

**NRI Institute of Information Science and Technology**

**Department of Computer Science and Engineering**

**LAB MANUAL**

**DATA WAREHOUSE AND DATA MINING**

**CS-705**

**Bachelor of Technology (B.Tech.)**

**VII Semester**

## 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

Relation: weather

| No. | outlook<br>Nominal | temperature<br>Numeric | humidity<br>Numeric | windy<br>Nominal | play<br>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             |

Undo OK Cancel

### Result:

This program has been successfully executed.

## EXPERIMENT NO:2

### 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:

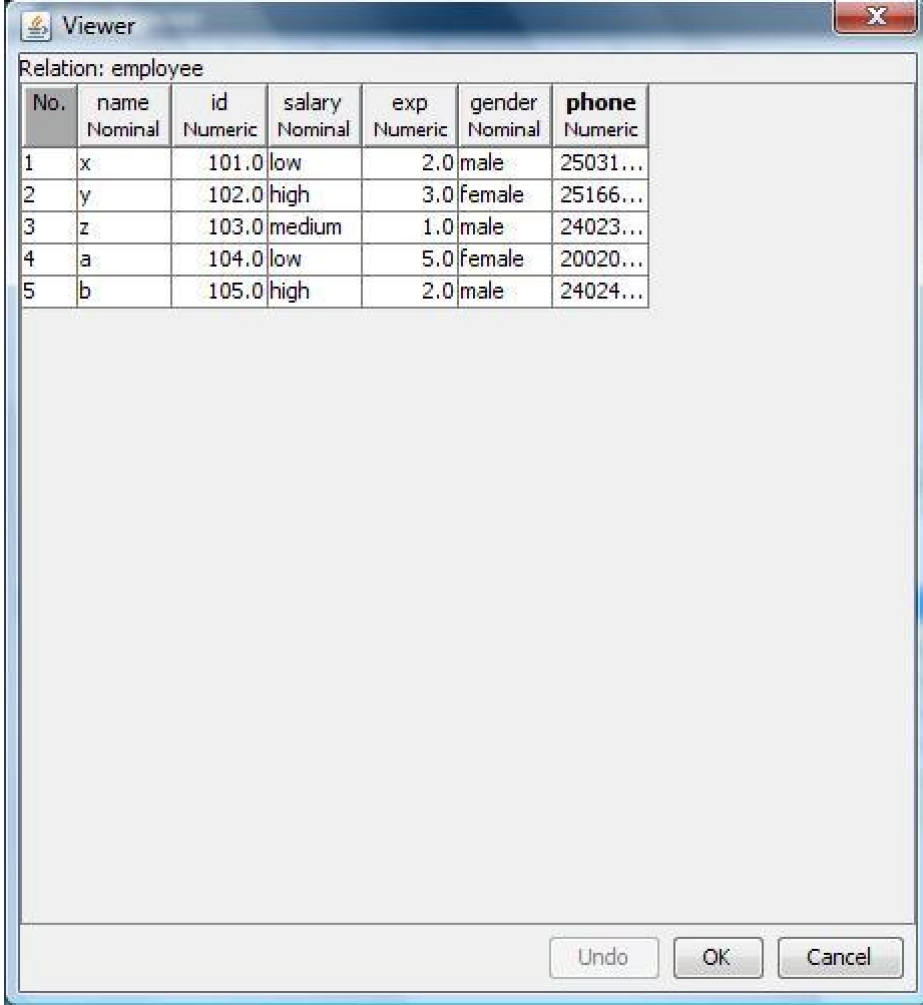
- 1) Open Start → Programs → Accessories → Notepad
- 2) Type the following training data set with the help of Notepad for Employee Table.

```
@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



The screenshot shows a 'Viewer' window with a table titled 'Relation: employee'. The table has 7 columns: 'No.', 'name', 'id', 'salary', 'exp', 'gender', and 'phone'. Each column has a data type listed below it: 'No.' (Numeric), 'name' (Nominal), 'id' (Numeric), 'salary' (Nominal), 'exp' (Numeric), 'gender' (Nominal), and 'phone' (Numeric). The table contains 5 rows of data.

| No. | name    | id      | salary  | exp     | gender  | phone    |
|-----|---------|---------|---------|---------|---------|----------|
|     | Nominal | Numeric | Nominal | Numeric | Nominal | Numeric  |
| 1   | x       | 101.0   | low     | 2.0     | male    | 25031... |
| 2   | y       | 102.0   | high    | 3.0     | female  | 25166... |
| 3   | z       | 103.0   | medium  | 1.0     | male    | 24023... |
| 4   | a       | 104.0   | low     | 5.0     | female  | 20020... |
