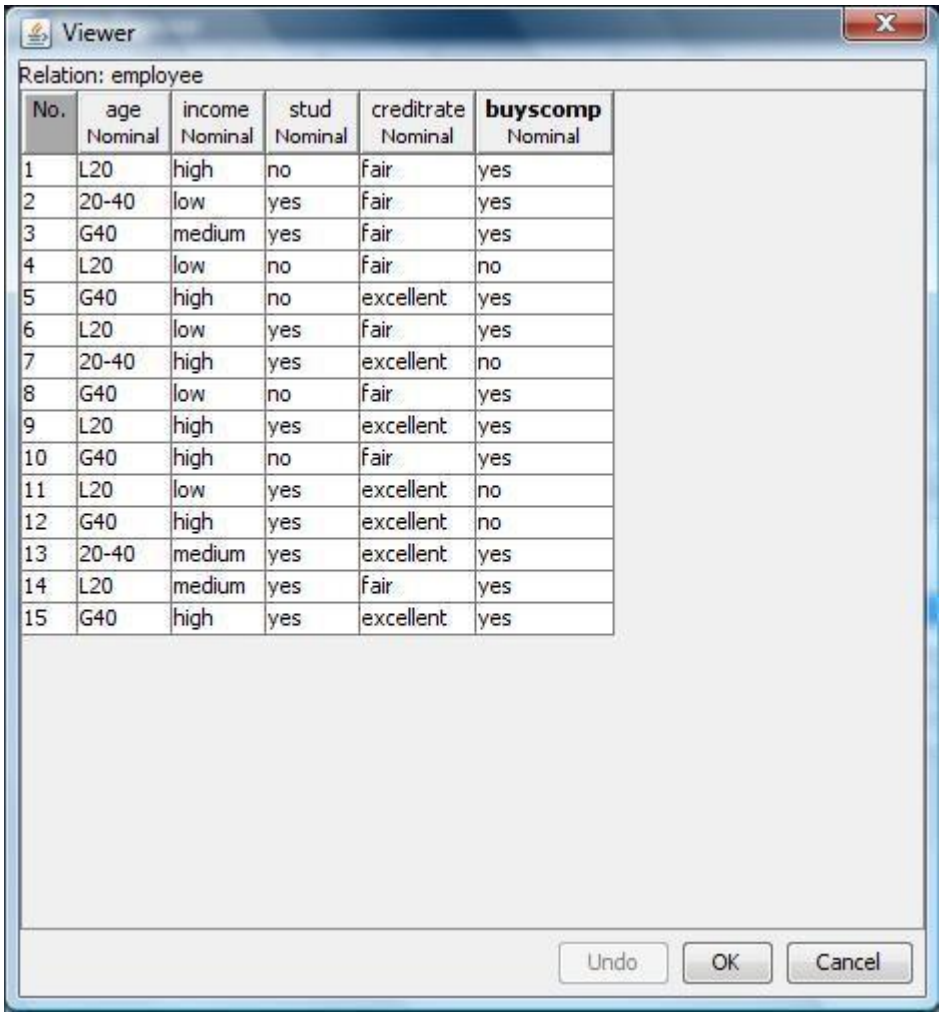
Procedure:

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.

