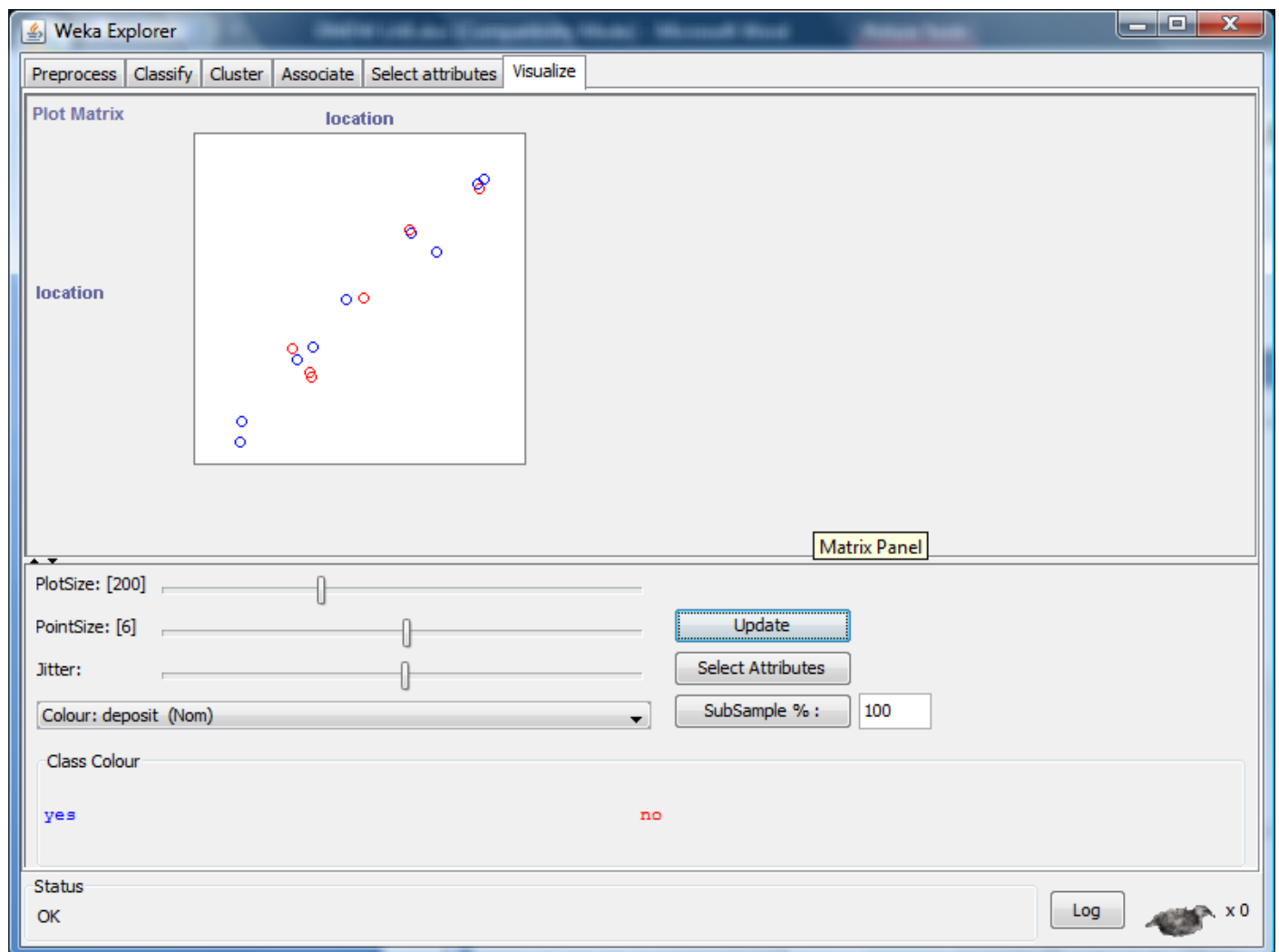


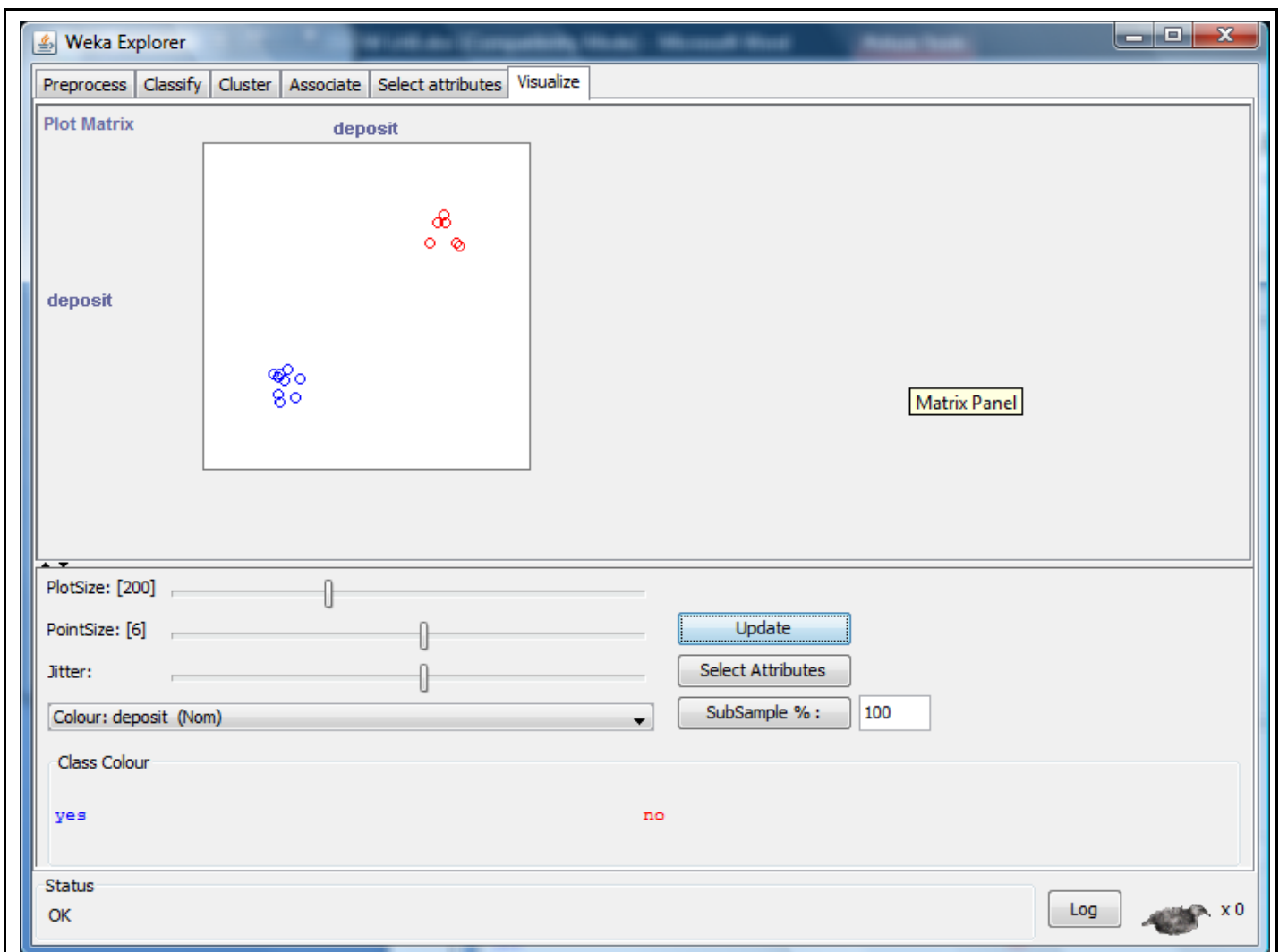
### Output:



16) After that we select the **Select Attribute button**, then select **Deposit attribute** and click 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:

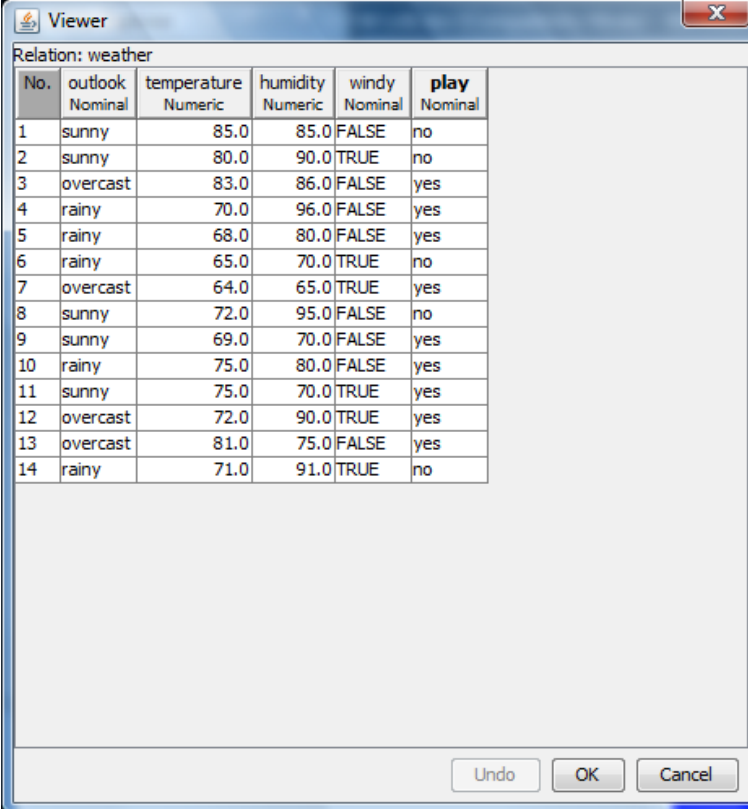
**Cross-validation**, sometimes called **rotation estimation**, is a technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in settings where the goal is prediction, and one wants to estimate how accurately a predictive model will 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



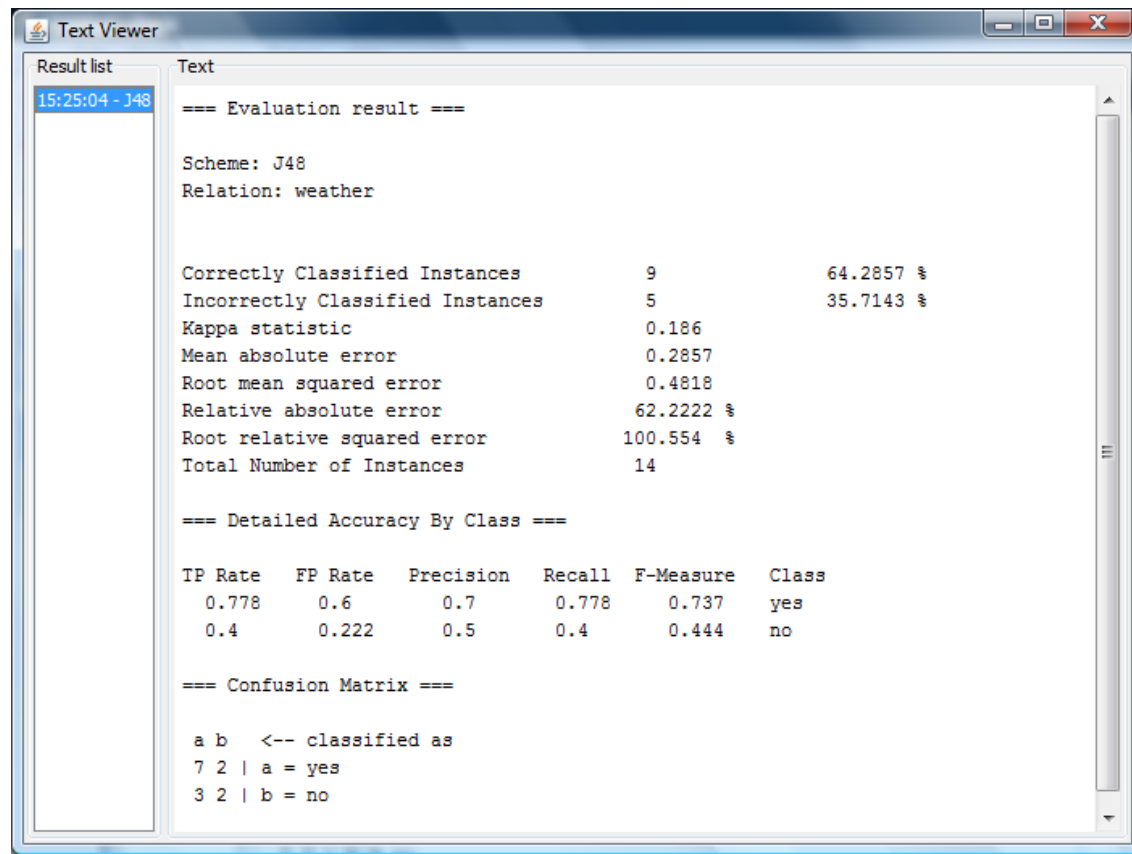
No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