| 5   | b       | 105.0   | high    | 2.0     | male    | 24024... |

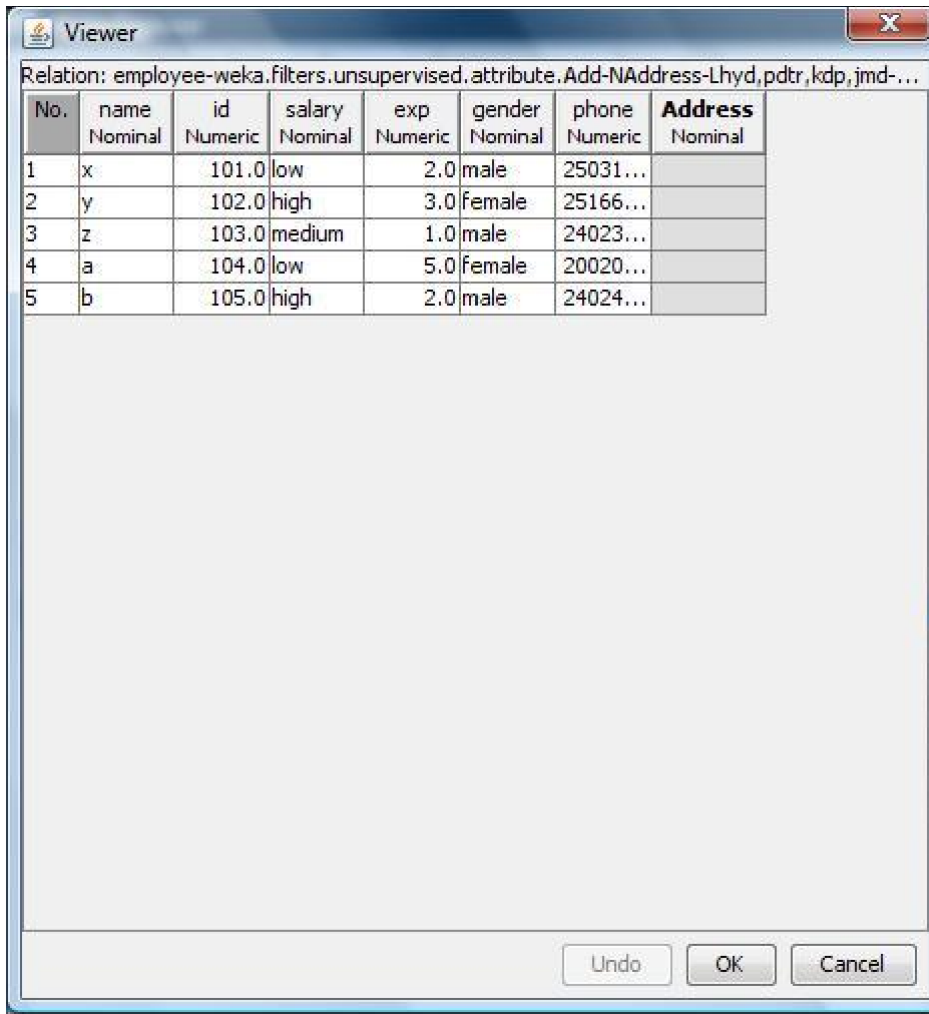
At the bottom of the window, there are three buttons: 'Undo', 'OK', and 'Cancel'.

## 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:



Relation: employee-weka.filters.unsupervised.attribute.Add-NAddress-Lhyd,pdtr,kdp,jmd-...

| No. | name    | id      | salary  | exp     | gender  | phone    | Address |
|-----|---------|---------|---------|---------|---------|----------|---------|
|     | Nominal | Numeric | Nominal | Numeric | Nominal | Numeric  | Nominal |
| 1   | x       | 101.0   | low     | 2.0     | male    | 25031... |         |
| 2   | y       | 102.0   | high    | 3.0     | female  | 25166... |         |
| 3   | z       | 103.0   | medium  | 1.0     | male    | 24023... |         |
| 4   | a       | 104.0   | low     | 5.0     | female  | 20020... |         |
| 5   | b       | 105.0   | high    | 2.0     | male    | 24024... |         |

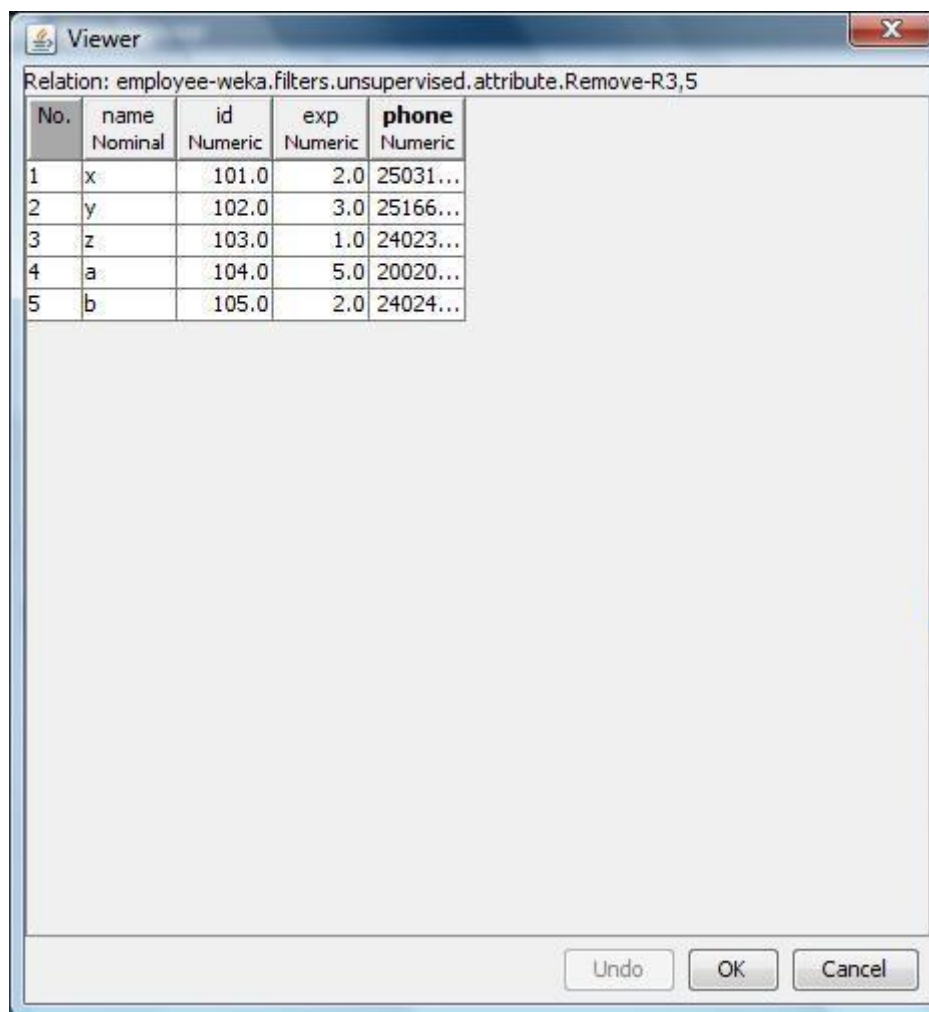
Undo OK Cancel

### 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:



Relation: employee-weka.filters.unsupervised.attribute.Remove-R3,5

| No. | name<br>Nominal | id<br>Numeric | exp<br>Numeric | phone<br>Numeric |
|-----|-----------------|---------------|----------------|------------------|
| 1   | x               | 101.0         | 2.0            | 25031...         |
| 2   | y               | 102.0         | 3.0            | 25166...         |
| 3   | z               | 103.0         | 1.0            | 24023...         |
| 4   | a               | 104.0         | 5.0            | 20020...         |
| 5   | b               | 105.0         | 2.0            | 24024...         |