Training Data Set → Buying Table

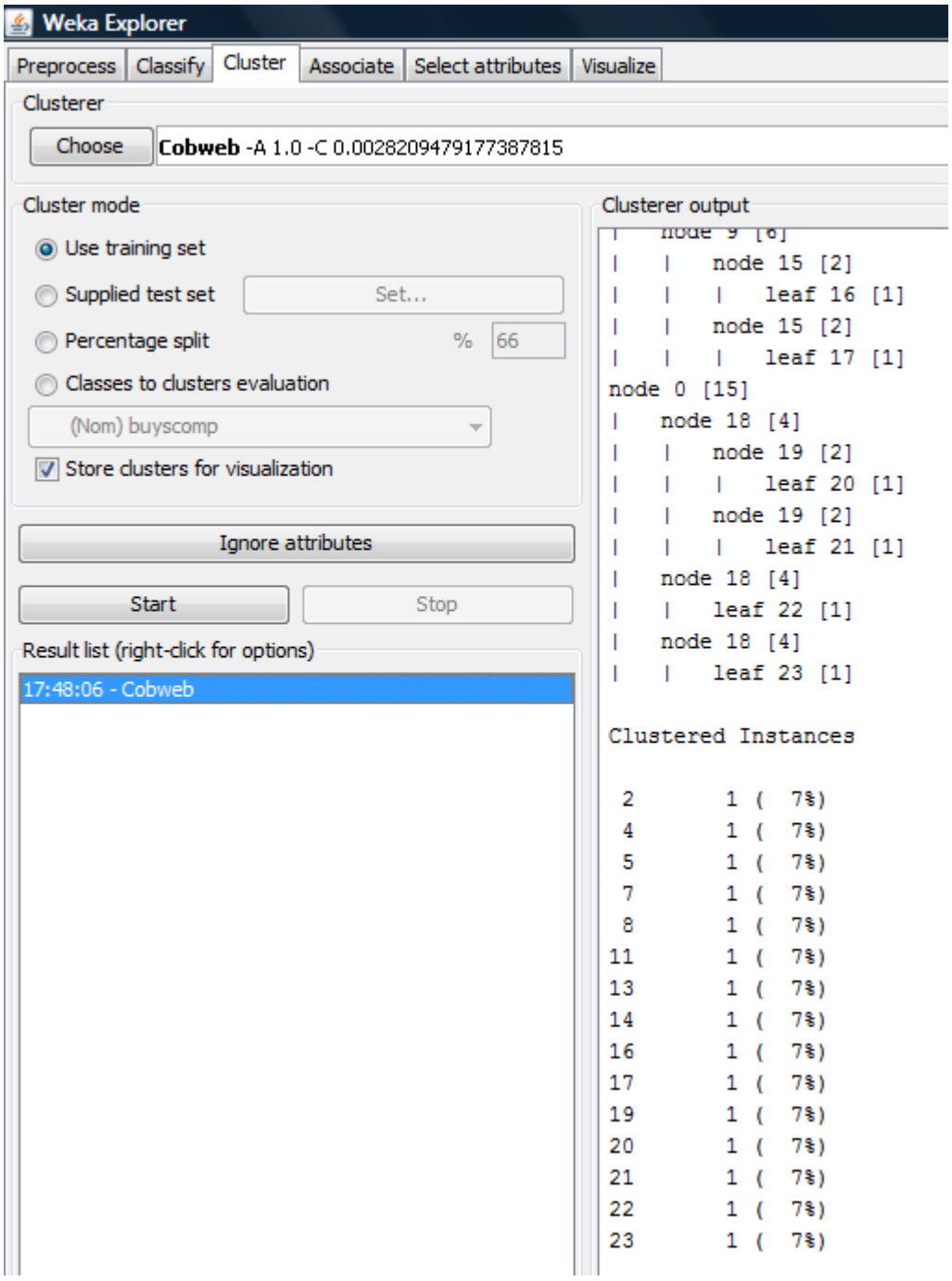


No.	age Nominal	income Nominal	stud Nominal	creditrate Nominal	**buyscomp** Nominal
1	L20	high	no	fair	yes
2	20-40	low	yes	fair	yes
3	G40	medium	yes	fair	yes
4	L20	low	no	fair	no
5	G40	high	no	excellent	yes
6	L20	low	yes	fair	yes
7	20-40	high	yes	excellent	no
8	G40	low	no	fair	yes
9	L20	high	yes	excellent	yes
10	G40	high	no	fair	yes
11	L20	low	yes	excellent	no
12	G40	high	yes	excellent	no
13	20-40	medium	yes	excellent	yes
14	L20	medium	yes	fair	yes
15	G40	high	yes	excellent	yes

Procedure:

1) Click **Start** -> **Programs** -> **Weka 3.4**
2) Click on **Explorer**.
3) Click on **open file** & then select **Buying.arff** file.
4) Click on **Cluster menu**. In this there are different algorithms are there.
5) Click on **Choose button** and then select **cobweb** algorithm.
6) Click on **Start button** and then **output** will be displayed on the screen.

Output:



Result:

The program has been successfully executed.

EXPERIMENT NO:17

Aim: Write a procedure for Clustering Weather data using EM Algorithm.

Description:

Cluster analysis or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Creation of Weather Table:

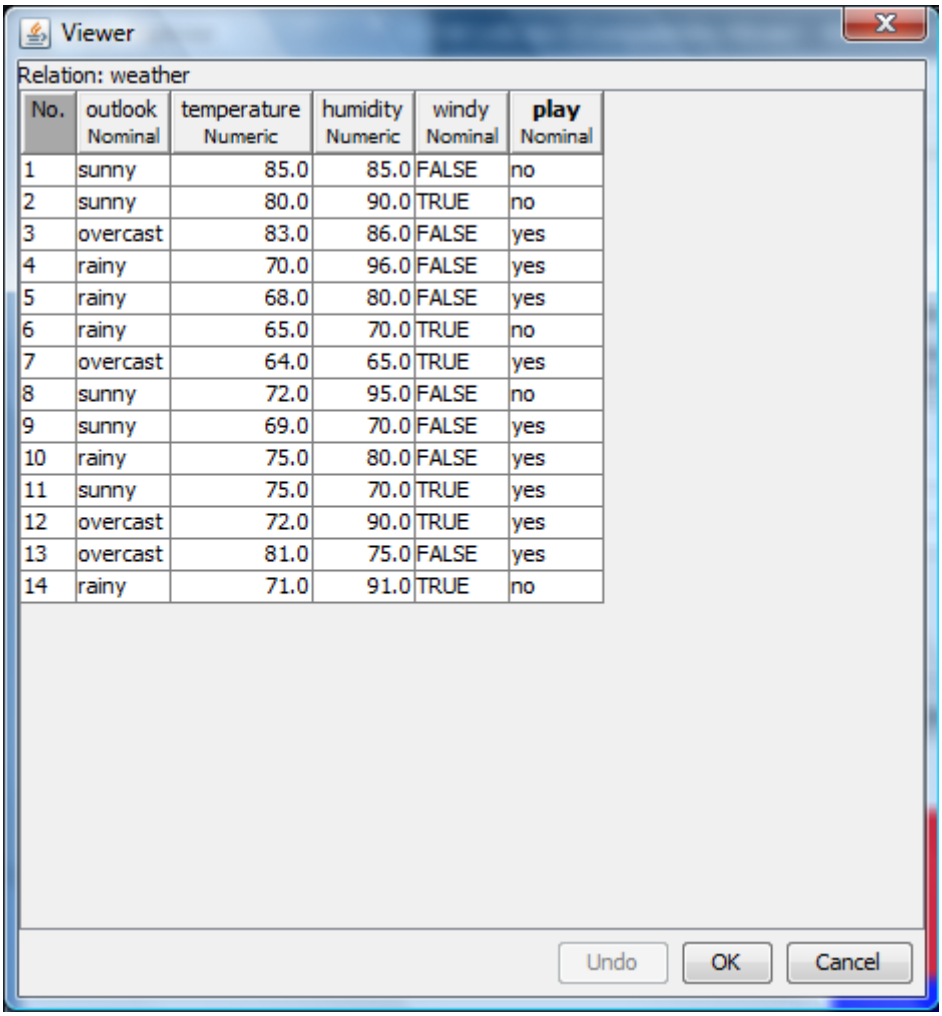
Procedure:

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



Procedure:

9) Click **Start** -> **Programs** -> **Weka 3.4**

10) Click on **Explorer**.

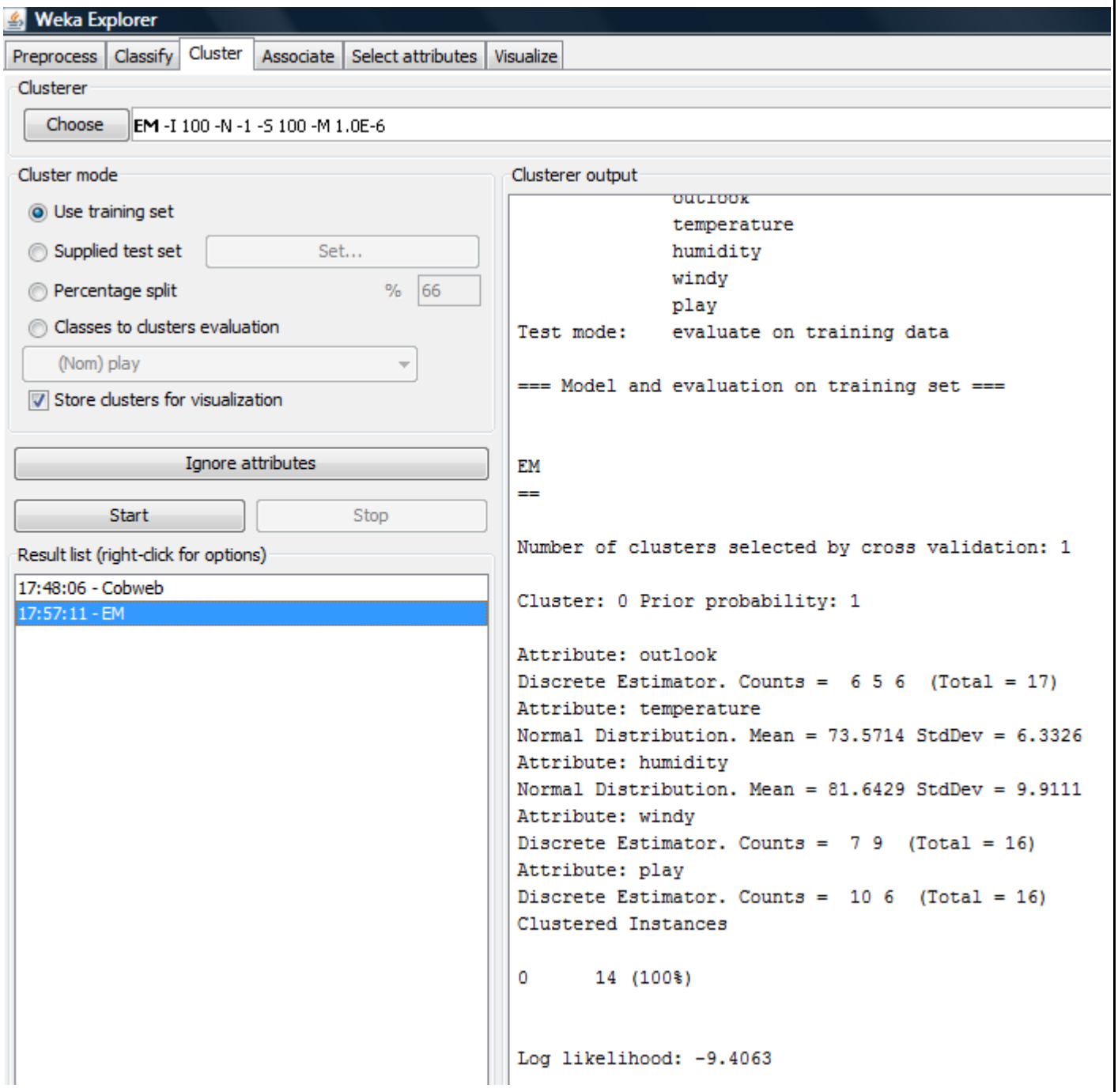
11) Click on **open file** & then select **Weather.arff** file.