#### **Procedure:**

- 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:**



```
Text Viewer
Result list
15:25:04 - J48
Text
=== Evaluation result ===

Scheme: J48
Relation: weather

Correctly Classified Instances      9           64.2857 %
Incorrectly Classified Instances    5           35.7143 %
Kappa statistic                    0.186
Mean absolute error                 0.2857
Root mean squared error             0.4818
Relative absolute error             62.2222 %
Root relative squared error         100.554 %
Total Number of Instances          14

=== Detailed Accuracy By Class ===

TP Rate    FP Rate    Precision    Recall    F-Measure    Class
0.778      0.6         0.7         0.778     0.737        yes
0.4        0.222      0.5         0.4       0.444        no

=== Confusion Matrix ===

a b  <-- classified as
7 2 | a = yes
3 2 | b = no
```

**Result:**

The program has been successfully executed.

## EXPERIMENT NO:16

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

### Description:

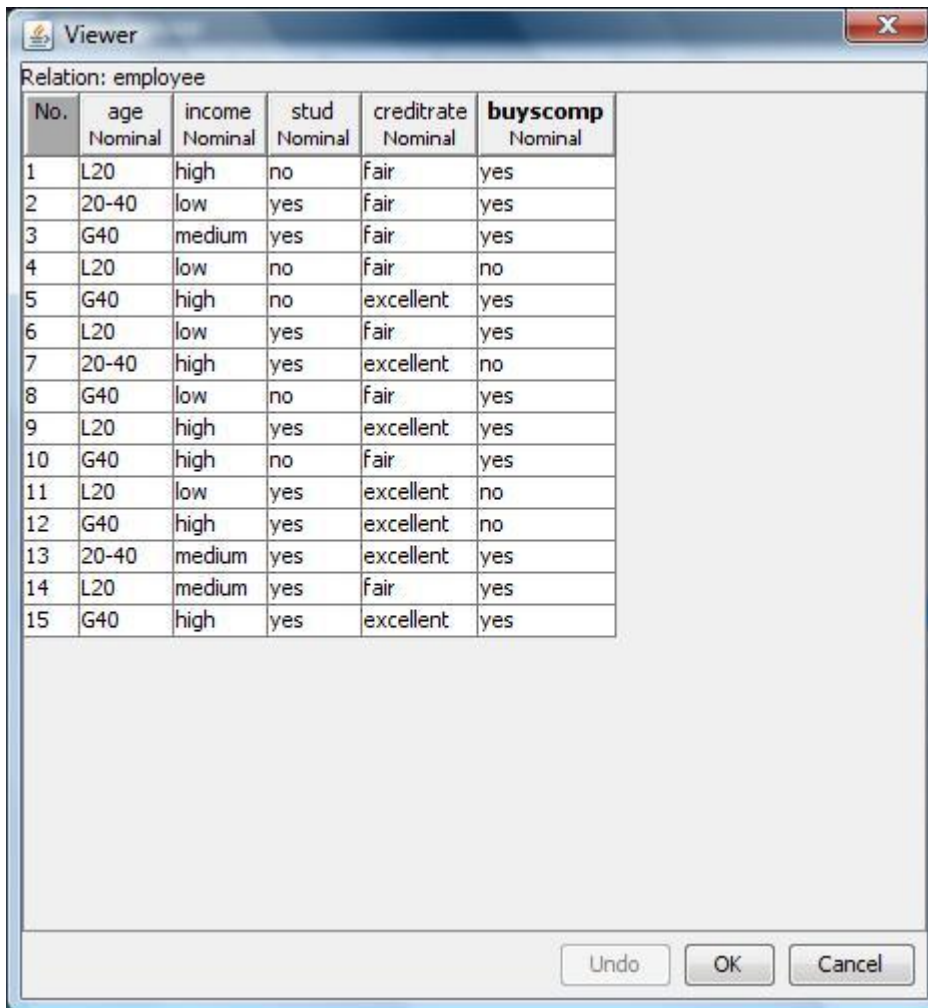
**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and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