Undo OK Cancel

## 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:**

Relation: employee-weka.filters.unsupervised.attribute.Normalize

| No. | name<br>Nominal | id<br>Numeric | salary<br>Nominal | exp<br>Numeric | gender<br>Nominal | phone<br>Numeric |
|-----|-----------------|---------------|-------------------|----------------|-------------------|------------------|
| 1   | x               | 0.0           | low               | 0.25           | male              | 25031...         |
| 2   | y               | 0.25          | high              | 0.5            | female            | 25166...         |
| 3   | z               | 0.5           | medium            | 0.0            | male              | 24023...         |
| 4   | a               | 0.75          | low               | 1.0            | female            | 20020...         |
| 5   | b               | 1.0           | high              | 0.25           | male              | 24024...         |

Undo OK Cancel

**Result:**

This program has been successfully executed.

## EXPERIMENT NO:3

### 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 canvas and connect them together in order to form a knowledge flow for processing and analyzing the data.

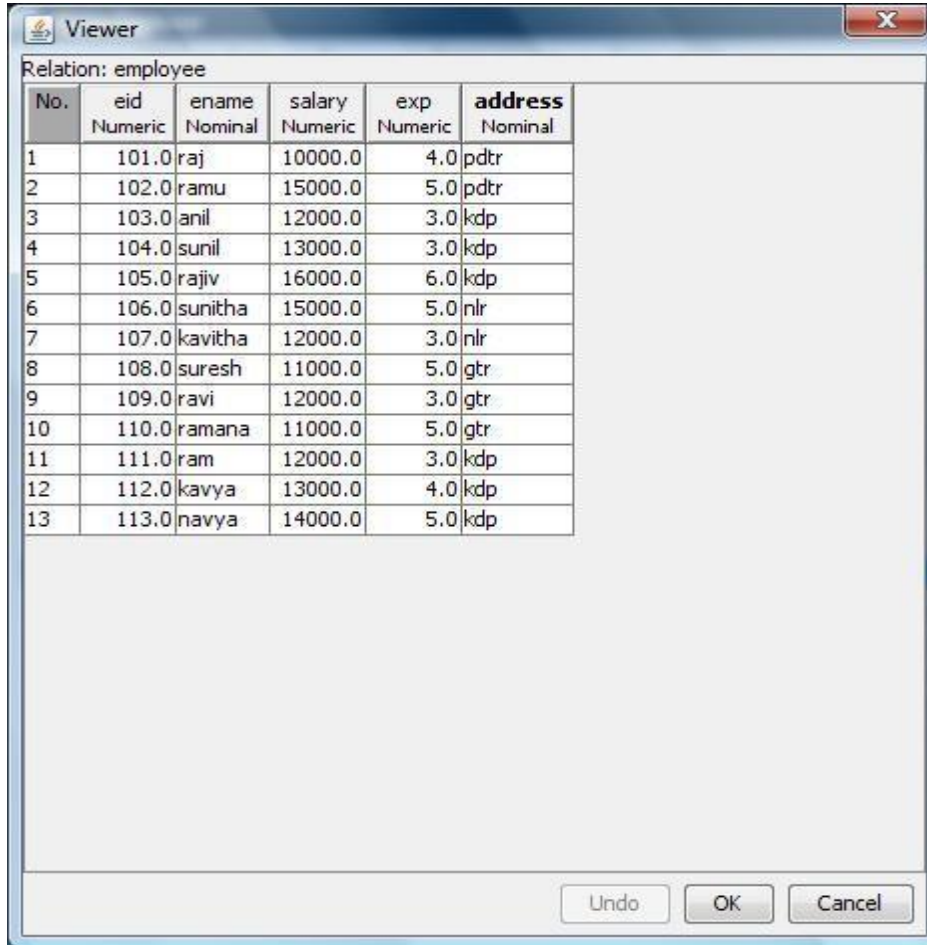
### Creation of Employee Table:

### 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



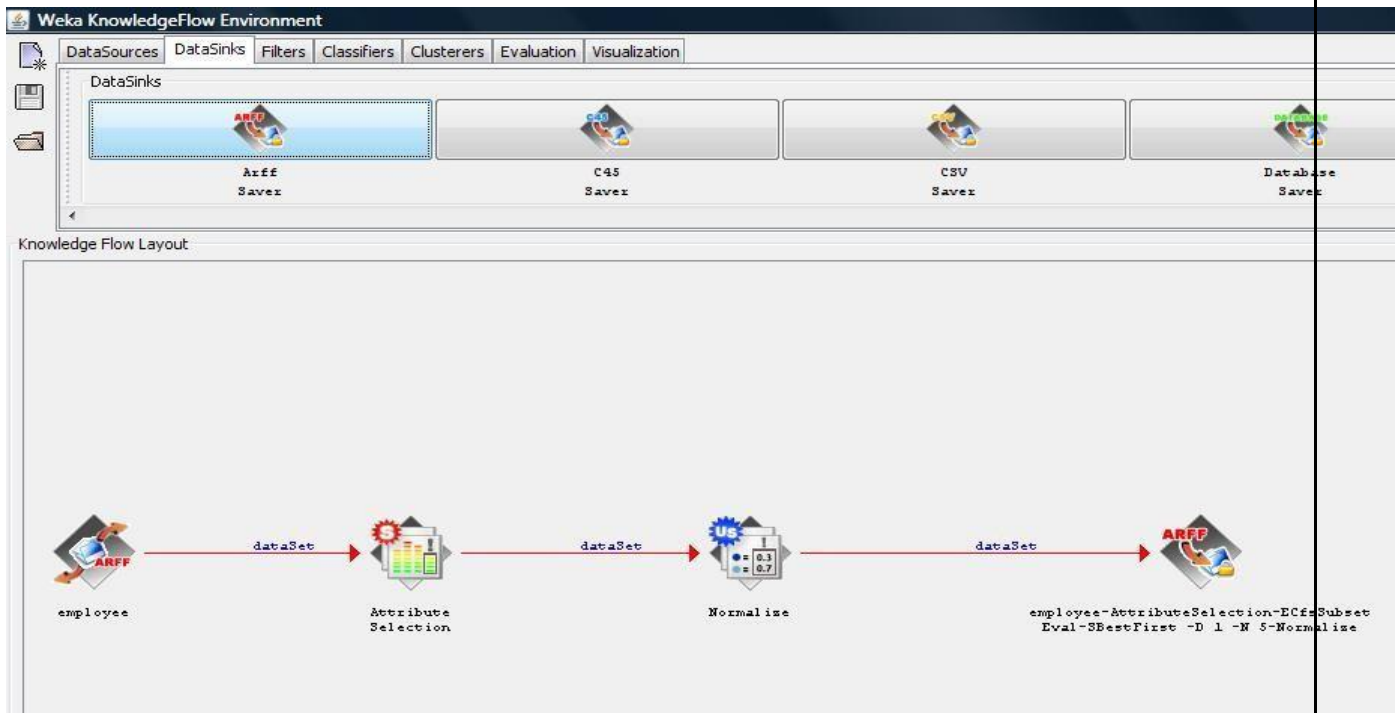
| No. | eid     | ename   | salary  | exp     | address |
|-----|---------|---------|---------|---------|---------|
|     | Numeric | Nominal | Numeric | Numeric | Nominal |
| 1   | 101.0   | raj     | 10000.0 | 4.0     | pdtr    |
| 2   | 102.0   | ramu    | 15000.0 | 5.0     | pdtr    |
