

NAME: ATHARVA AVINASH AMLE

ROLL NO: 1

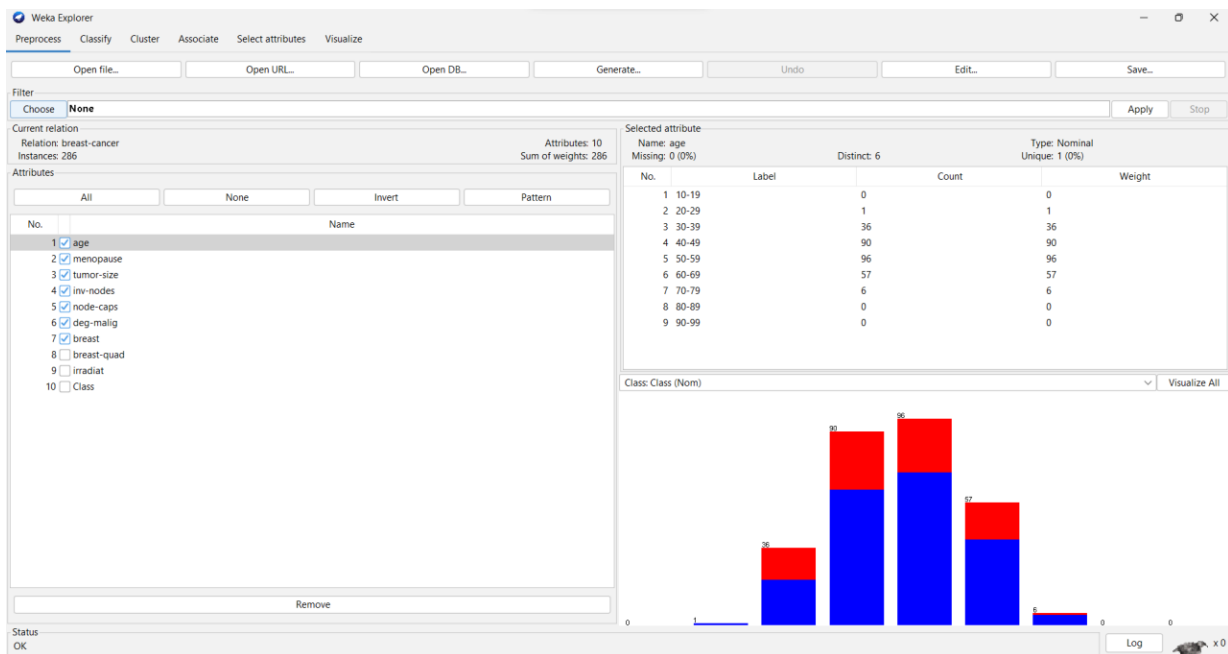
DIV: TE3

BATCH: A

EXPERIMENT - 6

Weka Classification using Naive Bayesian

Explorer:



Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options:

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 10
- ☐ Percentage split %: 66
-

(Nom) Class: **breast-cancer**

Start

Result list (right-click for options):

- 144916 - treesJ48
- 145257 - bayesNaiveBayes**

Classifier output:

```

=== Run information ===

Scheme:      weka.classifiers.bayes.NaiveBayes
Relation:    breast-cancer
Instances:   286
Attributes:  10
  age
  menopause
  tumor-size
  inv-nodes
  node-caps
  deg-malign
  breast
  breast-quadrant
  irradiat
  Class

Test mode:   evaluate on training data


=== Classifier model (full training set) ===

Naive Bayes Classifier

Attribute      no-recurrence-events    recurrence-events
                (0.7)          (0.3)
=====

age
10-19          1.0          1.0
20-29          2.0          1.0
30-39          22.0         16.0
40-49          64.0         28.0
50-59          72.0         26.0
60-69          41.0         18.0
70-79          6.0          2.0
80-89          1.0          1.0
90-99          1.0          1.0
[total]       210.0        94.0

```

Status: OK  x 0

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options:

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds: 10
- ☐ Percentage split %: 66
-

(Nom) Class: **breast-cancer**

Start

Result list (right-click for options):

- 144916 - treesJ48
- 145257 - bayesNaiveBayes**

Classifier output:

```


menopause
1140          6.0          3.0
ge40          95.0         36.0
premeno      103.0        49.0
[total]      204.0        88.0

tumor-size
0-4           8.0          2.0
5-9           5.0          1.0
10-14         28.0         2.0
15-19         24.0         8.0
20-24         35.0        17.0
25-29         37.0        19.0
30-34         36.0        26.0
35-39         13.0         8.0
40-44         17.0         7.0
45-49         3.0          2.0
50-54         6.0          4.0
55-59         1.0          1.0
[total]      213.0        97.0

inv-nodes
0-2           168.0        47.0
3-5           20.0         18.0
6-8           8.0          11.0
9-11          5.0          7.0
12-14         2.0          3.0
15-17         4.0          4.0
18-20         1.0          1.0
21-23         1.0          1.0
24-26         1.0          2.0
27-29         1.0          1.0
30-32         1.0          1.0
33-35         1.0          1.0
36-39         1.0          1.0
[total]      214.0        98.0

node-caps