The screenshot shows the 'Viewer' window in Weka, displaying a table titled 'Relation: employee'. The table has six columns: 'No.', 'age', 'income', 'stud', 'creditrate', and 'buyscomp'. Each column has a 'Nominal' data type. The table contains 15 rows of data. At the bottom of the window are 'Undo', 'OK', and 'Cancel' buttons.

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:

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 'Result list' shows a single entry: '17:48:06 - Cobweb'. The 'Clusterer output' pane displays the hierarchical structure of the clusters, and the 'Clustered Instances' pane shows a list of instances with their assigned cluster numbers and percentages.

**Clusterer output**

```
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**

Instance	Cluster	Percentage
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:17

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

**Description:**

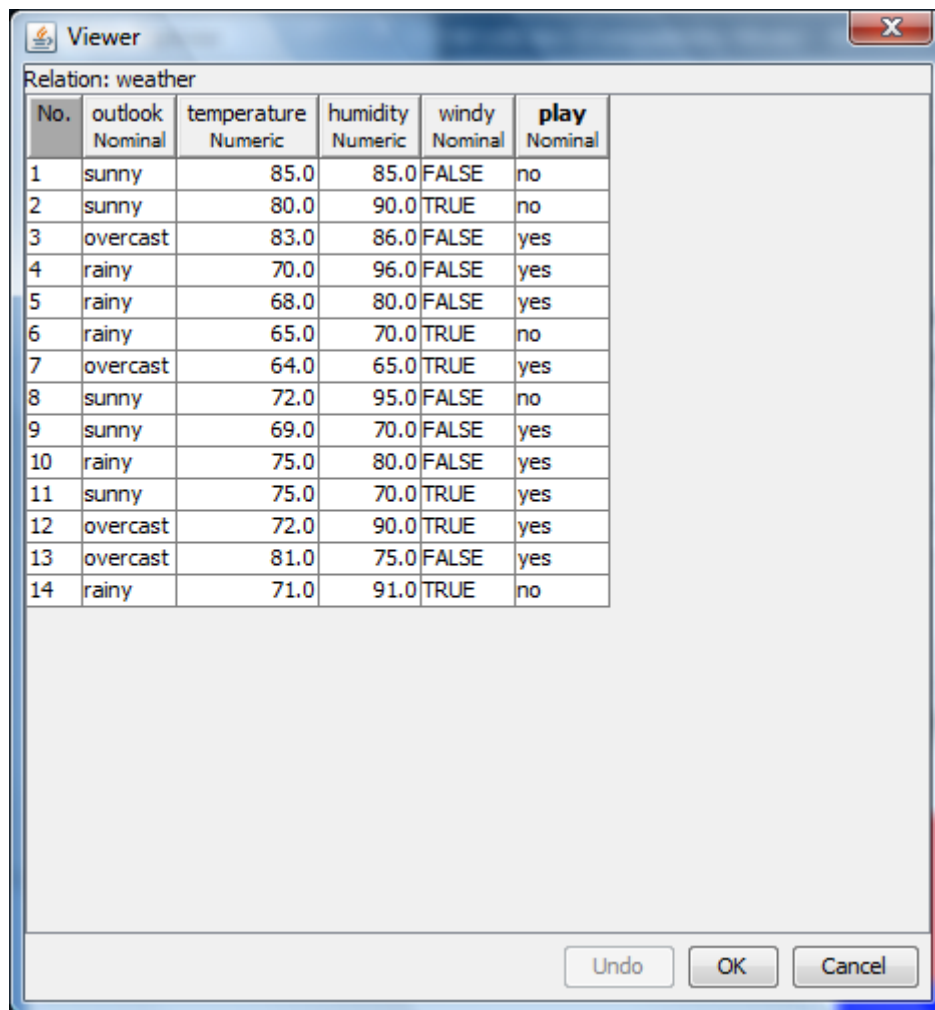
**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and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



The screenshot shows the 'Viewer' window in Weka, displaying the 'weather' relation. The window has a title bar with a close button. Below the title bar, the text 'Relation: weather' is shown. The main area contains a table with 6 columns: 'No.', 'outlook', 'temperature', 'humidity', 'windy', and 'play'. The 'outlook' column is labeled 'Nominal', 'temperature' is 'Numeric', 'humidity' is 'Numeric', 'windy' is 'Nominal', and 'play' is 'Nominal'. The table contains 14 rows of data. At the bottom right of the window are three buttons: 'Undo', 'OK', and 'Cancel'.

No.	outlook Nominal	temperature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	85.0	85.0	FALSE	no
