

CSA1618 DWDM-DE

EXPERIMENT-24

DATA SEGMENTATION BY COBWEB – HIERARCHIAL CLUSTERING ALGORITHM USING WEKA TOOL

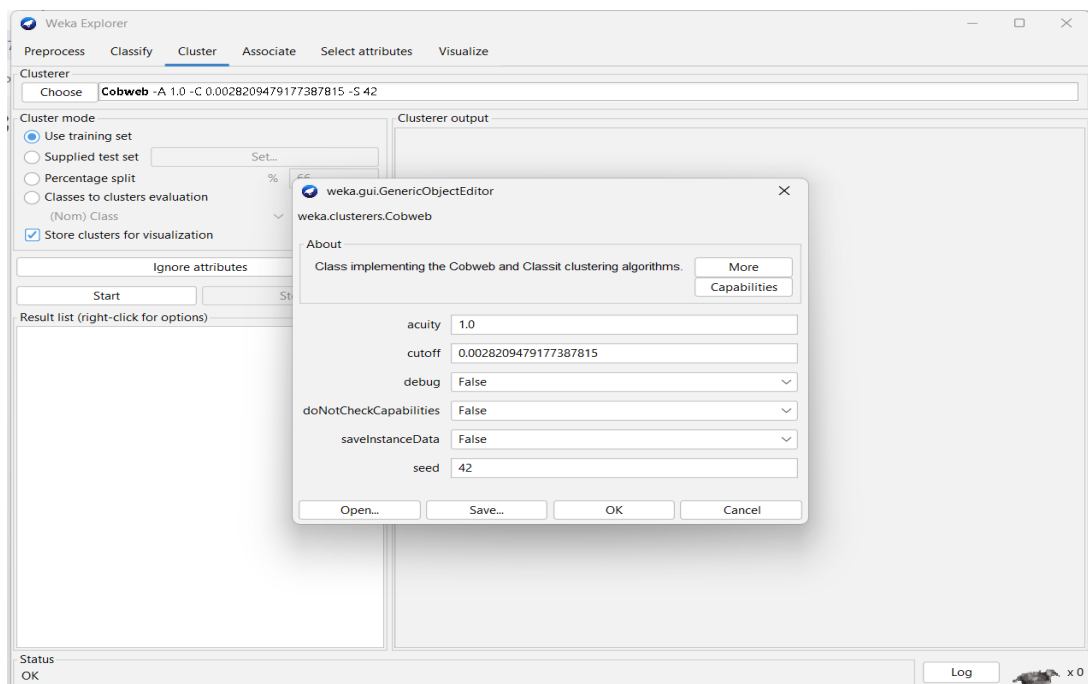
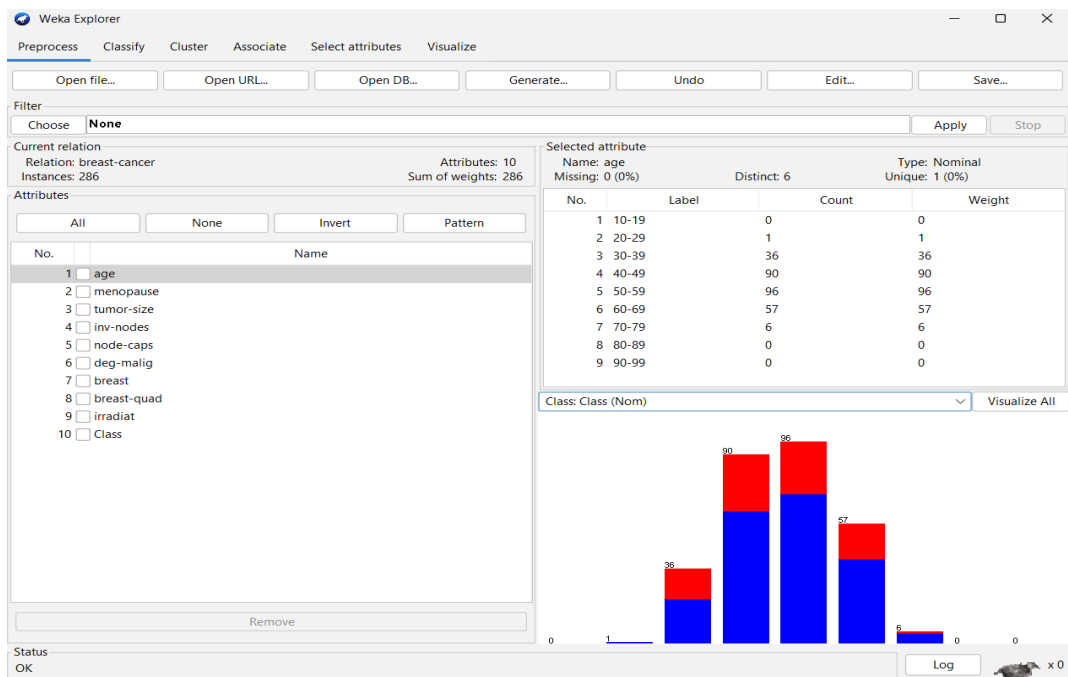
AIM:

To create data segmentation by cobweb-hierarchial clustering algorithm using weka tool.

PROCEDURE:

1. Download and install WEKA.
2. Open WEKA and Choose "Explorer" from the main menu.
3. Under Preprocess, Click on the open file button and select the dataset. Ensure that categorical attributes are in the correct format (nominal, not numeric).
4. Click on the "Cluster" tab. In the Cluster mode section, select "Use training set".
5. Click "Choose" (next to the cluster algorithm) and Select **Cobweb** (found under weka.clusterers).
6. Click on "**Cobweb**" to configure parameters: **acuity (Default = 1.0)**: Controls cluster granularity (higher values lead to fewer clusters). **cutoff (Default = 0.002)**: Defines the threshold for merging clusters (higher values result in more clusters).
7. Click "**OK**" and then "**Start**" to begin clustering. Click "Visualize" to see how the clusters are distributed. Save the file.





Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer
Choose **Cobweb -A 1.0 -C 0.0028209479177387815 -S 42**

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

Ignore attributes

Start Stop

Result list (right-click for options)
15:46:53 - Cobweb

Status OK

Clusterer output

```

=== Run information ===

Scheme:      weka.clusterers.Cobweb -A 1.0 -C 0.0028209479177387815 -S 42
Relation:    breast-cancer
Instances:   286
Attributes:  10
age
menopause
tumor-size
inv-nodes
node-caps
deg-malig
breast
breast-quad
irradiat
Class
Test mode:   evaluate on training data

=== Clustering model (full training set) ===

Number of merges: 100
Number of splits: 74
Number of clusters: 417

node 0 [286]
| node 1 [106]
| | node 2 [21]
| | | leaf 3 [1]
| | node 2 [21]
| | | node 4 [4]
| | | leaf 5 [1]
| | | node 4 [4]
| | | node 6 [2]
  
```

Log x0

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer
Choose **Cobweb -A 1.0 -C 0.0028209479177387815 -S 42**

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

Ignore attributes

Start Stop

Result list (right-click for options)
15:46:53 - Cobweb

Status OK

Clusterer output

```

| | | | node 412 [2]
| | | | leaf 414 [1]
| | | node 411 [5]
| | | leaf 415 [2]
| | | node 411 [5]
| | | leaf 416 [1]

Time taken to build model (full training data) : 0.15 seconds

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

Clustered Instances

5      1 ( 0%)
7      1 ( 0%)
8      1 ( 0%)
9      1 ( 0%)
12     1 ( 0%)
13     1 ( 0%)
14     1 ( 0%)
15     1 ( 0%)
16     1 ( 0%)
17     1 ( 0%)
20     1 ( 0%)
21     1 ( 0%)
23     1 ( 0%)
24     1 ( 0%)
25     2 ( 1%)
29     1 ( 0%)
30     1 ( 0%)
31     1 ( 0%)
32     1 ( 0%)
  