```

Status: OK  x 0

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds
- ☐ Percentage split %

(Nom) Class

Result list (right-click for options)

- 14:49:16 - treesJ48
- 14:52:57 - bayesNaiveBayes**

Classifier output

```
node-caps
yes          26.0          32.0
no           172.0         52.0
[total]      198.0         84.0

deg-malig
1            60.0          13.0
2            103.0         29.0
3            41.0          46.0
[total]      204.0         88.0

breast
left         104.0         50.0
right        99.0          37.0
[total]      203.0         87.0

breast-quad
left_up      72.0          27.0
left_low     76.0          36.0
right_up     21.0          14.0
right_low    19.0           7.0
central      10.0           5.0
[total]      206.0         89.0

irradiat
yes          38.0          32.0
no           165.0         55.0
[total]      203.0         87.0

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds
```

Status: OK x0

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options

- ☒ Use training set
- ☐ Supplied test set
- ☐ Cross-validation Folds
- ☐ Percentage split %

(Nom) Class

Result list (right-click for options)

- 14:49:16 - treesJ48
- 14:52:57 - bayesNaiveBayes**

Classifier output

```
node-caps
yes          38.0          32.0
no           165.0         55.0
[total]      203.0         87.0

Time taken to build model: 0 seconds

=== Evaluation on training set ===

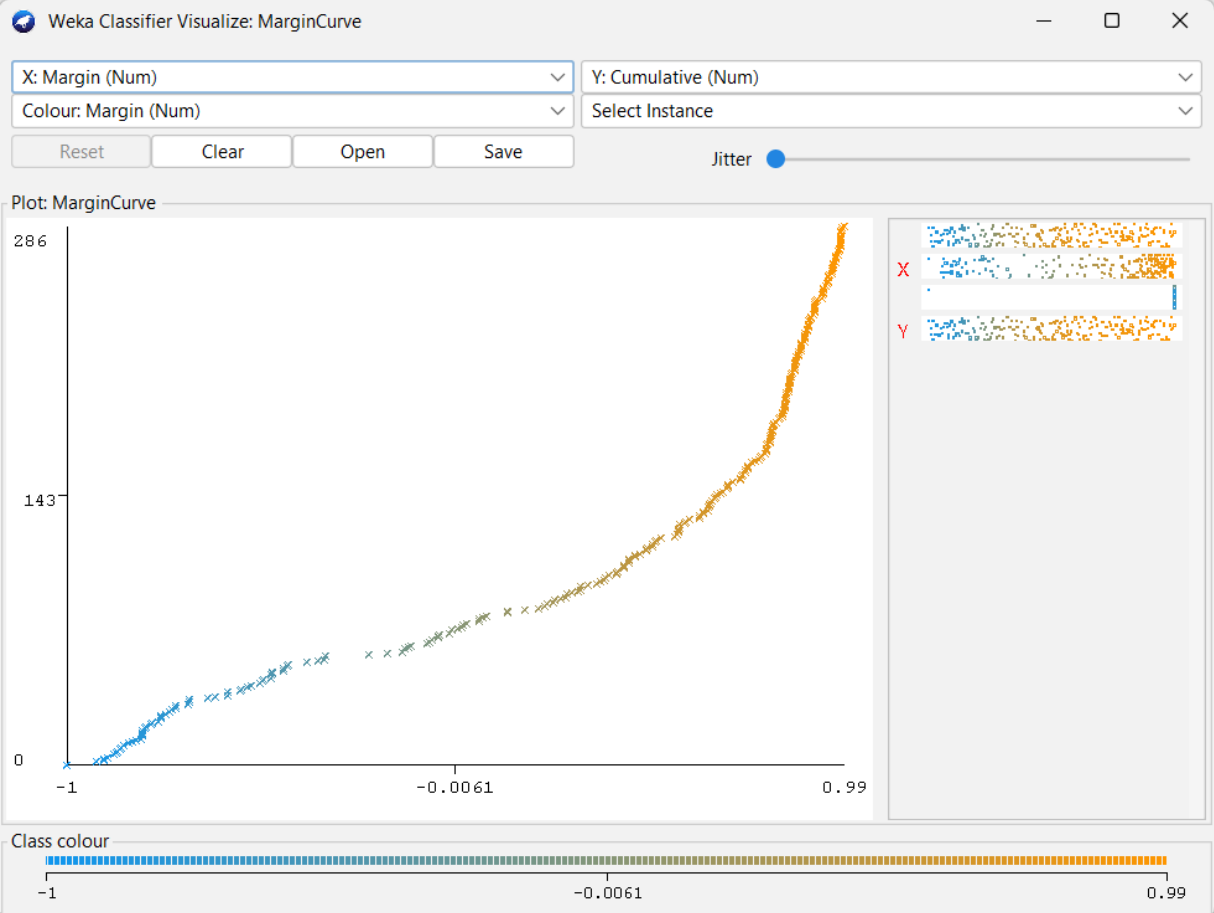
Time taken to test model on training data: 0 seconds

=== Summary ===
Correctly Classified Instances      215      75.1748 %
Incorrectly Classified Instances    71      24.8252 %
Kappa statistic                    0.3693
Mean absolute error                 0.3012
Root mean squared error             0.4278
Relative absolute error             72.0082 %
Root relative squared error         93.6095 %
Total Number of Instances          286

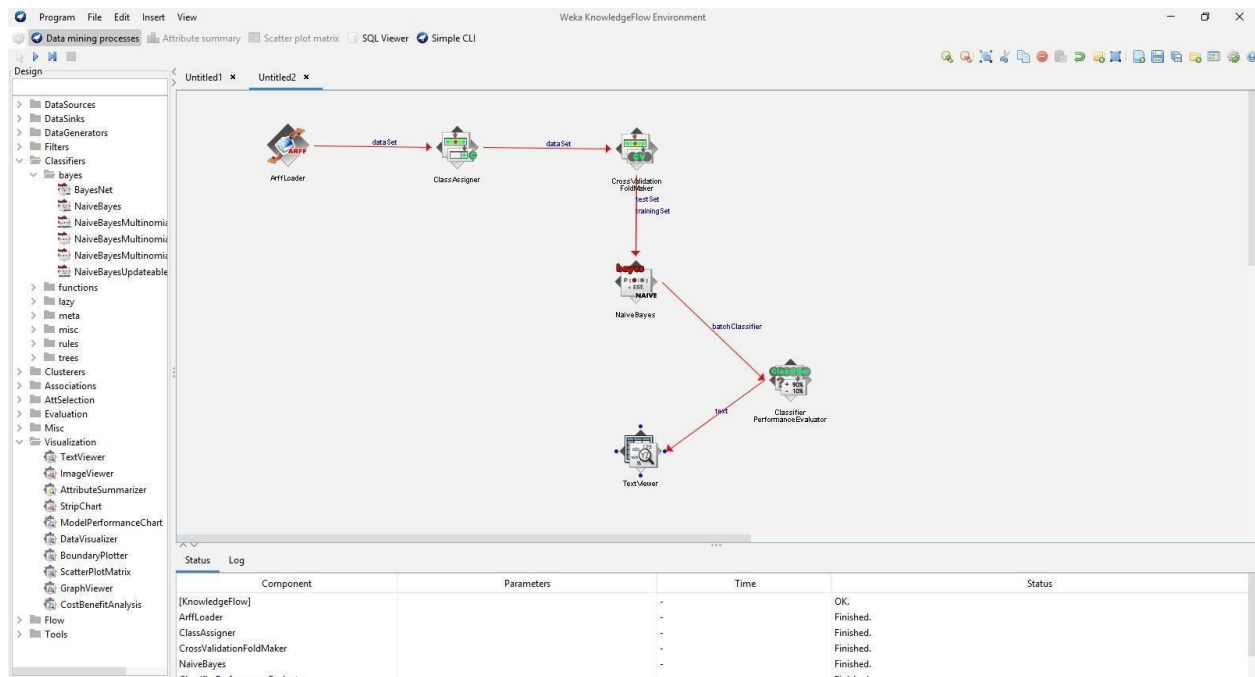
=== Detailed Accuracy By Class ===
              TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
              0.866    0.518    0.798      0.866    0.831      0.374    0.760    0.879    no-recurrence-events
              0.482    0.134    0.603      0.482    0.536      0.374    0.760    0.610    recurrence-events
Weighted Avg.  0.752    0.404    0.740      0.752    0.743      0.374    0.760    0.799

=== Confusion Matrix ===
  a  b  <-- classified as
174 27 | a = no-recurrence-events
 44 41 | b = recurrence-events
```