| 3   | 103.0   | 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 **Normalize** 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 **Attribute 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**.

|    | A  | B | C | D | E | F | G | H | I | J |
|----|--|---|---|---|---|---|---|---|---|---|
| 1  | @relation 'employee-weka.filters.supervised.attribute.AttributeSelection-Eweka.attributeSelection.Cf |   |   |   |   |   |   |   |   |   |
| 2  |  |   |   |   |   |   |   |   |   |   |
| 3  | @attribute eid numeric   |   |   |   |   |   |   |   |   |   |
| 4  | @attribute ename {raj,ramu,anil,sunil,rajiv,sunitha,kavitha,suresh,ravi,ramana,ram,kavya,navya}      |   |   |   |   |   |   |   |   |   |
| 5  | @attribute address {pdtr,kdp,nlr,gtr}  |   |   |   |   |   |   |   |   |   |
| 6  |  |   |   |   |   |   |   |   |   |   |
| 7  | @data  |   |   |   |   |   |   |   |   |   |
| 8  | 0,raj,pdtr   |   |   |   |   |   |   |   |   |   |
| 9  | 0.083333,ramu,pdtr   |   |   |   |   |   |   |   |   |   |
| 10 | 0.166667,anil,kdp  |   |   |   |   |   |   |   |   |   |
| 11 | 0.25,sunil,kdp   |   |   |   |   |   |   |   |   |   |
| 12 | 0.333333,rajiv,kdp   |   |   |   |   |   |   |   |   |   |
| 13 | 0.416667,sunitha,nlr   |   |   |   |   |   |   |   |   |   |
| 14 | 0.5,kavitha,nlr  |   |   |   |   |   |   |   |   |   |
| 15 | 0.583333,suresh,gtr  |   |   |   |   |   |   |   |   |   |
| 16 | 0.666667,ravi,gtr  |   |   |   |   |   |   |   |   |   |
| 17 | 0.75,ramana,gtr  |   |   |   |   |   |   |   |   |   |
| 18 | 0.833333,ram,kdp   |   |   |   |   |   |   |   |   |   |
| 19 | 0.916667,kavya,kdp   |   |   |   |   |   |   |   |   |   |
| 20 | 1,navya,kdp  |   |   |   |   |   |   |   |   |   |
| 21 |  |   |   |   |   |   |   |   |   |   |



