

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:

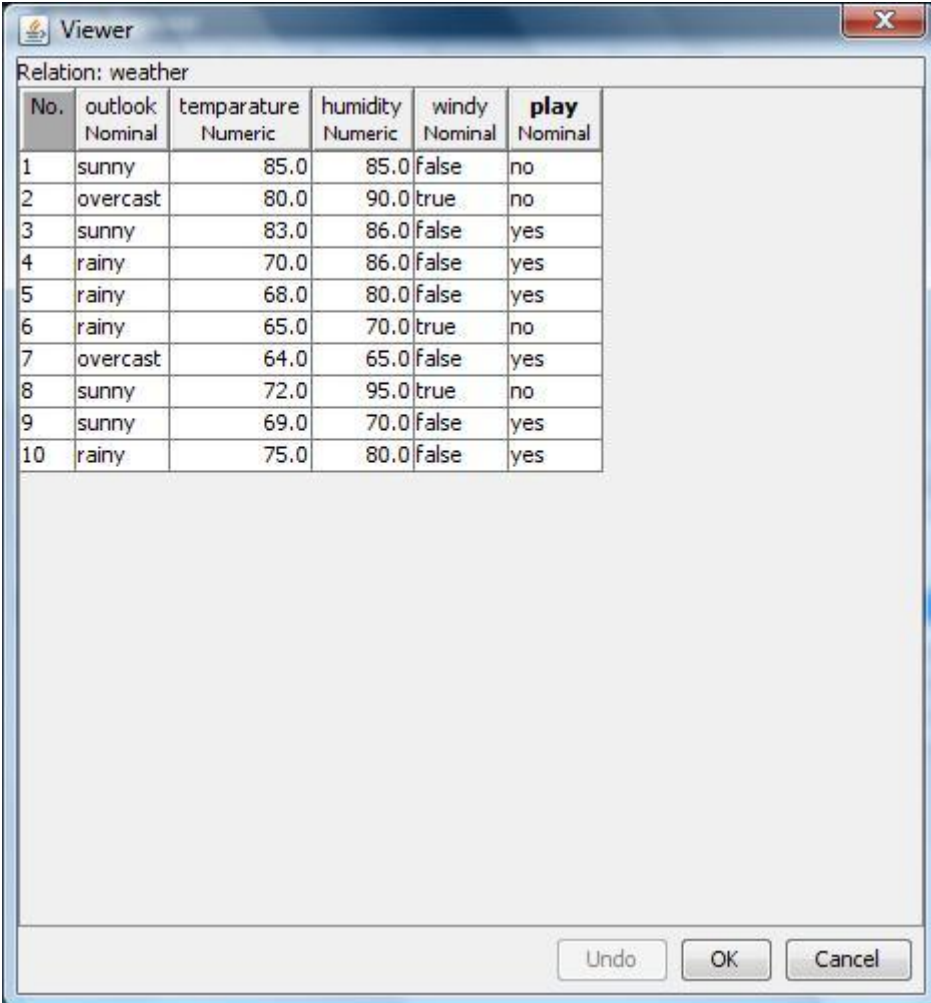
- 1) Open Start → Programs → Accessories → Notepad
- 2) Type the following training data set with the help of Notepad for Employee Table.

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

Undo OK Cancel

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:

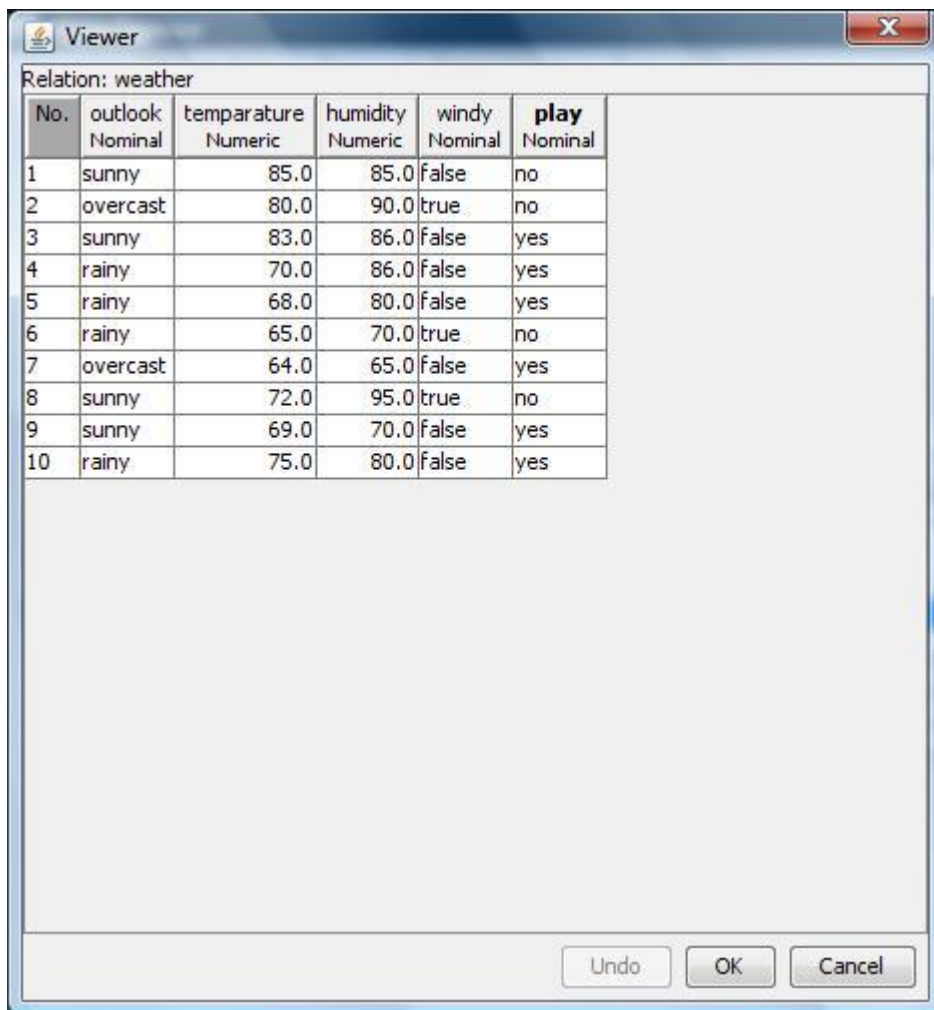
- 1) Open Start → Programs → Accessories → Notepad
- 2) Type the following training data set with the help of Notepad for Weather Table.

```
@relation weather
@attribute outlook {sunny,rainy,overcast}
@attribute temperature 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



Relation: weather

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

Undo OK Cancel

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 temprature 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.

Viewer

Relation: weather

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

Undo OK 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 **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:

Viewer

Relation: weather-weka.filters.unsupervised.attribute.Add-Nclimate-LNominal-Clast

No.	outlook Nominal	temparature Numeric	humidity Numeric	windy Nominal	play Nominal	climate 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	

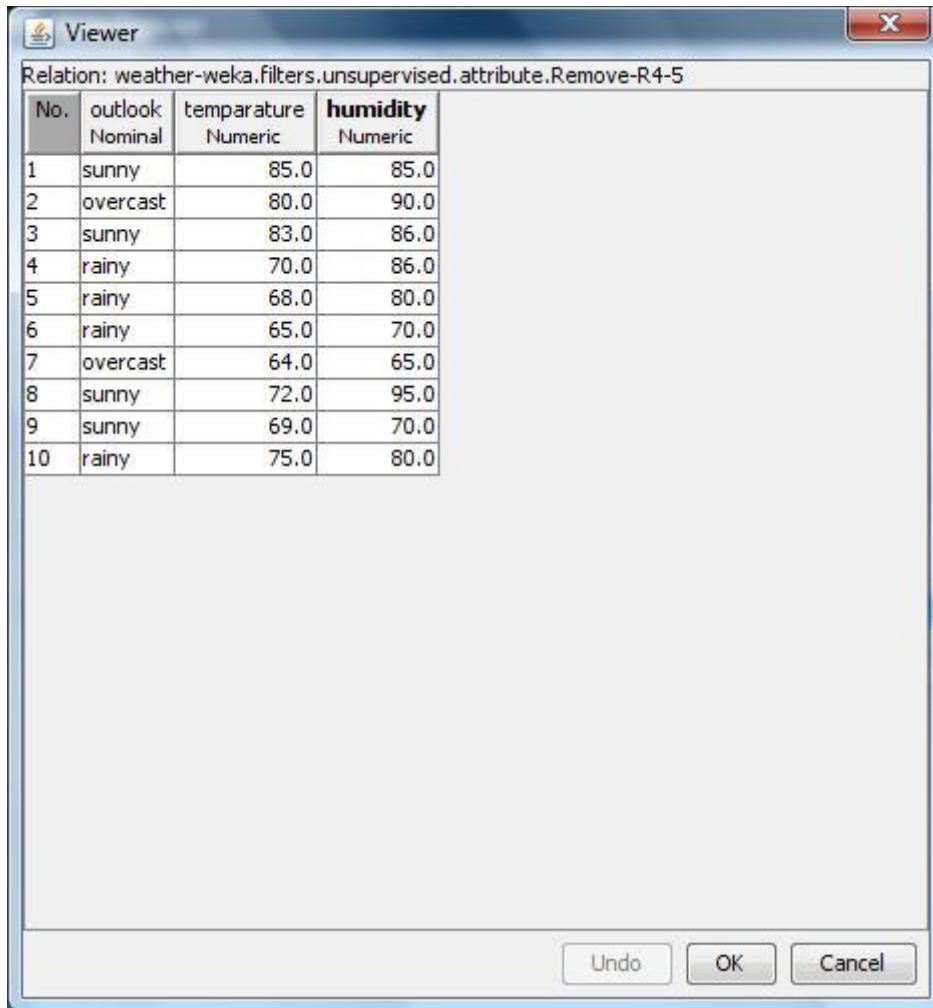
Undo OK Cancel

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:



Relation: weather-weka.filters.unsupervised.attribute.Remove-R4-5

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

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 **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 **temperature, 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 TEMPERATURE, HUMIDITY:

Viewer

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

No.	outlook Nominal	temparature Numeric	humidity Numeric	windy Nominal	play Nominal
1	sunny	1.0	0.6666...	false	no
2	overcast	0.7619047...	0.8333...	true	no
3	sunny	0.9047619...	0.7	false	yes
4	rainy	0.2857142...	0.7	false	yes
5	rainy	0.1904761...	0.5	false	yes
6	rainy	0.0476190...	0.1666...	true	no
7	overcast	0.0	0.0	false	yes
8	sunny	0.3809523...	1.0	true	no
9	sunny	0.2380952...	0.1666...	false	yes
10	rainy	0.5238095...	0.5	false	yes

Undo OK Cancel

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:

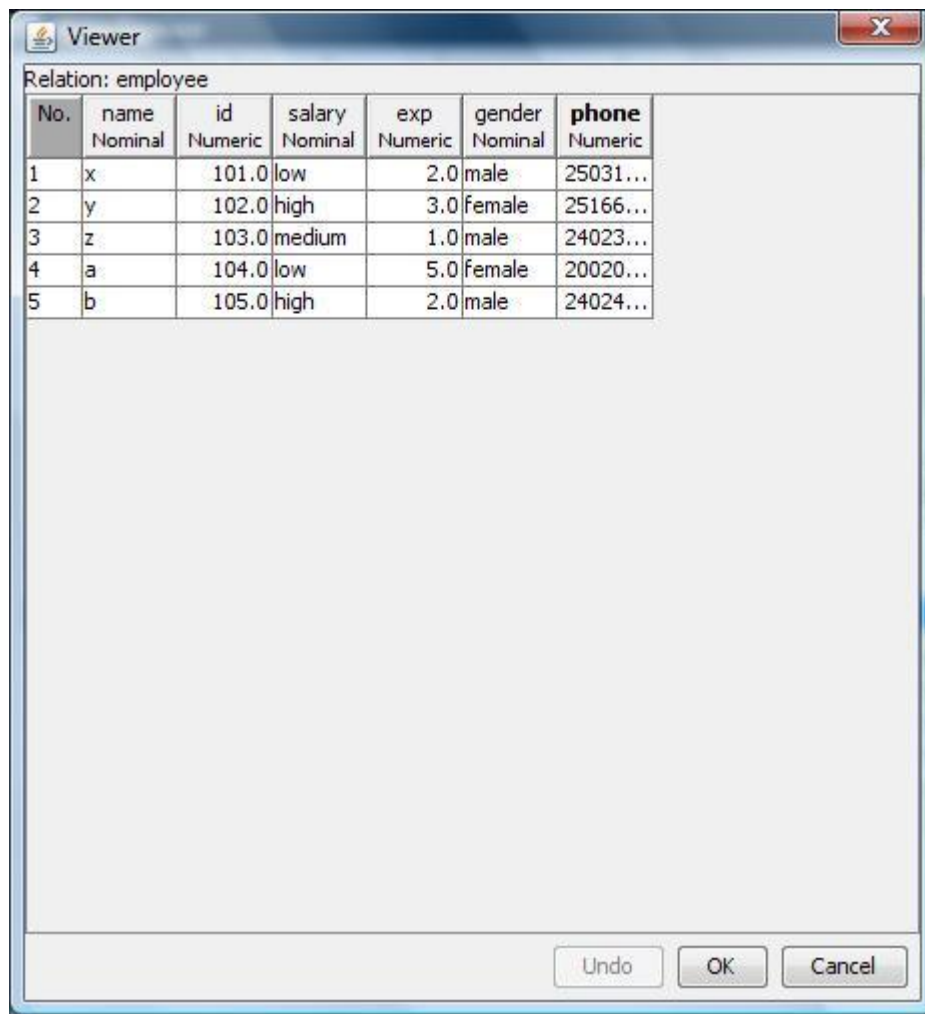
- 1) Open Start → Programs → Accessories → Notepad
- 2) Type the following training data set with the help of Notepad for Employee Table.