```

Log x0

OBSERVATION:

==== Run information ====

Scheme: weka.clusterers.Cobweb -A 1.0 -C 0.0028209479177387815 -S 42

Relation: breast-cancer

Instances: 286

Attributes: 10

age

menopause

tumor-size

inv-nodes

node-caps

deg-malig

breast

breast-quad

irradiat

Class

Test mode: evaluate on training data

==== Clustering model (full training set) ====

Number of merges: 100

Number of splits: 74

Number of clusters: 417

node 0 [286]

| node 1 [106]

| | node 2 [21]

| | | leaf 3 [1]

| | node 2 [21]

| | | node 4 [4]

| | | | leaf 5 [1].....

| | | node 411 [5]

| | | | leaf 415 [2]

| | | node 411 [5]

| | | | leaf 416 [1]

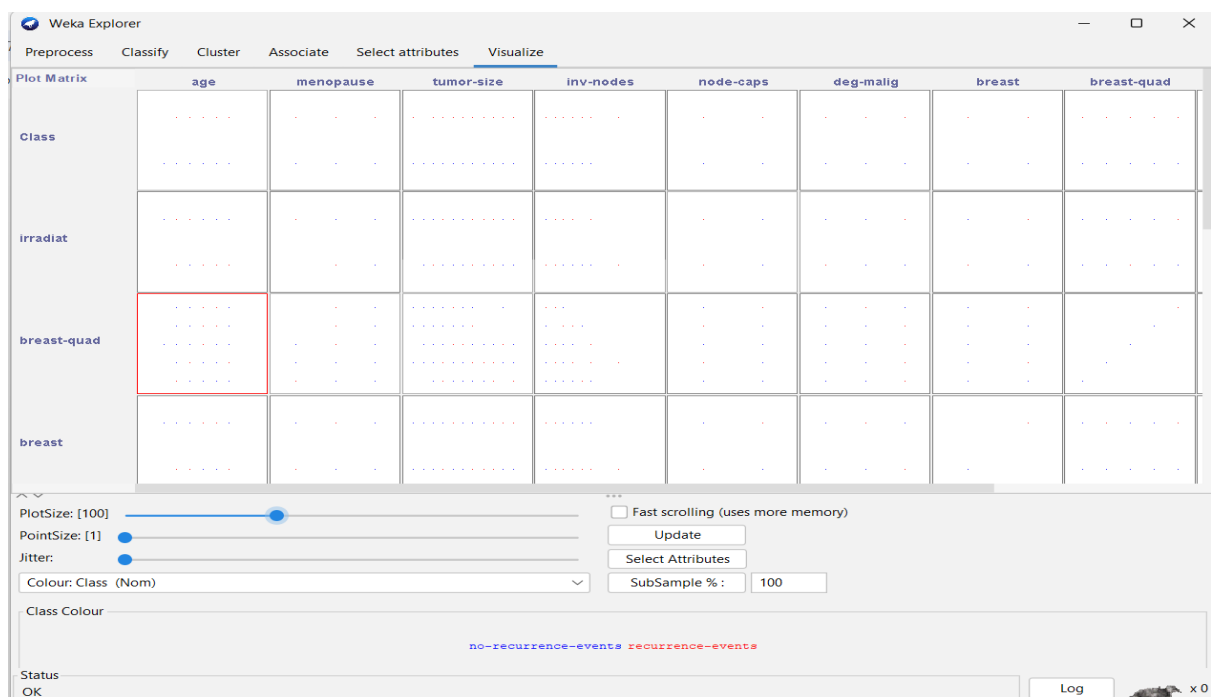
Time taken to build model (full training data) : 0.15 seconds