**Result:**

This program has been successfully executed.

## EXPERIMENT NO:4

**Aim:** Finding Association Rules for Employee data.

### Description:

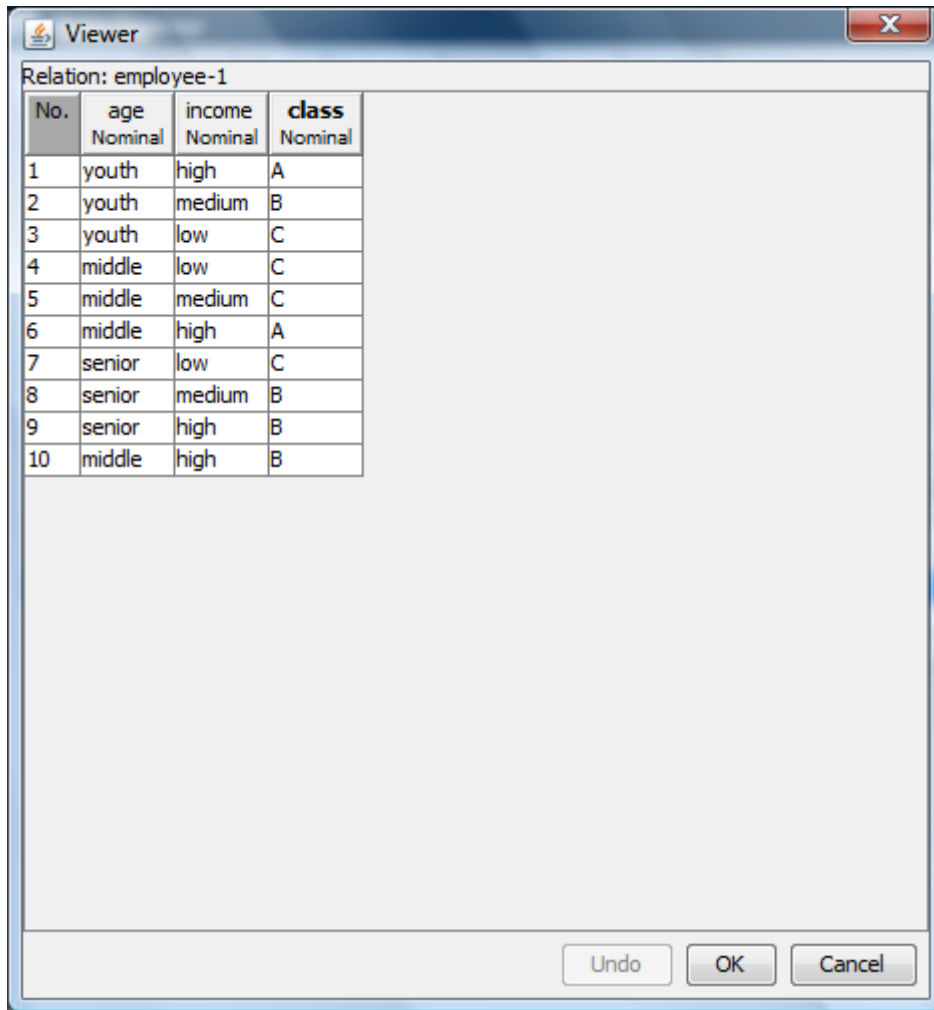
In data mining, **association rule learning** is a popular and well researched method for discovering interesting relations between variables in large databases. It can be described as analyzing and presenting strong rules discovered in databases using different measures of interestingness. In market 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



The screenshot shows a 'Viewer' window with the title 'Relation: employee-1'. It displays a table with 10 rows and 4 columns. The columns are labeled 'No.', 'age', 'income', and 'class'. The 'age' and 'income' columns have 'Nominal' data types listed below them. The 'class' column also has 'Nominal' listed below it. The data rows are as follows:

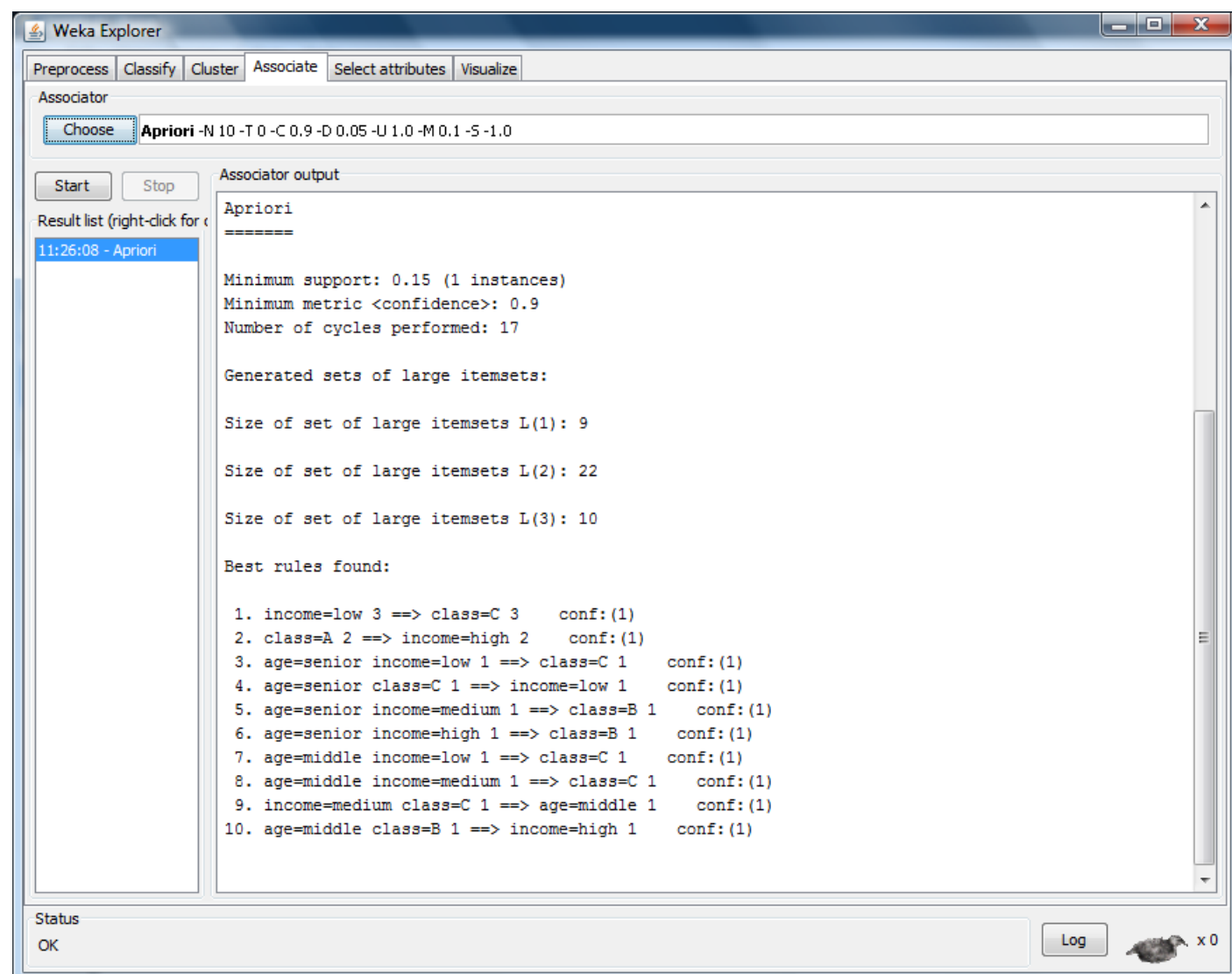
| No. | age    | income | class |
|-----|--------|--------|-------|
| 1   | youth  | high   | A     |
| 2   | youth  | medium | B     |
| 3   | youth  | low    | C     |
| 4   | middle | low    | C     |
| 5   | middle | medium | C     |
| 6   | middle | high   | A     |
| 7   | senior | low    | C     |
| 8   | senior | medium | B     |
| 9   | senior | high   | B     |
| 10  | middle | high   | B     |

At the bottom of the window, there are three buttons: 'Undo', 'OK', and 'Cancel'.

### 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:5

**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:

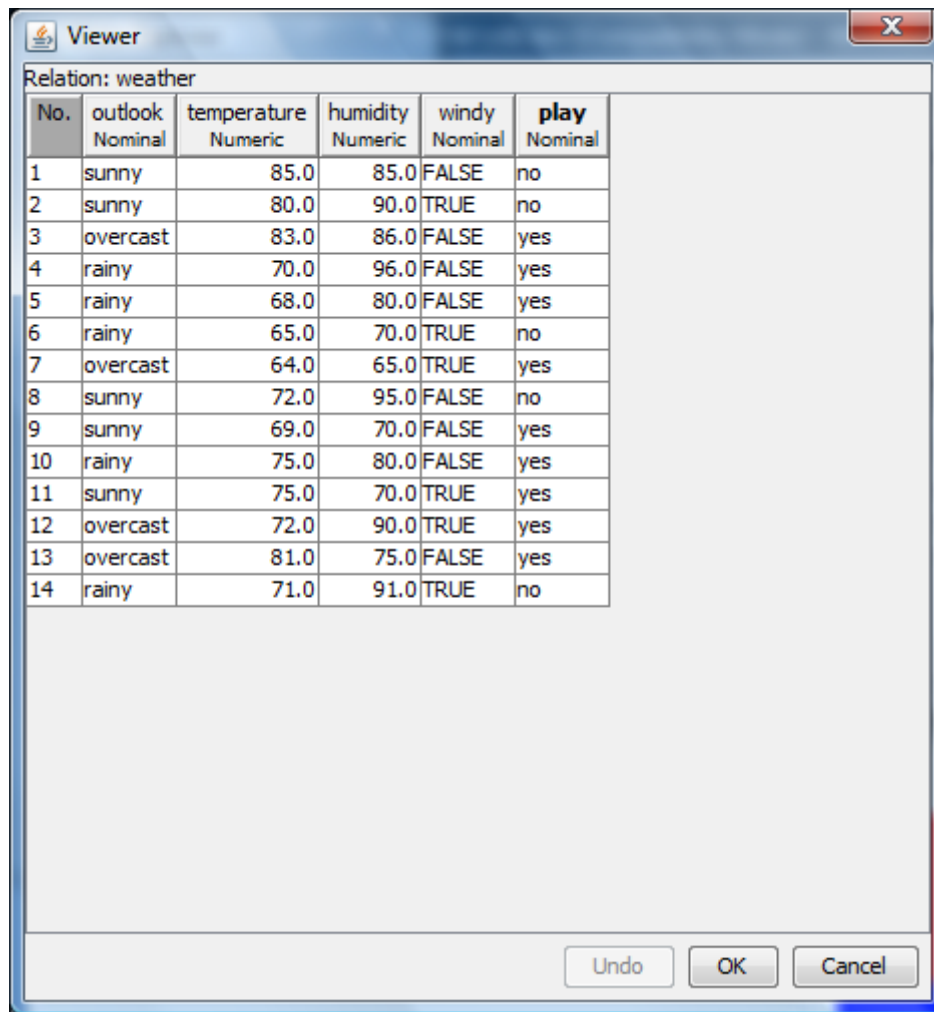
- 1) Open Start → Programs → Accessories → Notepad
- 2) Type the following training data set with the help of Notepad for Weather Table.