Status: OK x0



Knowledge Base:



Text Viewer

Result list

14:27:55.528 - NaiveBayes

Text

=== Evaluation result ===

Scheme: NaiveBayes

Relation: breast-cancer

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	205	71.6783 %
Incorrectly Classified Instances	81	28.3217 %
Kappa statistic	0.2857	
Mean absolute error	0.3272	
Root mean squared error	0.4534	
Relative absolute error	78.2086 %	
Root relative squared error	96.1872 %	
Total Number of Instances	286	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.836	0.545	0.778	0.636	0.696	0.288	0.701	0.637	no-recurrence-events
	0.435	0.164	0.529	0.435	0.477	0.288	0.701	0.514	recurrence-events
Weighted Avg.	0.717	0.446	0.704	0.717	0.708	0.288	0.701	0.741	

=== Confusion Matrix ===

	a	b	Classified as
168 33	a	a = no-recurrence-events	
48 37	b	b = recurrence-events	

Close Settings Clear results

i) Weka Clustering using K-means: Explorer:

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer: Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

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):

- 14:59:28 - SimpleKMeans

Status: OK

Cluster output:

```
=== Run information ===
#Scheme:      weka.clusterers.SimpleKMeans -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10
Relation:      breast-cancer
Instances:      286
Attributes:     10
age
menopause
tumor-size
inv-nodes
node-caps
deg-malign
breast
breast-quad
irradiat
Class
Test mode:      evaluate on training data

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

KMeans
=====

Number of iterations: 3
Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:

Cluster#
```