```
@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



The screenshot shows a 'Viewer' window in Weka. The title bar says 'Viewer'. Below the title bar, it says 'Relation: employee'. The main area contains a table with 7 columns: 'No.', 'name', 'id', 'salary', 'exp', 'gender', and 'phone'. Each column has a data type listed below it: 'No.' is Numeric, 'name' is Nominal, 'id' is Numeric, 'salary' is Nominal, 'exp' is Numeric, 'gender' is Nominal, and 'phone' is 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...

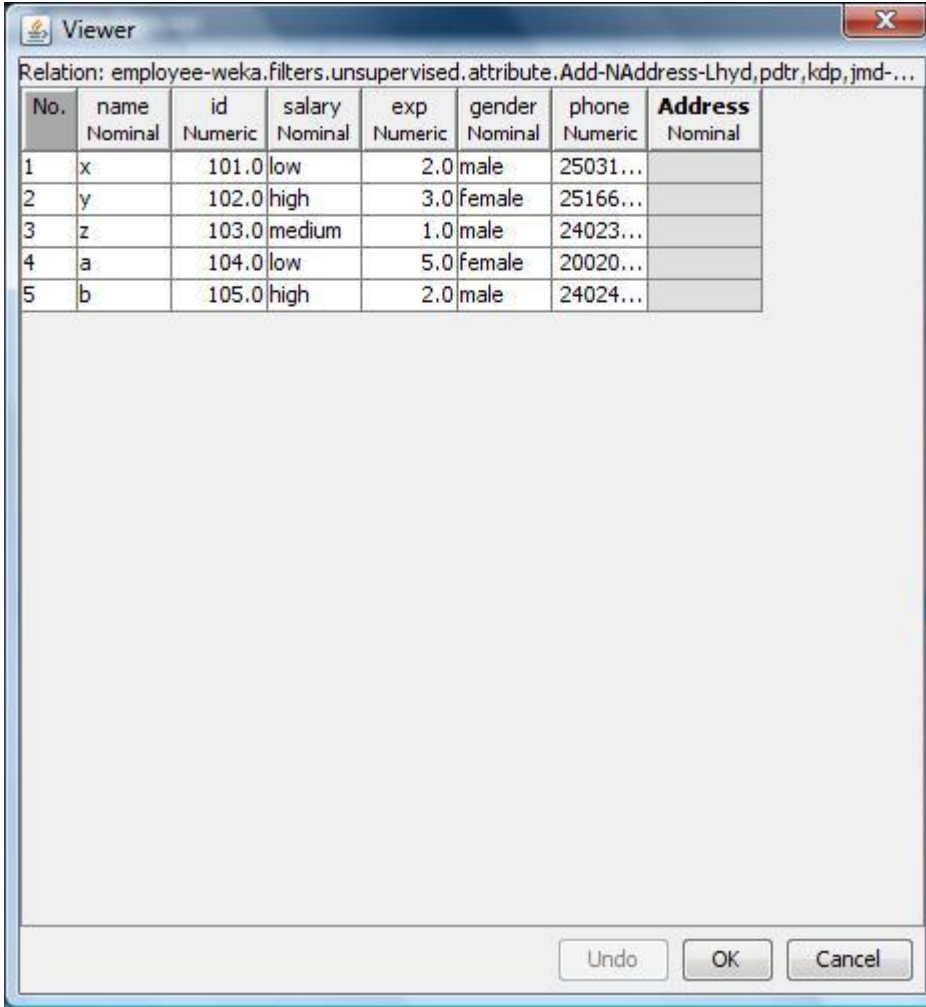
At the bottom of the window, there are three buttons: 'Undo', 'OK', and 'Cancel'.