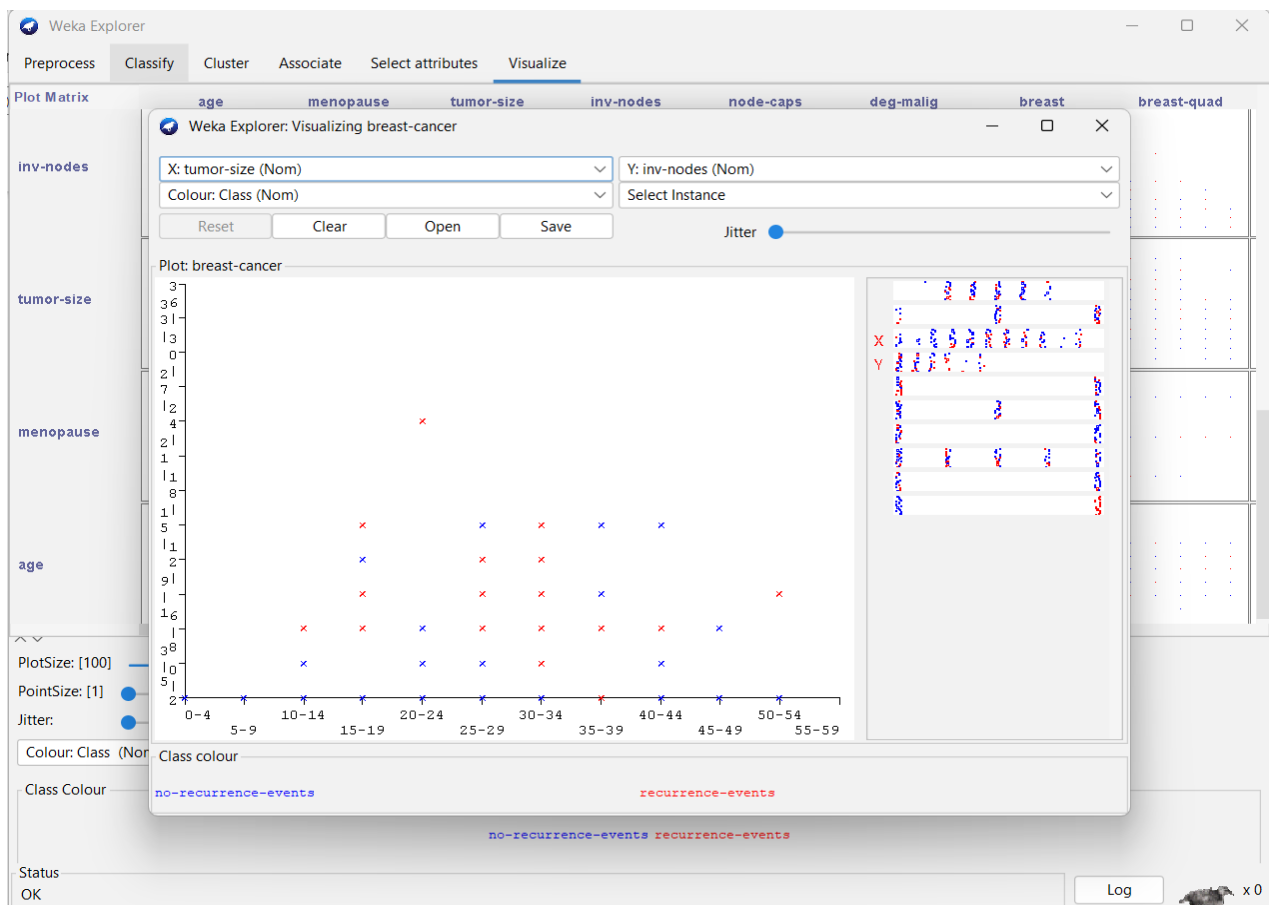
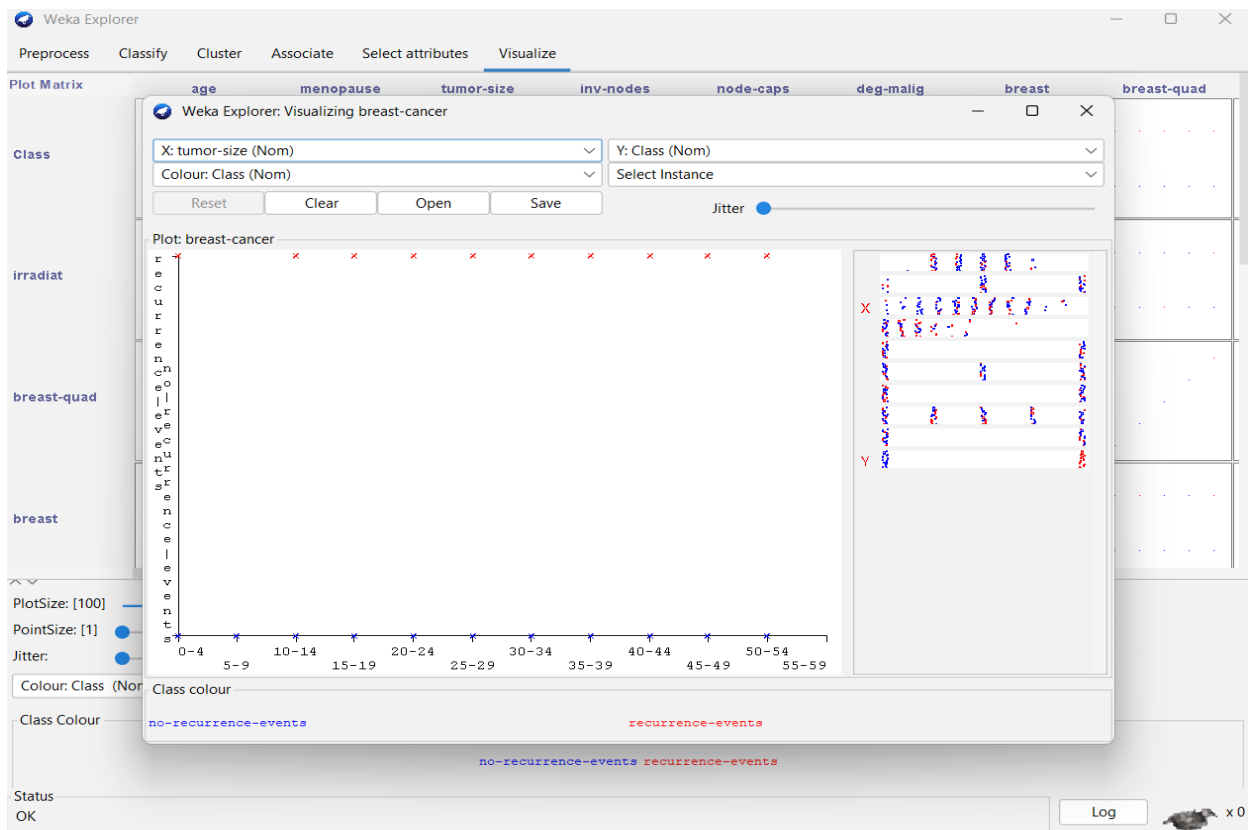
=== Model and evaluation on training set ===

Clustered Instances

5 1 (0%)
7 1 (0%)
8 1 (0%)
9 1 (0%)
12 1 (0%)
13 1 (0%)
14 1 (0%)
15 1 (0%)
16 1 (0%)
17 1 (0%)
20 1 (0%)
21 1 (0%)
23 1 (0%)
24 1 (0%)
25 2 (1%).....

PLOT:





RESULT:

Thus, the data analysis of cobweb hierarchial clustering algorithm using weka tools has been analyzed and observed successfully.