Weka Explorer

Preprocess Classify **Cluster** Associate Select attributes Visualize

Clusterer: Choose **SimpleKMeans** -init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -N 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

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):

- 14:59:28 - SimpleKMeans

Status: OK

Cluster output:

```
Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events
Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:

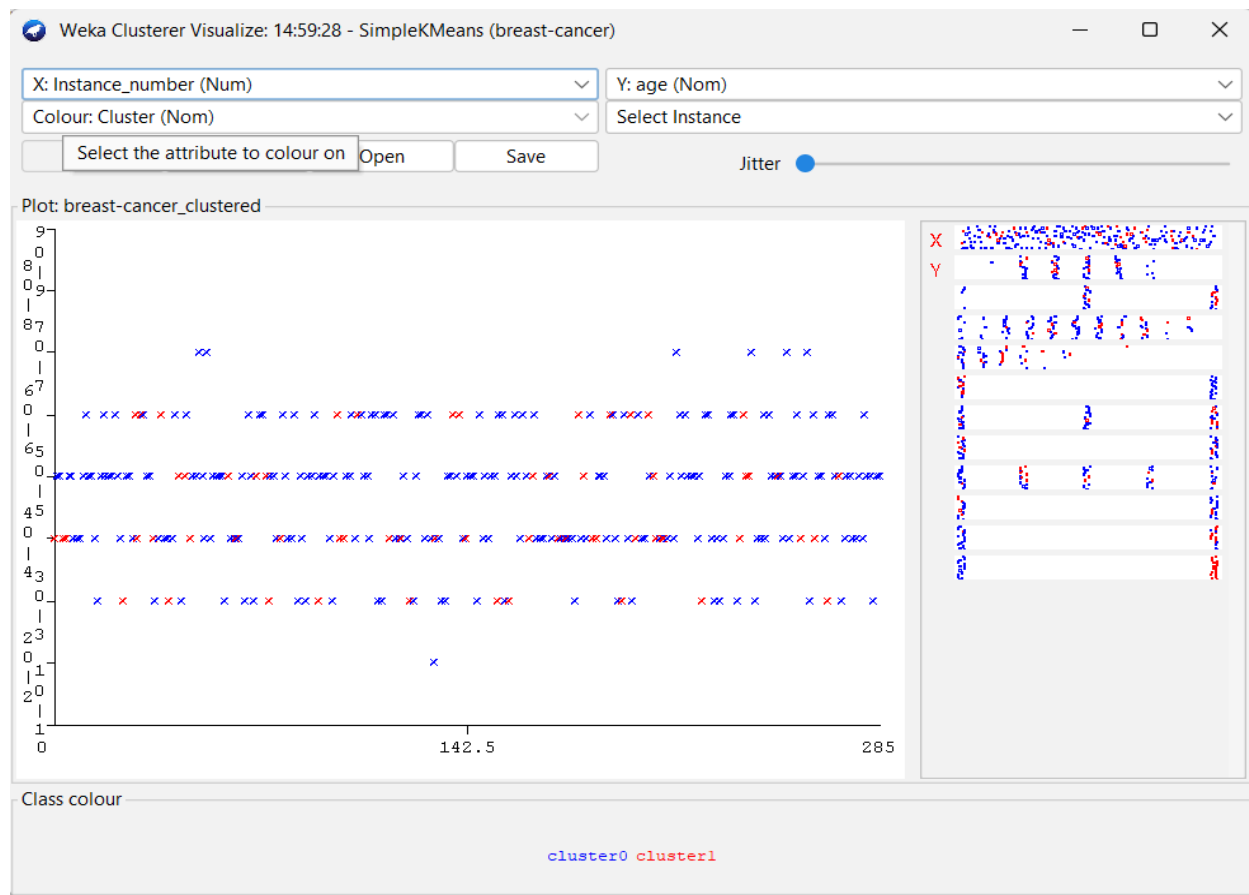
Attribute          Full Data          Cluster#          0          1
(286.0)            (225.0)            (61.0)
=====
age                50-59              50-59              40-49
menopause          premeno            premeno            premeno
tumor-size         30-34              25-25              30-34
inv-nodes          0-2                0-2                0-2
node-caps          no                 no                 yes
deg-malign         2                  2                  3
breast             left               left               left
breast-quad        left_low            left_low            left_low
irradiat           no                 no                 no
Class              no-recurrence-events no-recurrence-events recurrence-events

Time taken to build model (full training data) : 0.01 seconds

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

Clustered Instances

0      225 ( 79%)
1       61 ( 21%)
```



Knowledge base:



Test Viewer

Result list

14:38:49.744 - SimpleMeans

Test

=== Evaluation result for training instances ===

Scheme: SimpleMeans-init 0 -max-candidates 100 -periodic-pruning 10000 -min-density 2.0 -t1 -1.25 -t2 -1.0 -W 2 -A "weka.core.EuclideanDistance -R first-last" -I 500 -num-slots 1 -S 10