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:



The screenshot shows the Weka Viewer window. The title bar says 'Viewer'. The relation path is 'employee-weka.filters.unsupervised.attribute.Add-NAddress-Lhyd,pdtr,kdp,jmd-...'. The table has 8 columns: No., name, id, salary, exp, gender, phone, and Address. The first five rows of data are visible.

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...	

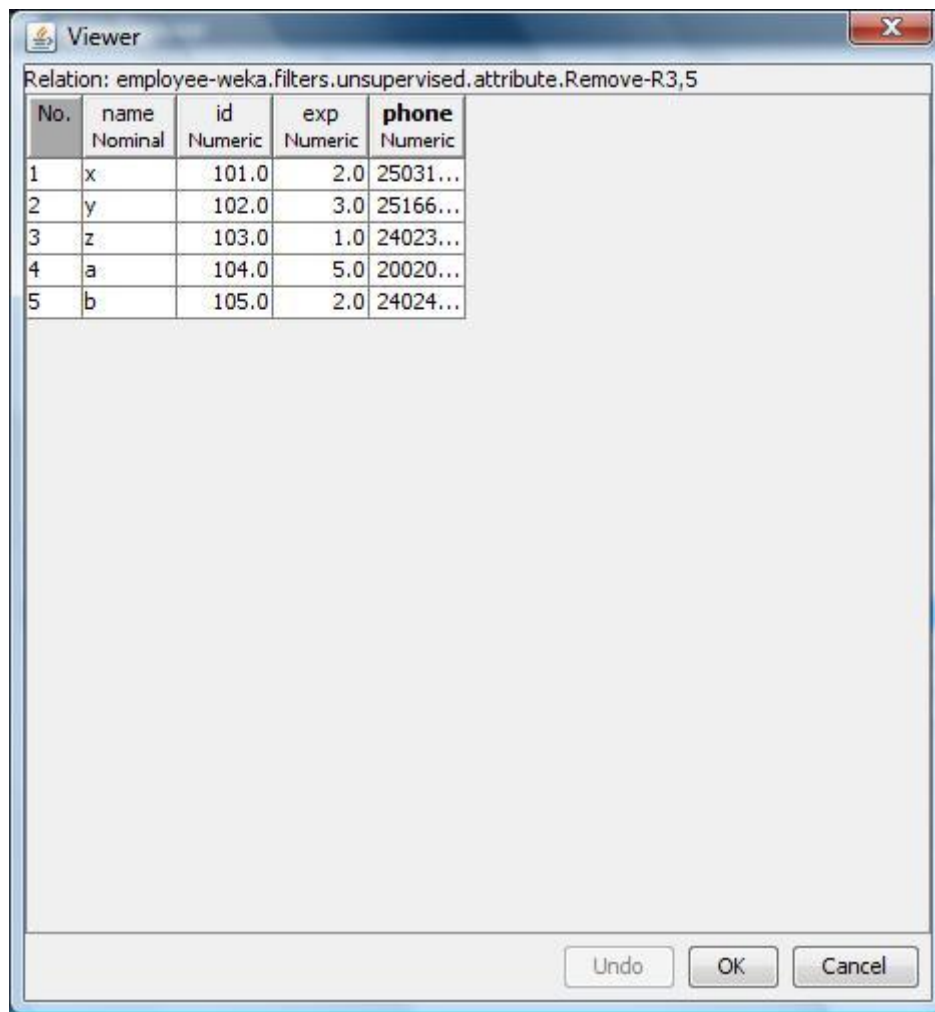
At the bottom of the window are buttons for 'Undo', 'OK', and 'Cancel'.