12) Click on **Cluster menu**. In this there are different algorithms are there.

13) Click on **Choose button** and then select **EM** algorithm.

14) Click on **Start button** and then **output** will be displayed on the screen.

Output:



Result:

The program has been successfully executed.

EXPERIMENT NO:18

Aim: Write a procedure for Banking data using Farthest First Algorithm.

Description:

Cluster analysis or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Creation of Banking Table:

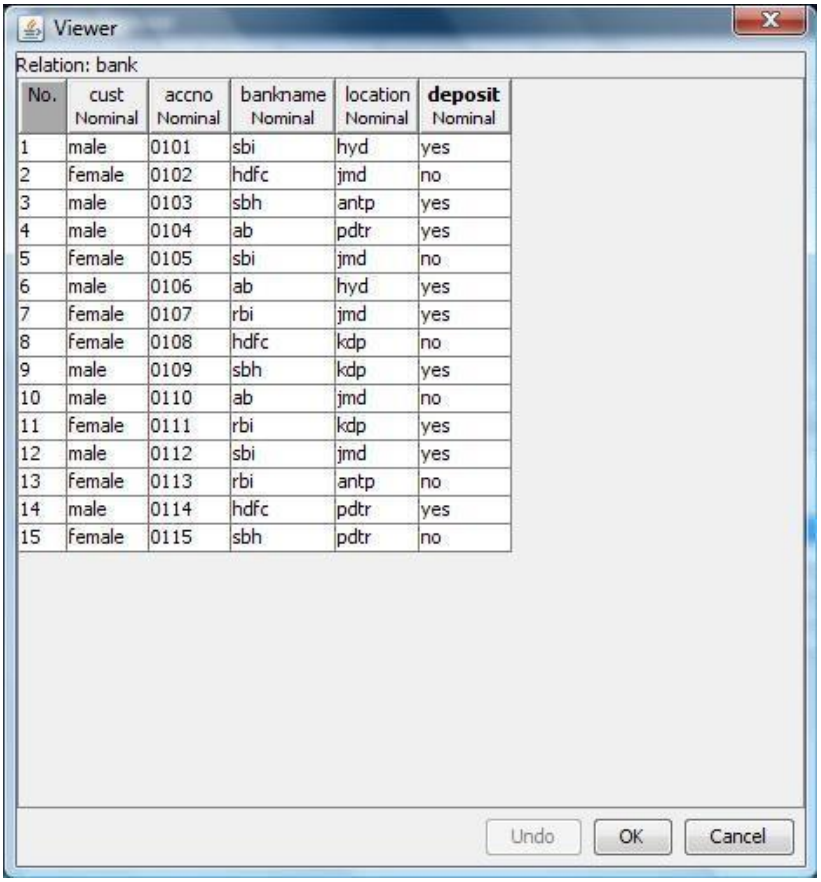
Procedure:

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

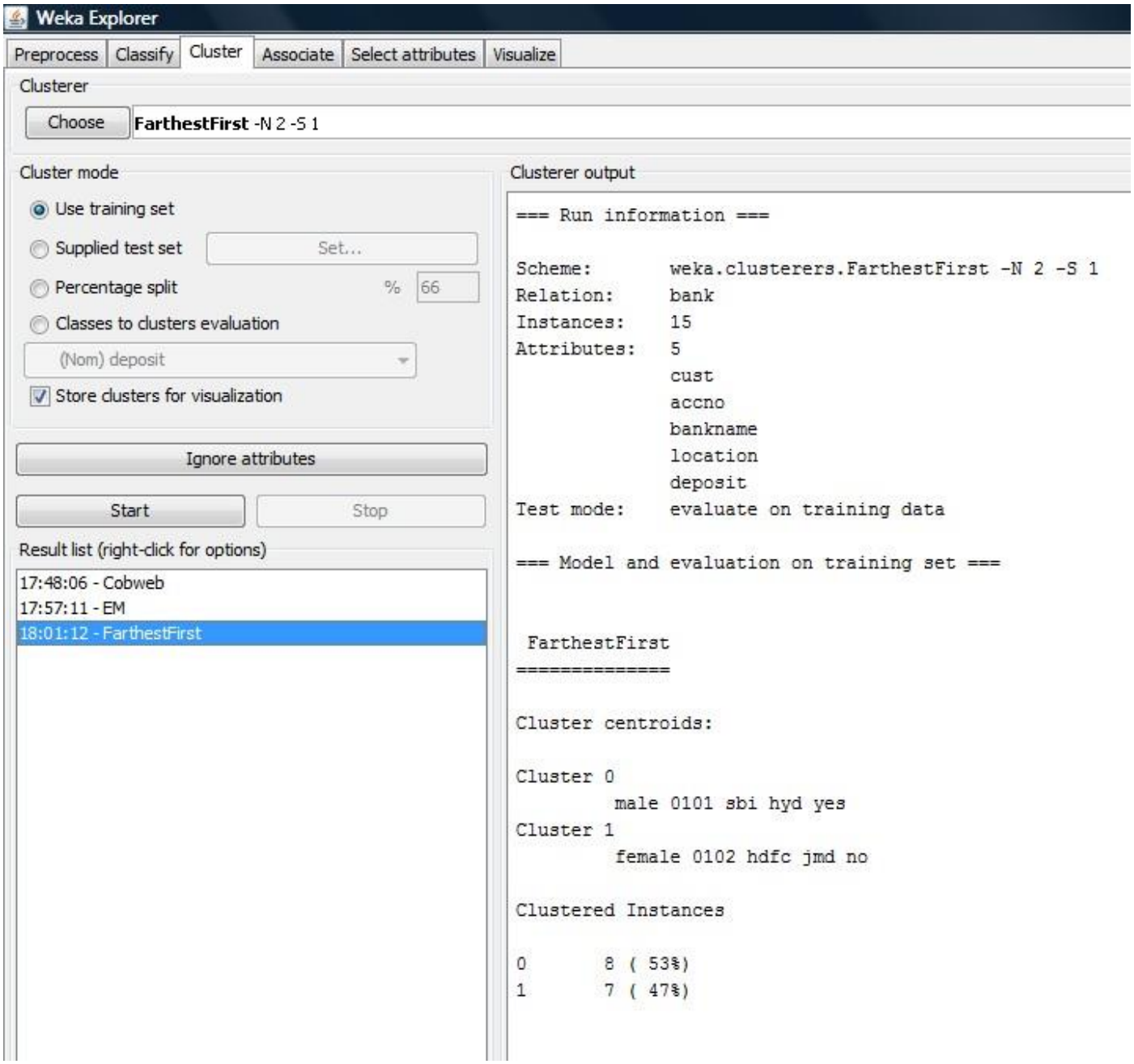


No.	cust Nominal	accno Nominal	bankname Nominal	location Nominal	**deposit** Nominal
1	male	0101	sbi	hyd	yes
2	female	0102	hdfc	jmd	no
3	male	0103	sbh	antp	yes
4	male	0104	ab	pdtr	yes
5	female	0105	sbi	jmd	no
6	male	0106	ab	hyd	yes
7	female	0107	rbi	jmd	yes
8	female	0108	hdfc	kdp	no
9	male	0109	sbh	kdp	yes
10	male	0110	ab	jmd	no
11	female	0111	rbi	kdp	yes
12	male	0112	sbi	jmd	yes
13	female	0113	rbi	antp	no
14	male	0114	hdfc	pdtr	yes
15	female	0115	sbh	pdtr	no

Procedure:

1) Click **Start** -> **Programs** -> **Weka 3.4**
2) Click on **Explorer**.
3) Click on **open file** & then select **Banking.arff** file.
4) Click on **Cluster menu**. In this there are different algorithms are there.
5) Click on **Choose button** and then select **FarthestFirst** algorithm.
6) Click on **Start button** and then **output** will be displayed on the screen.

Output:



Result:

The program has been successfully executed.

EXPERIMENT NO:19

Aim: Write a procedure for Employee data using Make Density Based Cluster Algorithm.