Relation: breast-cancer

kMeans

=====

Number of iterations: 3

Within cluster sum of squared errors: 1177.0

Initial starting points (random):

Cluster 0: 50-59,premeno,10-14,0-2,no,2,right,left_up,no,no-recurrence-events

Cluster 1: 40-49,premeno,15-19,0-2,yes,3,right,left_up,no,recurrence-events

Missing values globally replaced with mean/mode

Final cluster centroids:

Attribute	Full Data	Cluster#	
		0	1
	(286.0)	(225.0)	(61.0)

=====

age	50-59	50-59	40-49
menopause	premeno	premeno	premeno
tumor-size	30-34	25-29	30-34
inv-nodes	0-2	0-2	0-2
node-caps	no	no	yes
deg-malg	2	2	3
breast	left	left	left
breast-quad	left_low	left_low	left_low
irradiat	no	no	no
Class	no-recurrence-events	no-recurrence-events	recurrence-events

Clustered Instances

0	225 (79%)
1	61 (21%)

Close

Settings

Clear results

ii) Weka Classification using Decision tree: Explorer:

Weka Explorer
Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **DecisionStump**

Test options
☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Class
Start Stop

Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
15:02:07 - trees.DecisionStump

Classifier output

```
=== Run information ===  
  
Scheme:      weka.classifiers.trees.DecisionStump  
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  
  
=== Classifier model (full training set) ===  
  
Decision Stump  
  
Classifications  
  
deg-malig = 3 : recurrence-events  
deg-malig != 3 : no-recurrence-events  
deg-malig is missing : no-recurrence-events  
  
Class distributions  
  
deg-malig = 3  
no-recurrence-events      recurrence-events  
0.47058823529411764      0.5294117647058824  
deg-malig != 3  
no-recurrence-events      recurrence-events  
0.80095950248756219      0.1990497512437812  
deg-malig is missing  
no-recurrence-events      recurrence-events  
0.7027972027972028      0.2972027972027972
```

Status: OK Log x 0

Weka Explorer
Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **DecisionStump**

Test options
☒ Use training set
☐ Supplied test set Set...
☐ Cross-validation Folds 10
☐ Percentage split % 66
More options...

(Nom) Class
Start Stop

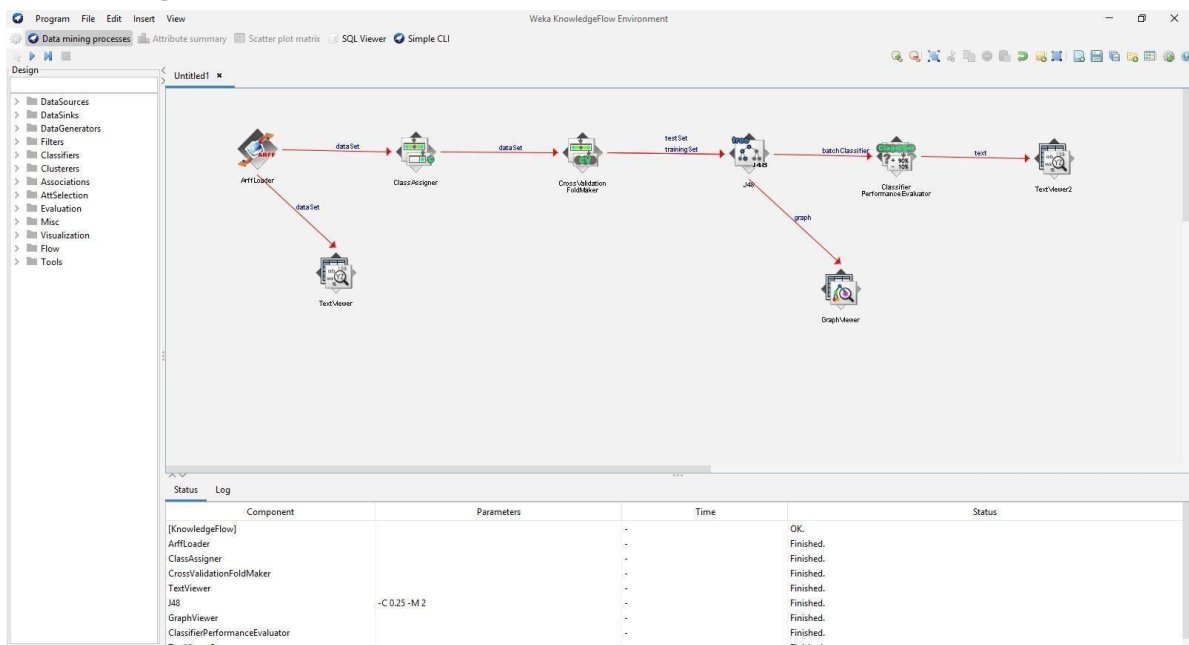
Result list (right-click for options)
14:49:16 - trees.J48
14:52:57 - bayes.NaiveBayes
15:02:07 - trees.DecisionStump