2	sunny	80.0	90.0	TRUE	no
3	overcast	83.0	86.0	FALSE	yes
4	rainy	70.0	96.0	FALSE	yes
5	rainy	68.0	80.0	FALSE	yes
6	rainy	65.0	70.0	TRUE	no
7	overcast	64.0	65.0	TRUE	yes
8	sunny	72.0	95.0	FALSE	no
9	sunny	69.0	70.0	FALSE	yes
10	rainy	75.0	80.0	FALSE	yes
11	sunny	75.0	70.0	TRUE	yes
12	overcast	72.0	90.0	TRUE	yes
13	overcast	81.0	75.0	FALSE	yes
14	rainy	71.0	91.0	TRUE	no

### Procedure:

- 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 shows the Weka Explorer Clusterer window. The 'Clusterer' tab is selected, and the 'EM' algorithm is chosen. The 'Cluster mode' section has 'Use training set' selected. The 'Clusterer output' pane displays the results of the EM algorithm.

**Clusterer**

Choose **EM** -I 100 -N -1 -S 100 -M 1.0E-6

**Cluster mode**

- ☒ Use training set
- ☐ Supplied test set
- ☐ Percentage split %
- ☐ Classes to clusters evaluation
- 
- ☒ Store clusters for visualization

**Ignore attributes**

**Start** **Stop**

**Result list (right-click for options)**

- 17:48:06 - Cobweb
- 17:57:11 - EM

**Clusterer output**

```
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
```

## 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

Viewer

Relation: bank

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

Undo OK Cancel

#### 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:

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'FarthestFirst -N 2 -S 1'. Under 'Cluster mode', 'Use training set' is selected, and 'Store clusters for visualization' is checked. The 'Result list' on the left shows three entries: '17:48:06 - Cobweb', '17:57:11 - EM', and '18:01:12 - FarthestFirst', with the last one selected. The 'Clusterer output' pane on the right displays the following information:

```
=== Run information ===

Scheme:      weka.clusterers.FarthestFirst -N 2 -S 1
Relation:    bank
Instances:   15
Attributes:  5
              cust
              accno
              bankname
              location
              deposit

Test mode:   evaluate on training data

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

FarthestFirst
=====

Cluster centroids:

Cluster 0
    male 0101 sbi hyd yes
Cluster 1
    female 0102 hdfc jmd no

Clustered Instances

0      8 ( 53%)
1      7 ( 47%)
```

## Result:

The program has been successfully executed.

## EXPERIMENT NO:19

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

### Description:

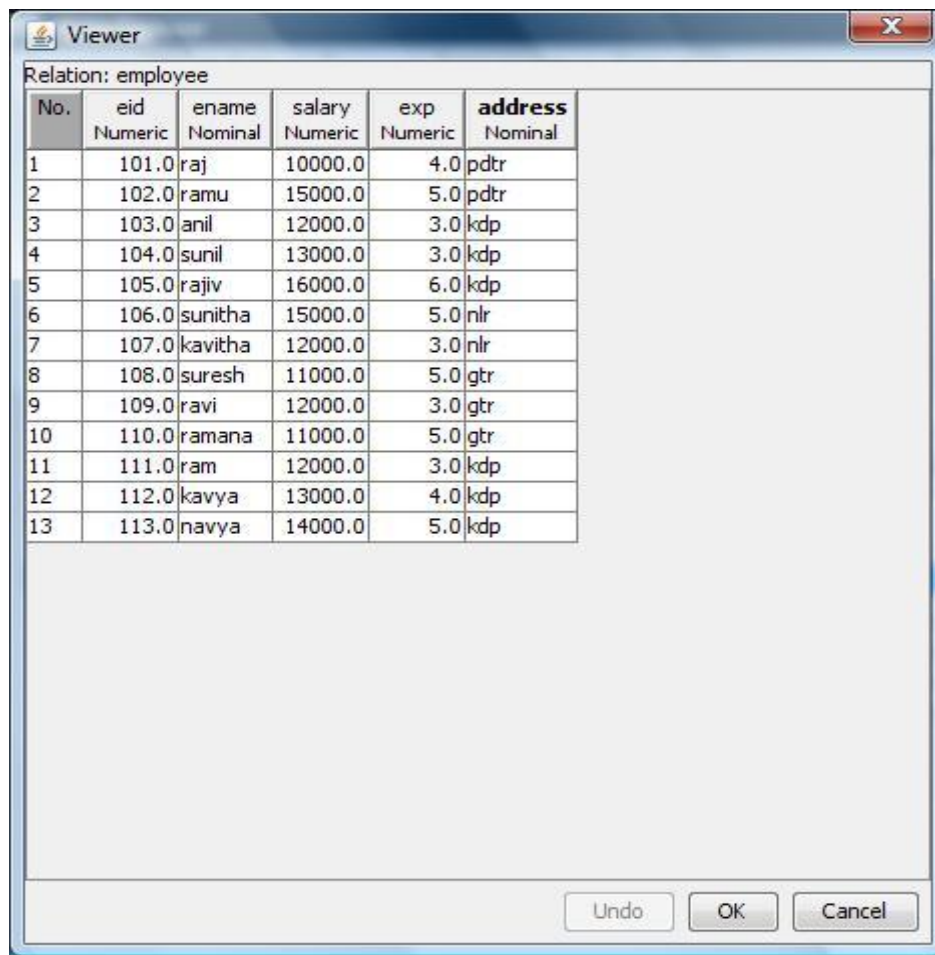
**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and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



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

#### Procedure:

- 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:

The screenshot shows the Weka Explorer interface with the 'Cluster' tab selected. The 'Clusterer' dropdown is set to 'MakeDensityBasedClusterer' with command line options: `-M 1.0E-6 -W weka.clusterers.SimpleKMeans -- -N 2 -S 10`. The 'Cluster mode' section has 'Use training set' selected. The 'Cluster output' pane displays the results for two clusters.

**Cluster mode**

- ☒ Use training set
- ☐ Supplied test set (Set...)
- ☐ Percentage split (% 66)
- ☐ Classes to clusters evaluation (Nom) address
- ☒ Store clusters for visualization

**Clusterer output**

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 list (right-click for options)**

- 17:48:06 - Cobweb
- 17:57:11 - EM
- 18:01:12 - FarthestFirst
- 18:11:26 - MakeDensityBasedClusterer