Description:

Cluster analysis or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Creation of Employee Table:

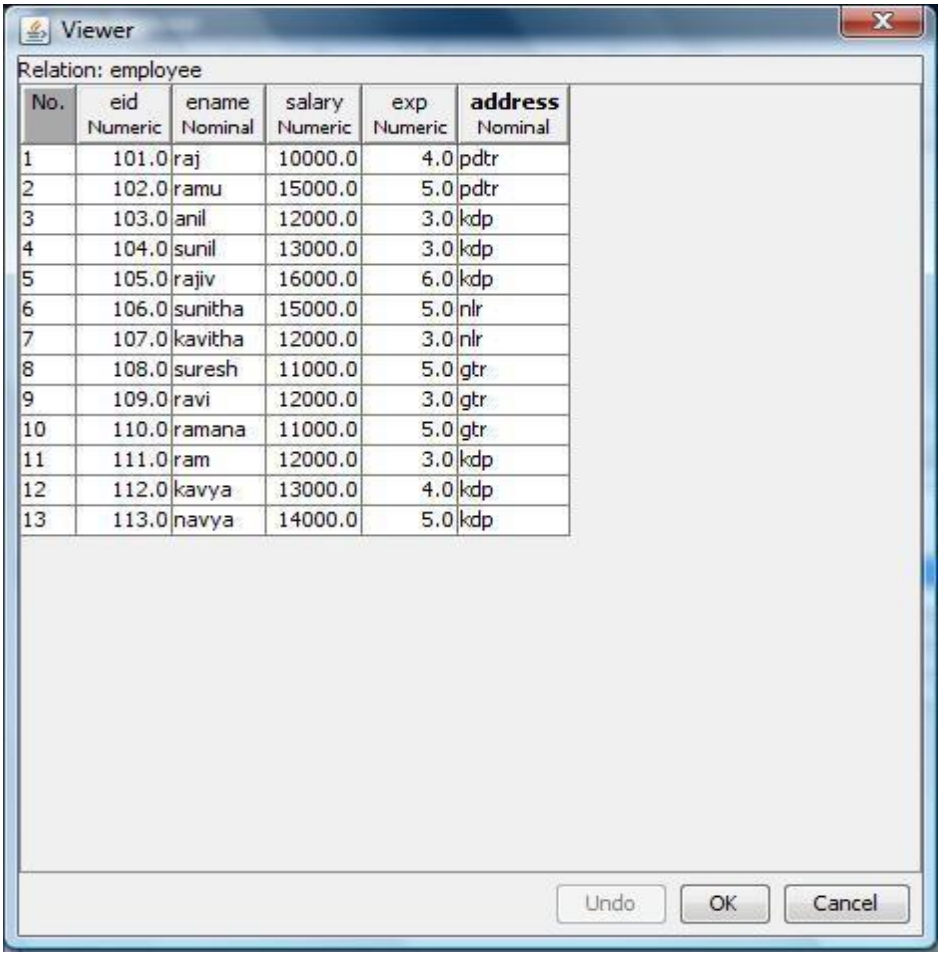
Procedure:

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.