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:



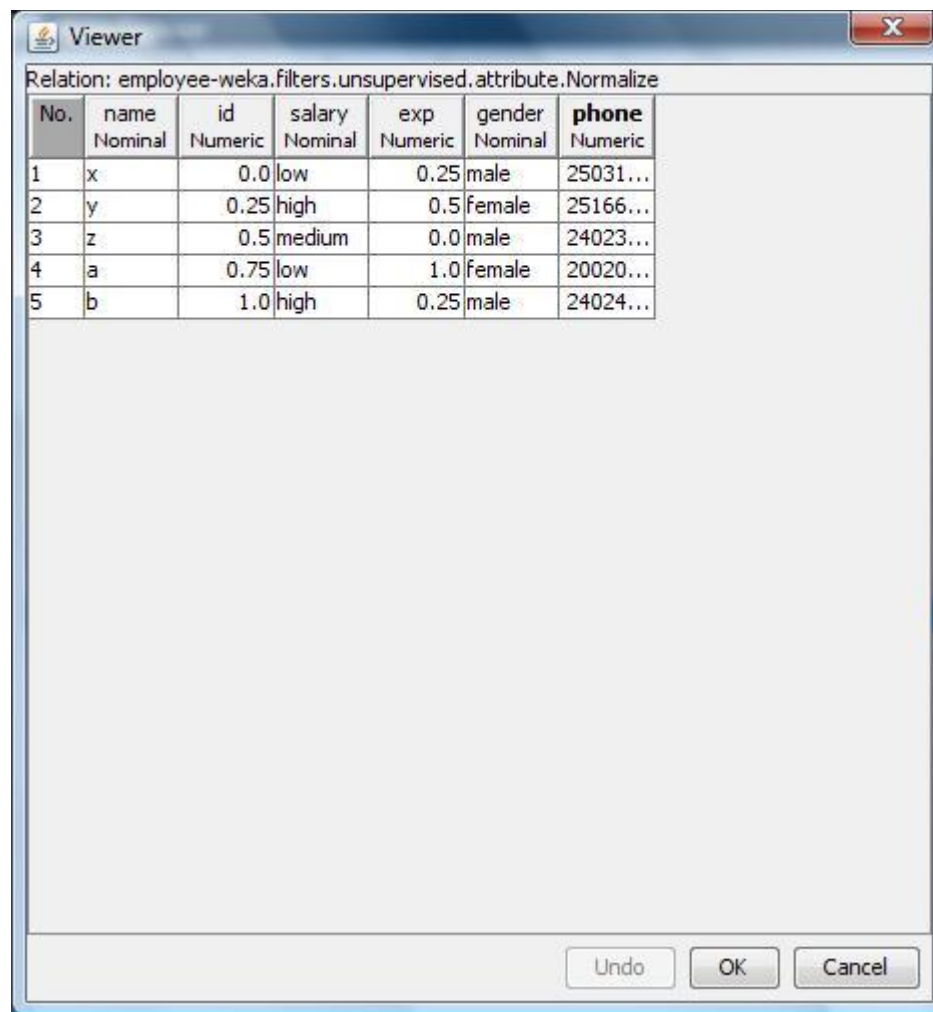
No.	name Nominal	id Numeric	exp Numeric	phone Numeric
1	x	101.0	2.0	25031...
2	y	102.0	3.0	25166...
3	z	103.0	1.0	24023...
4	a	104.0	5.0	20020...
5	b	105.0	2.0	24024...

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:



The screenshot shows a 'Viewer' window from the Weka data mining software. The title bar reads 'Relation: employee-weka.filters.unsupervised.attribute.Normalize'. The window contains a table with 7 columns: 'No.', 'name', 'id', 'salary', 'exp', 'gender', and 'phone'. Each column has a data type specified below it: 'name' is Nominal, 'id' is Numeric, 'salary' is Nominal, 'exp' is Numeric, 'gender' is Nominal, and 'phone' is Numeric. There are 5 rows of data. At the bottom of the window are 'Undo', 'OK', and 'Cancel' buttons.

No.	name Nominal	id Numeric	salary Nominal	exp Numeric	gender Nominal	phone 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...

Result:

This program has been successfully executed.