## Result:

The program has been successfully executed.

## EXPERIMENT NO:20

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

### **Description:**

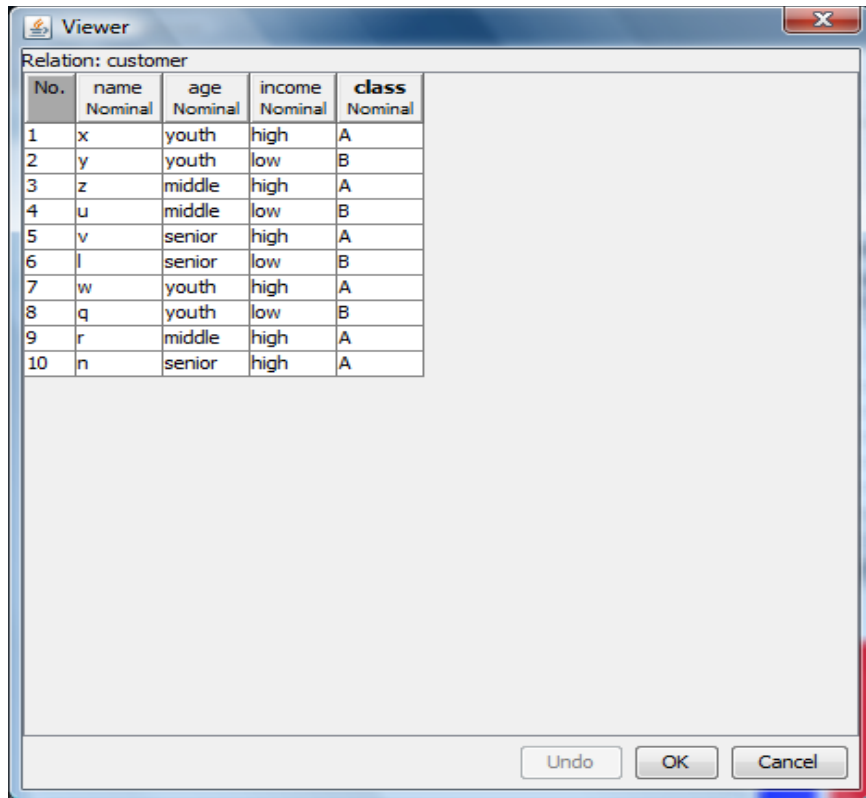
**Cluster analysis** or **clustering** is the task of assigning a set of objects into groups (called **clusters**) so that the objects in the same cluster are more similar (in some sense or another) to each other than to those in other clusters. Clustering is a main task of explorative data mining, and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



The screenshot shows a 'Viewer' window in Weka. The title bar says 'Viewer'. Below the title bar, it says 'Relation: customer'. The main area contains a table with 5 columns: 'No.', 'name', 'age', 'income', and 'class'. Each column has a data type listed below it: 'No.' is 'Nominal', 'name' is 'Nominal', 'age' is 'Nominal', 'income' is 'Nominal', and 'class' is 'Nominal'. The table contains 10 rows of data. At the bottom right of the window, there are three buttons: 'Undo', 'OK', and 'Cancel'.

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

### 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 application with the 'Cluster' tab selected. The 'SimpleKMeans' algorithm is chosen with parameters -N 2 -S 10. The 'Cluster mode' section has 'Use training set' selected, and 'Store clusters for visualization' is checked. The 'Cluster output' pane displays the following information:

```
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%)
```

The 'Result list' on the left shows a list of recent operations, with '18:17:12 - SimpleKMeans' selected and highlighted in blue.

## Result:

The program has been successfully executed.