```
@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



The screenshot shows a window titled "Viewer" with a close button (X) in the top right corner. The window displays a table with the following data:

| No. | outlook<br>Nominal | temperature<br>Numeric | humidity<br>Numeric | windy<br>Nominal | play<br>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              |

At the bottom of the window, there are three buttons: "Undo", "OK", and "Cancel".

### 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:

The screenshot displays the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' window is open, showing the 'EM' algorithm with parameters '-I 100 -N -1 -S 100 -M 1.0E-6'. The 'Cluster mode' section is configured with 'Use training set' selected, 'Percentage split' at 66%, and 'Store clusters for visualization' checked. The 'Clusterer output' window shows the following text:

```
outlook
temperature
humidity
windy
play
Test mode: evaluate on training data

=== Model and evaluation on training set ===

EM
==

Number of clusters selected by cross validation: 1

Cluster: 0 Prior probability: 1

Attribute: outlook
Discrete Estimator. Counts = 6 5 6 (Total = 17)
Attribute: temperature
Normal Distribution. Mean = 73.5714 StdDev = 6.3326
Attribute: humidity
Normal Distribution. Mean = 81.6429 StdDev = 9.9111
Attribute: windy
Discrete Estimator. Counts = 7 9 (Total = 16)
Attribute: play
Discrete Estimator. Counts = 10 6 (Total = 16)
Clustered Instances

0      14 (100%)

Log likelihood: -9.4063
```

The 'Result list' on the left shows two entries: '17:48:06 - Cobweb' and '17:57:11 - EM', with the latter selected.

## Result:

The program has been successfully executed.

## EXPERIMENT NO:6

**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:

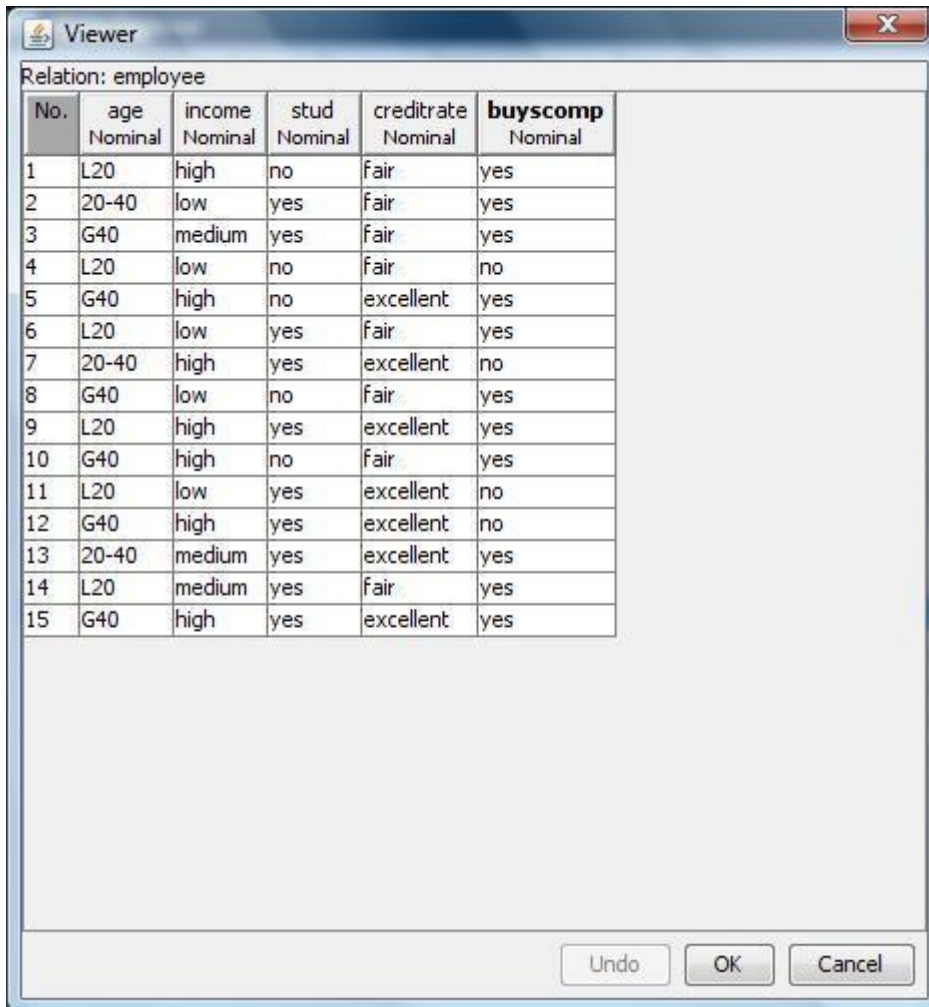
### Procedure:

- 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



The screenshot shows a 'Viewer' window with a table of data. The table has 6 columns: 'No.', 'age', 'income', 'stud', 'creditrates', and 'buyscomp'. Each column has a 'Nominal' data type. The data is as follows:

| No. | age   | income | stud | creditrates | buyscomp |
|-----|-------|--------|------|-------------|----------|
| 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      |

At the bottom of the window, there are three buttons: 'Undo', 'OK', and 'Cancel'.

### 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:**

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'Cobweb -A 1.0 -C 0.0028209479177387815'. Under 'Cluster mode', 'Use training set' is selected, and 'Store clusters for visualization' is checked. The 'Ignore attributes' button is visible. The 'Start' button has been clicked, and the 'Result list' shows '17:48:06 - Cobweb'. The 'Clusterer output' pane displays a hierarchical tree structure of nodes and leaves, and the 'Clustered Instances' pane shows a list of instances with their cluster assignments.

```
node 9 [6]
| | node 15 [2]
| | | leaf 16 [1]
| | node 15 [2]
| | | leaf 17 [1]
node 0 [15]
| node 18 [4]
| | node 19 [2]
| | | leaf 20 [1]
| | node 19 [2]
| | | leaf 21 [1]
| node 18 [4]
| | leaf 22 [1]
| node 18 [4]
| | leaf 23 [1]

Clustered Instances

2      1 ( 7%)
4      1 ( 7%)
5      1 ( 7%)
7      1 ( 7%)
8      1 ( 7%)
11     1 ( 7%)
13     1 ( 7%)
14     1 ( 7%)
16     1 ( 7%)
17     1 ( 7%)
19     1 ( 7%)
20     1 ( 7%)
21     1 ( 7%)
22     1 ( 7%)
23     1 ( 7%)
```

**Result:**

The program has been successfully executed.

## EXPERIMENT NO:7

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

### Description:

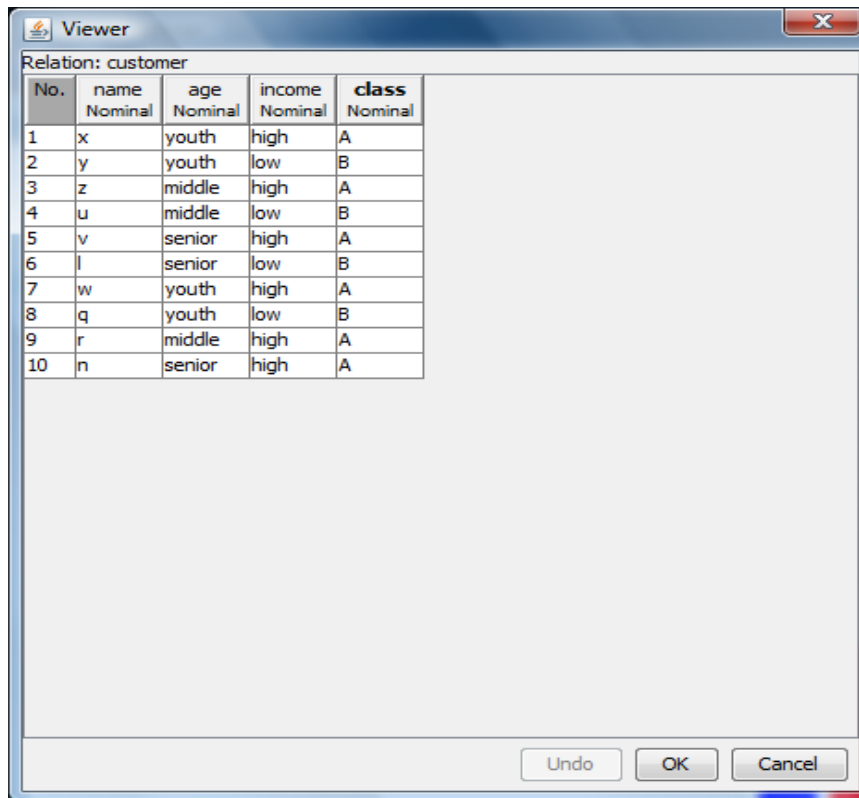
**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and 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



The screenshot shows a window titled "Viewer" with a close button (X) in the top right corner. Below the title bar, it says "Relation: customer". The main area contains a table with 10 rows and 5 columns. The columns are labeled "No.", "name", "age", "income", and "class". Each column has a data type listed below it: "Nominal" for "name", "age", and "class", and "Nominal" for "income". The data rows are as follows:

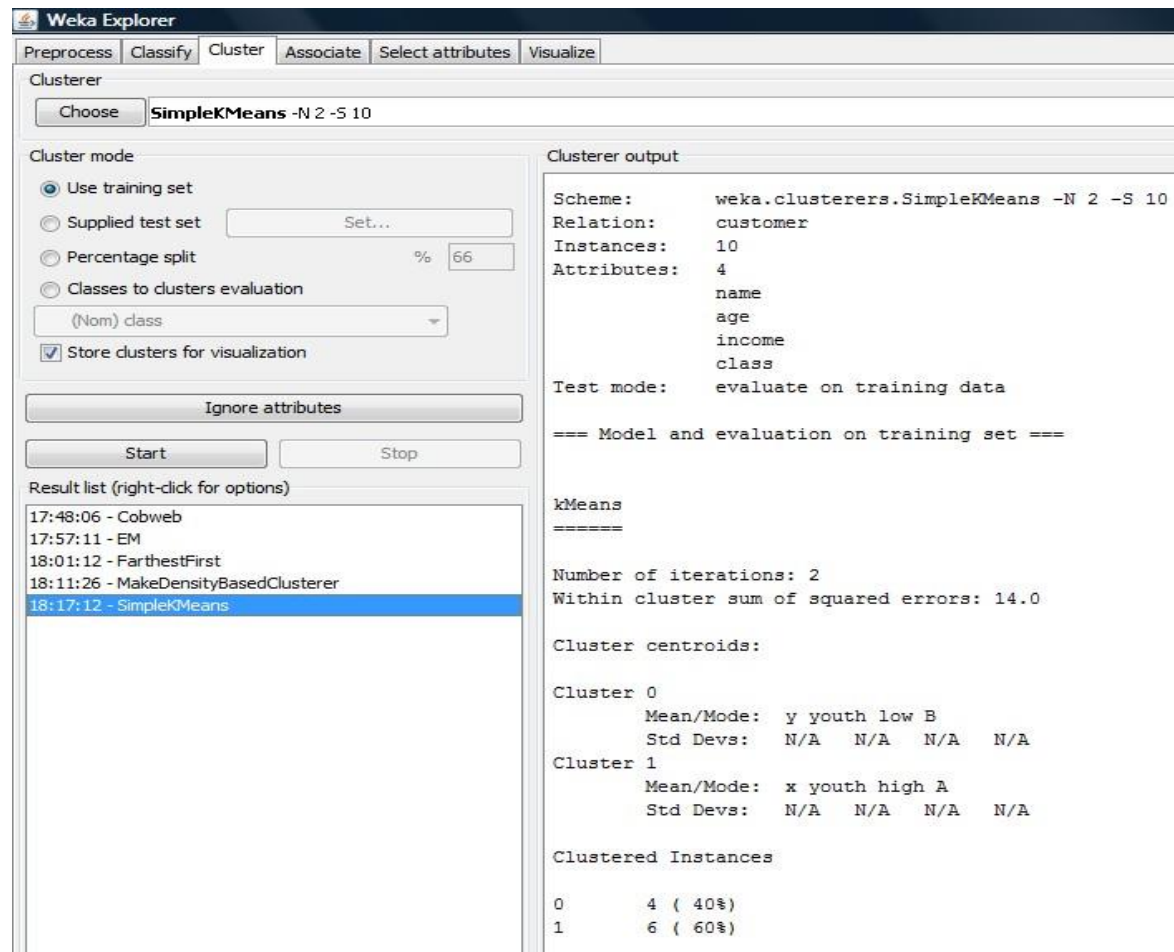
| No. | name | age    | income | class |
|-----|------|--------|--------|-------|
| 1   | x    | youth  | high   | A     |
| 2   | y    | youth  | low    | B     |
| 3   | z    | middle | high   | A     |
| 4   | u    | middle | low    | B     |
| 5   | v    | senior | high   | A     |
| 6   | l    | senior | low    | B     |
| 7   | w    | youth  | high   | A     |
| 8   | q    | youth  | low    | B     |
| 9   | r    | middle | high   | A     |
| 10  | n    | senior | high   | A     |

At the bottom of the window, there are three buttons: "Undo", "OK", and "Cancel".

### 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:



The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'SimpleKMeans -N 2 -S 10' algorithm is chosen. The 'Cluster mode' section has 'Use training set' selected, and 'Store clusters for visualization' is checked. The 'Cluster output' pane displays the following text:

```
Scheme:      weka.clusterers.SimpleKMeans -N 2 -S 10
Relation:    customer
Instances:   10
Attributes:  4
             name
             age
             income
             class

Test mode:   evaluate on training data

=== Model and evaluation on training set ===

kMeans
=====

Number of iterations: 2
Within cluster sum of squared errors: 14.0

Cluster centroids:

Cluster 0
  Mean/Mode:  y youth low B
  Std Devs:   N/A  N/A  N/A  N/A
Cluster 1
  Mean/Mode:  x youth high A
  Std Devs:   N/A  N/A  N/A  N/A

Clustered Instances

0      4 ( 40%)
1      6 ( 60%)
```

## Result:

The program has been successfully executed.

## EXPERIMENT NO:8

**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

Relation: employee

| No. | eid<br>Numeric | ename<br>Nominal | salary<br>Numeric | exp<br>Numeric | address<br>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                |

Undo OK Cancel

**Procedure:**

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

## Output:

The screenshot displays the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'MakeDensityBasedClusterer -M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -N 2 -S 10'. The 'Cluster mode' section has 'Use training set' selected, with 'Store clusters for visualization' checked. The 'Result list' shows the execution of 'MakeDensityBasedClusterer' at 18:11:26. The 'Clusterer output' pane displays the following results:

```
Cluster: 0 Prior probability: 0.6667
Attribute: eid
Normal Distribution. Mean = 107.7778 StdDev = 3.4247
Attribute: ename
Discrete Estimator. Counts = 1 1 2 2 2 2 2 1 2 1 2 2 2 (Total = 22)
Attribute: salary
Normal Distribution. Mean = 13222.2222 StdDev = 1396.645
Attribute: exp
Normal Distribution. Mean = 3.8889 StdDev = 1.0999
Attribute: address
Discrete Estimator. Counts = 1 7 3 2 (Total = 13)

Cluster: 1 Prior probability: 0.3333
Attribute: eid
Normal Distribution. Mean = 105.25 StdDev = 3.8324
Attribute: ename
Discrete Estimator. Counts = 2 2 1 1 1 1 1 2 1 2 1 1 1 (Total = 17)
Attribute: salary
Normal Distribution. Mean = 11750 StdDev = 1920.2864
Attribute: exp
Normal Distribution. Mean = 4.75 StdDev = 0.433
Attribute: address
Discrete Estimator. Counts = 3 1 1 3 (Total = 8)
Clustered Instances

0      9 ( 69%)
1      4 ( 31%)

Log likelihood: -16.52967
```

## Result:

The program has been successfully executed.