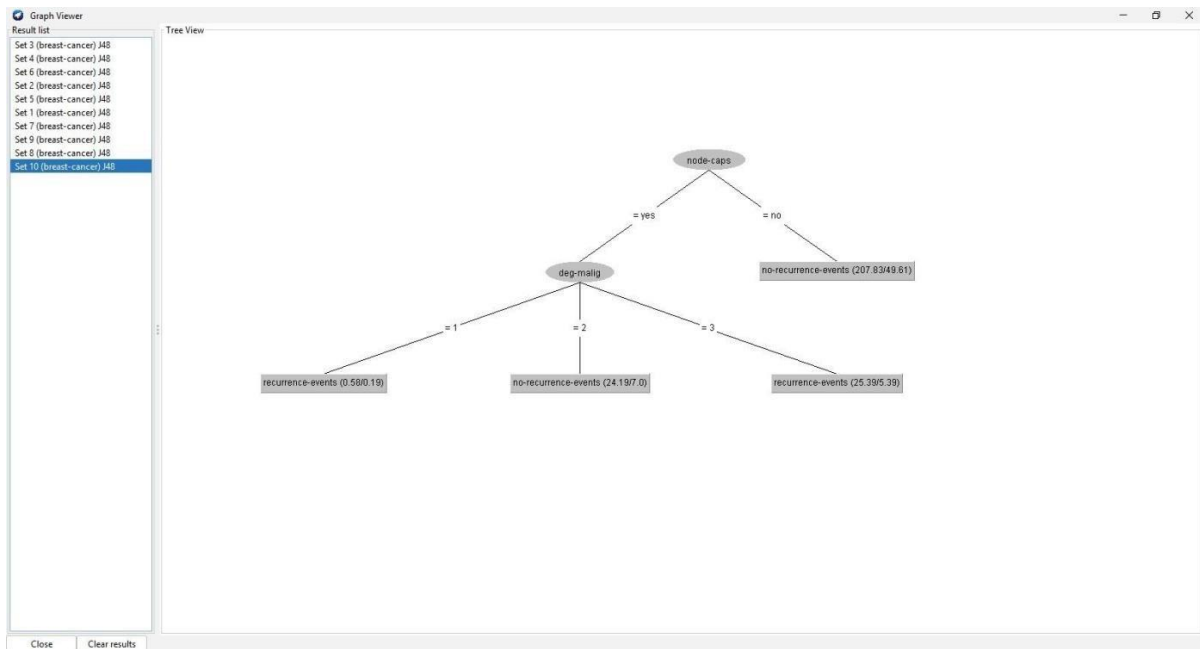
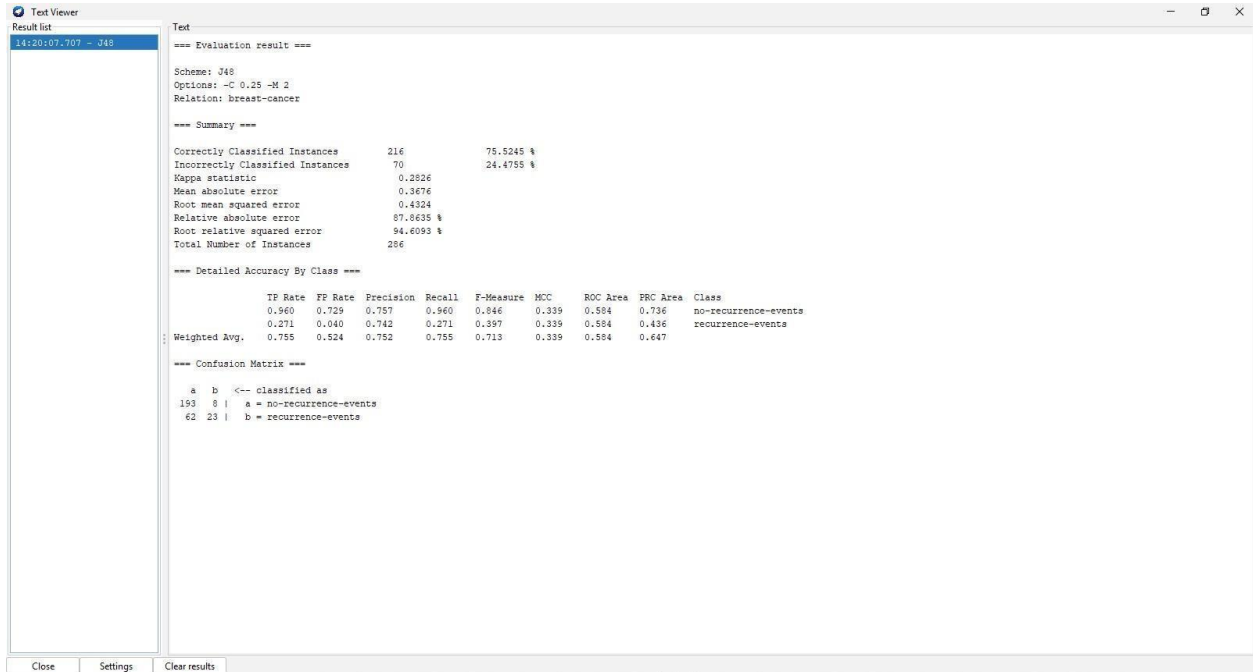
Classifier output

```
no-recurrence-events      recurrence-events  
0.80095950248756219      0.1990497512437812  
deg-malig is missing  
no-recurrence-events      recurrence-events  
0.7027972027972028      0.2972027972027972  
  
Time taken to build model: 0 seconds  
  
=== Evaluation on training set ===  
  
Time taken to test model on training data: 0 seconds  
  
=== Summary ===  
  
Correctly Classified Instances      206      72.028 %  
Incorrectly Classified Instances    80      27.972 %  
Kappa statistic                    0.3304  
Mean absolute error                 0.3721  
Root mean squared error             0.4314  
Relative absolute error             88.9615 %  
Root relative squared error         94.3834 %  
Total Number of Instances          286  
  
=== Detailed Accuracy By Class ===  
  
      TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class  
0.801  0.471  0.801    0.801  0.801    0.330  0.665  0.781  no-recurrence-events  
0.529  0.199  0.529    0.529  0.529    0.330  0.665  0.420  recurrence-events  
Weighted Avg.  0.720  0.390  0.720    0.720  0.720    0.330  0.665  0.674  
  
=== Confusion Matrix ===  
  
  a  b  <-- classified as  
161 40 | a = no-recurrence-events  
40 45 | b = recurrence-events
```