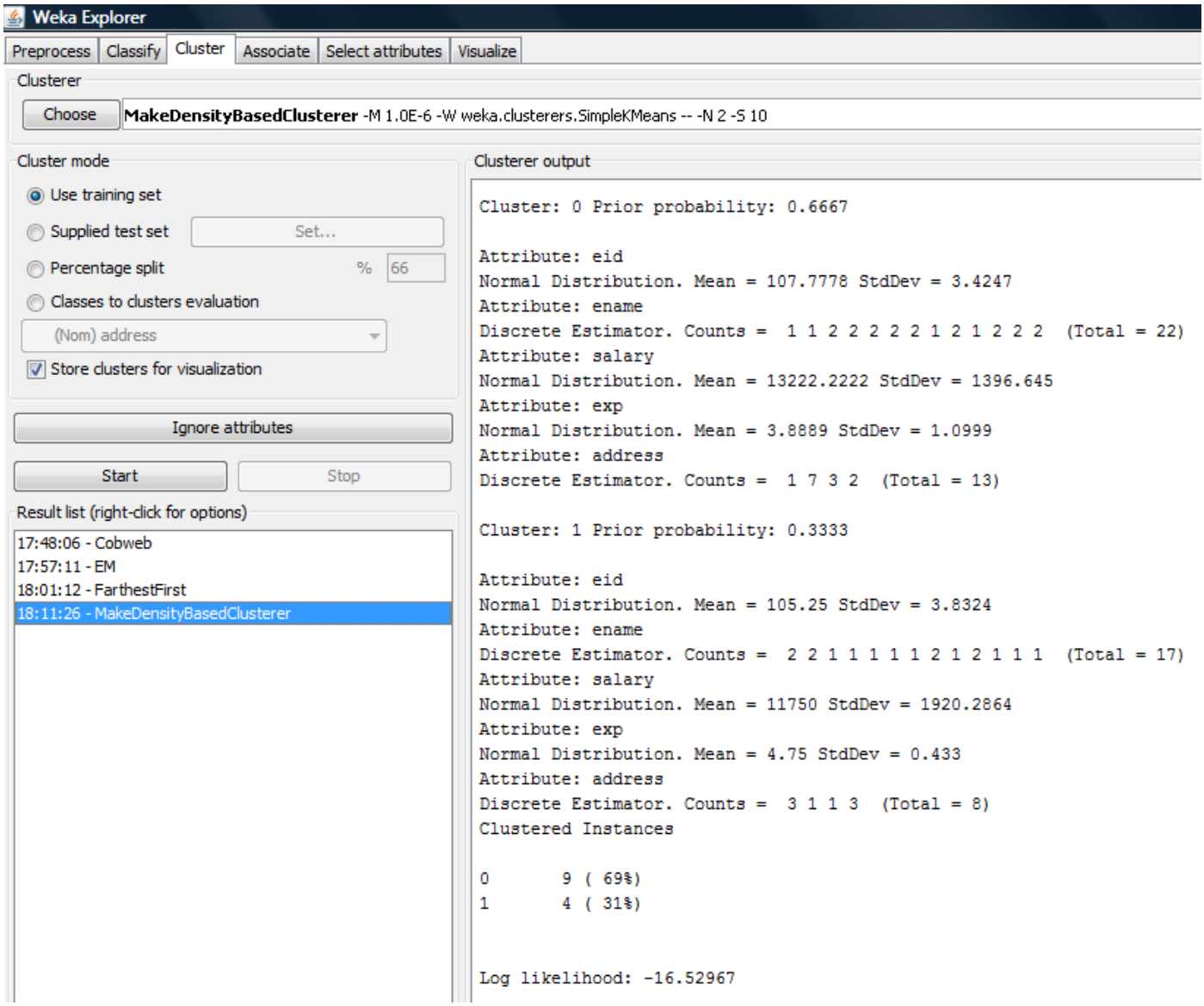
Training Data Set → Employee Table



Procedure:

1) Click **Start** -> **Programs** -> **Weka 3.4**
2) Click on **Explorer**.
3) Click on **open file** & then select **Employee.arff** file.
4) Click on **Cluster menu**. In this there are different algorithms are there.
5) Click on **Choose button** and then select **MakeDensityBasedClusterer** algorithm.
6) Click on **Start button** and then **output** will be displayed on the screen.

Output:



Result:

The program has been successfully executed.

EXPERIMENT NO:20

Aim: Write a procedure for Clustering Customer data using Simple KMeans Algorithm.

Description:

Cluster analysis or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and a common technique for statistical data analysis used in many fields, including machine learning, pattern recognition, image analysis, information retrieval, and bioinformatics.

Creation of Customer Table:

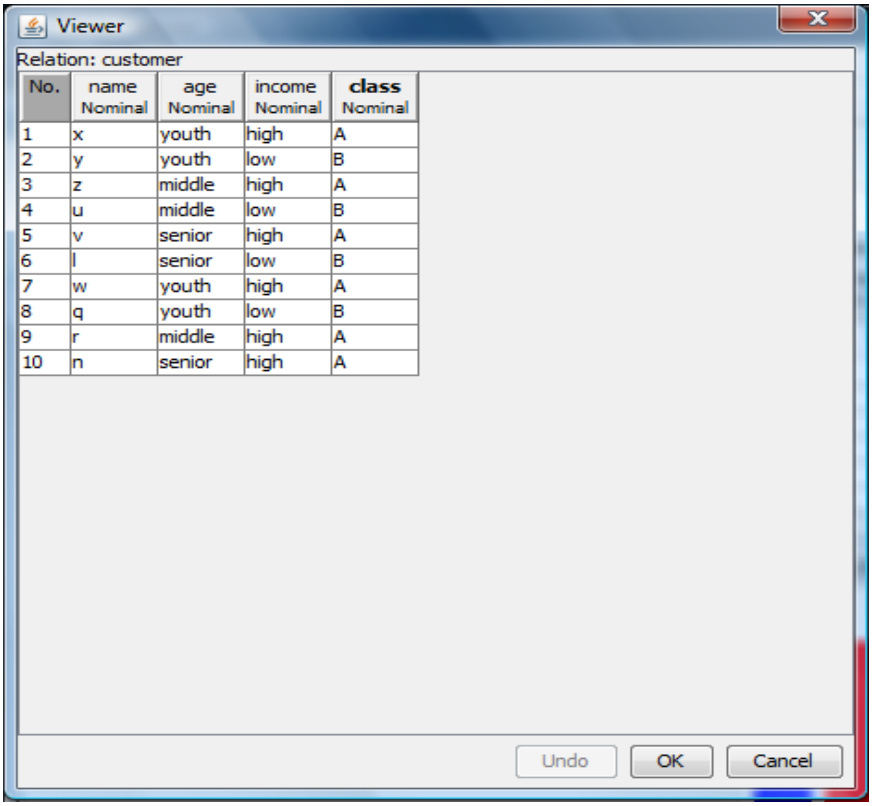
Procedure:

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

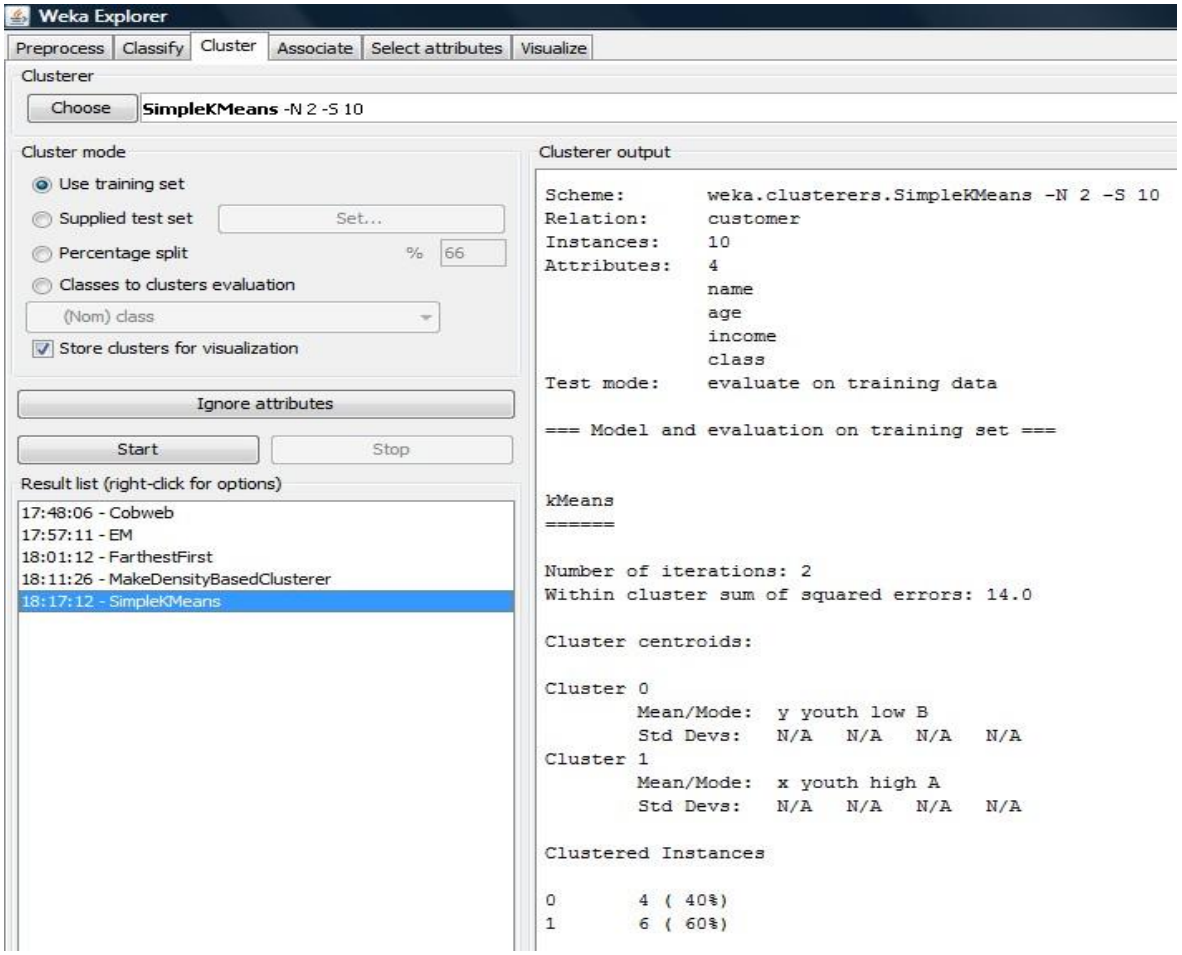
Training Data Set → Customer Table



Procedure:

1) Click **Start** -> **Programs** -> **Weka 3.4**
2) Click on **Explorer**.
3) Click on **open file** & then select **Customer.arff** file.
4) Click on **Cluster menu**. In this there are different algorithms are there.
5) Click on **Choose button** and then select **SimpleKMeans** algorithm.
6) Click on **Start button** and then **output** will be displayed on the screen.

Output:



Result:

The program has been successfully executed.