Status: OK Log x 0

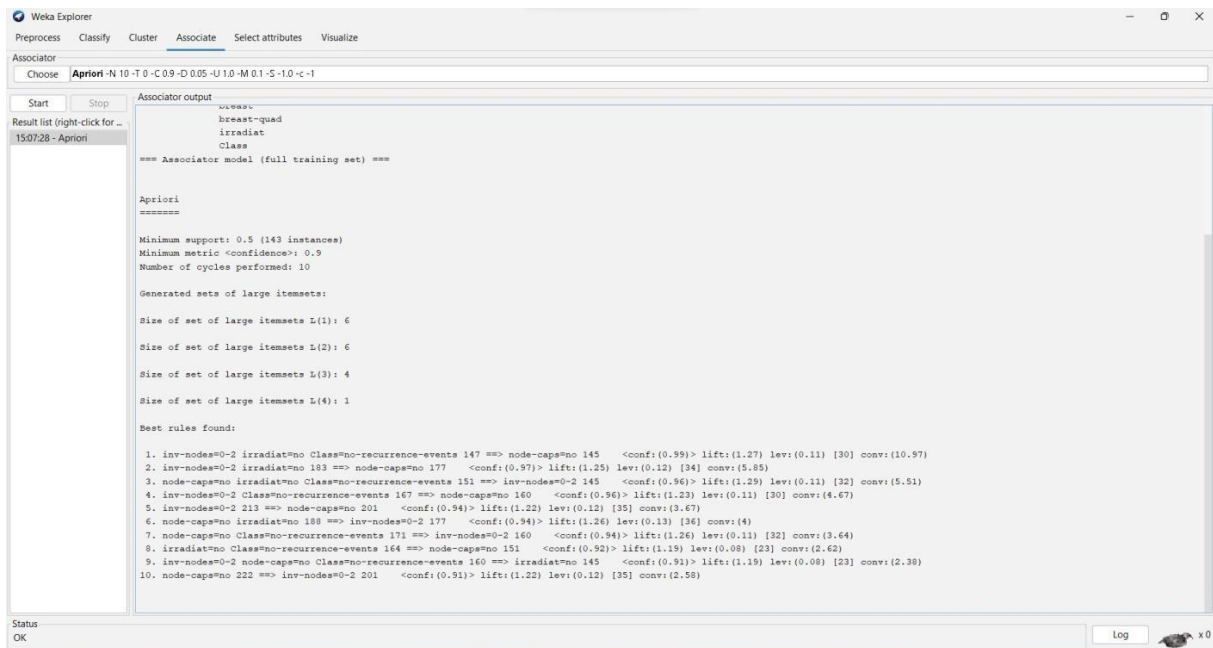
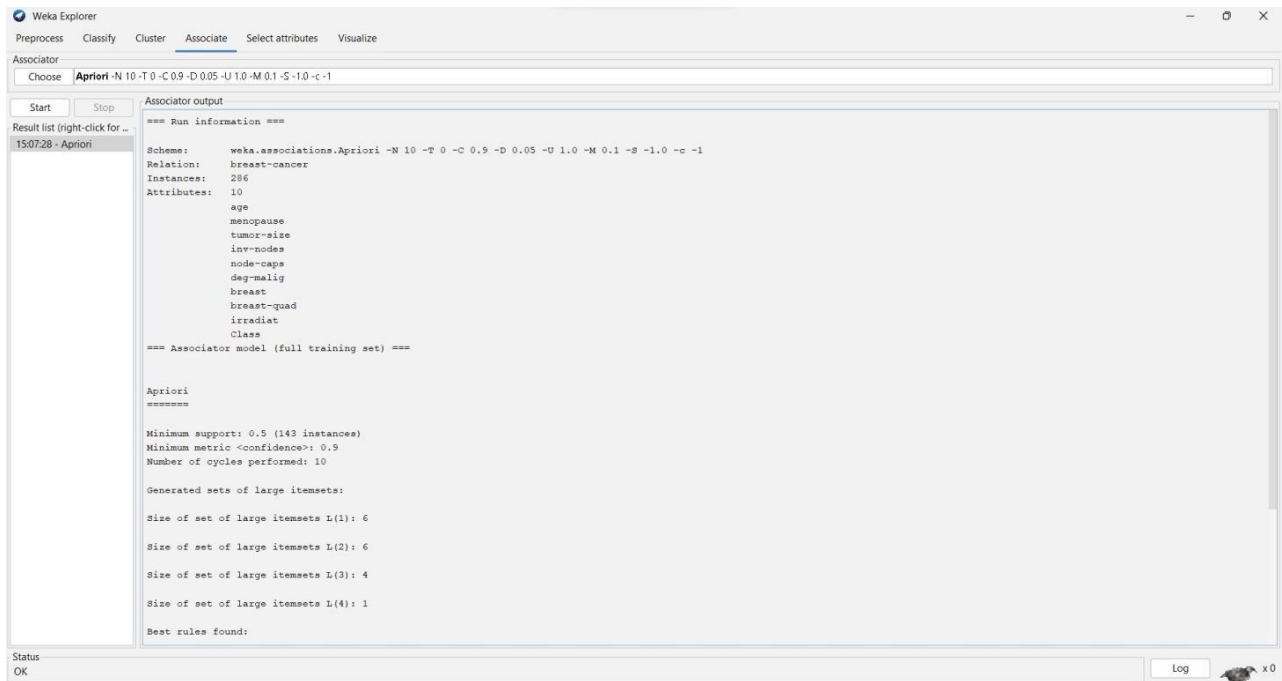
Knowledge base:





iii) Weka Association using Apriori :

Explorer:



Knowledge Base:

Weka KnowledgeFlow Environment

Program File Edit Insert View

Data mining processes Attribute summary Scatter plot matrix SQL Viewer Simple CLI

Design

Untitled1

Workflow diagram showing the process flow: ArffLoader -> TrainingSetMaker -> Apriori -> TextViewer.

Log

Component	Parameters	Time	Status
[KnowledgeFlow]		-	OK.
ArffLoader		-	Finished.
TrainingSetMaker		-	Finished.
Apriori	-N 10 -T 0 -C 0.8 -D 0.05 -U 1.0 -M 0.2 -S -1.0 -c -1	-	Finished.
TextViewer		-	Finished.

Text Viewer

Result list

15:05:00.082 - Model: Apri...
15:06:46.715 - Model: Apri...

Text

=== Associator model ===

Scheme: Apriori
Relation: supermarket

Apriori

Minimum support: 0.3 (1386 instances)
Minimum metric <confidence>: 0.8
Number of cycles performed: 14

Generated sets of large itemsets:

Size of set of large itemsets L(1): 25
Size of set of large itemsets L(2): 69
Size of set of large itemsets L(3): 20

Best rules found:

1. biscuits=t vegetables=t 1764 ==> bread and cake=t 1487 <conf:(0.84)> lift:(1.17) lev:(0.05) [217] conv:(1.78)
2. total=high 1679 ==> bread and cake=t 1413 <conf:(0.84)> lift:(1.17) lev:(0.04) [204] conv:(1.76)
3. biscuits=t milk-cream=t 1767 ==> bread and cake=t 1485 <conf:(0.84)> lift:(1.17) lev:(0.05) [213] conv:(1.75)
4. biscuits=t fruit=t 1837 ==> bread and cake=t 1541 <conf:(0.84)> lift:(1.17) lev:(0.05) [218] conv:(1.73)
5. biscuits=t frozen food=t 1810 ==> bread and cake=t 1510 <conf:(0.83)> lift:(1.16) lev:(0.04) [207] conv:(1.69)
6. frozen food=t fruit=t 1561 ==> bread and cake=t 1548 <conf:(0.83)> lift:(1.16) lev:(0.05) [208] conv:(1.66)
7. frozen food=t milk-cream=t 1826 ==> bread and cake=t 1516 <conf:(0.83)> lift:(1.15) lev:(0.04) [201] conv:(1.65)
8. baking needs=t milk-cream=t 1907 ==> bread and cake=t 1580 <conf:(0.83)> lift:(1.15) lev:(0.04) [207] conv:(1.63)
9. milk-cream=t fruit=t 2038 ==> bread and cake=t 1684 <conf:(0.83)> lift:(1.15) lev:(0.05) [217] conv:(1.61)
10. baking needs=t biscuits=t 1764 ==> bread and cake=t 1456 <conf:(0.83)> lift:(1.15) lev:(0.04) [186] conv:(1.6)

Close Settings Clear results