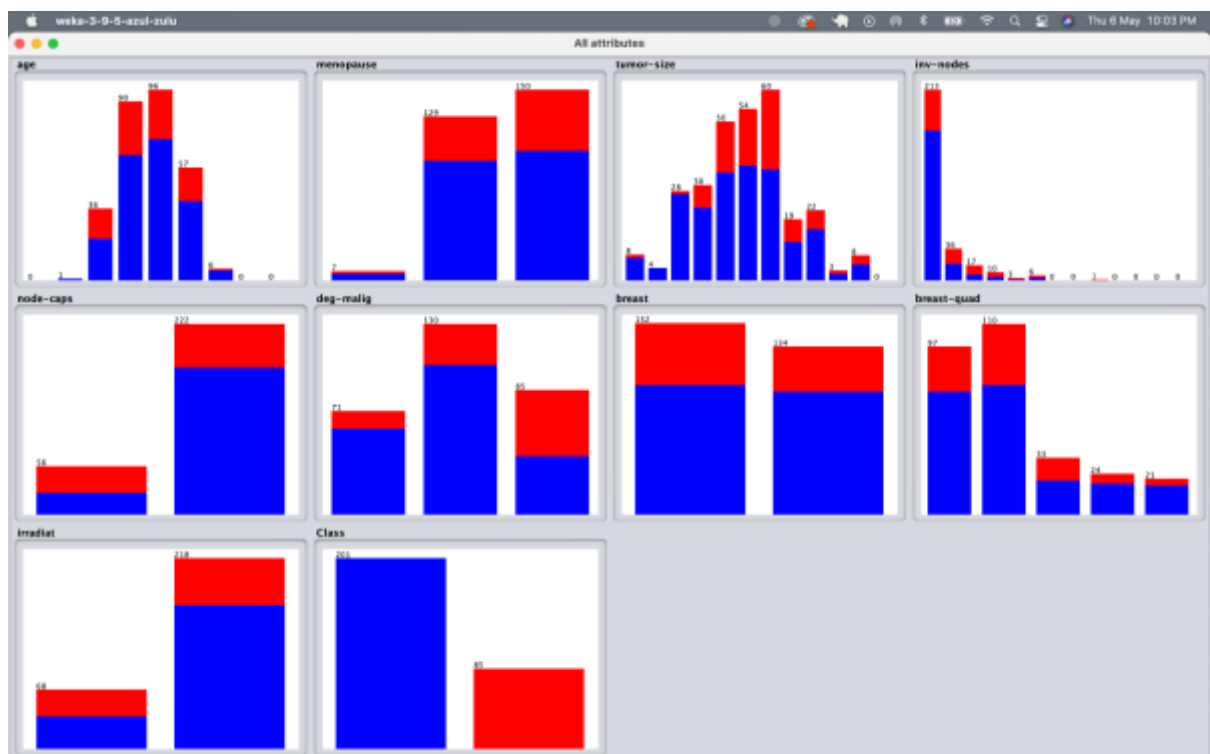
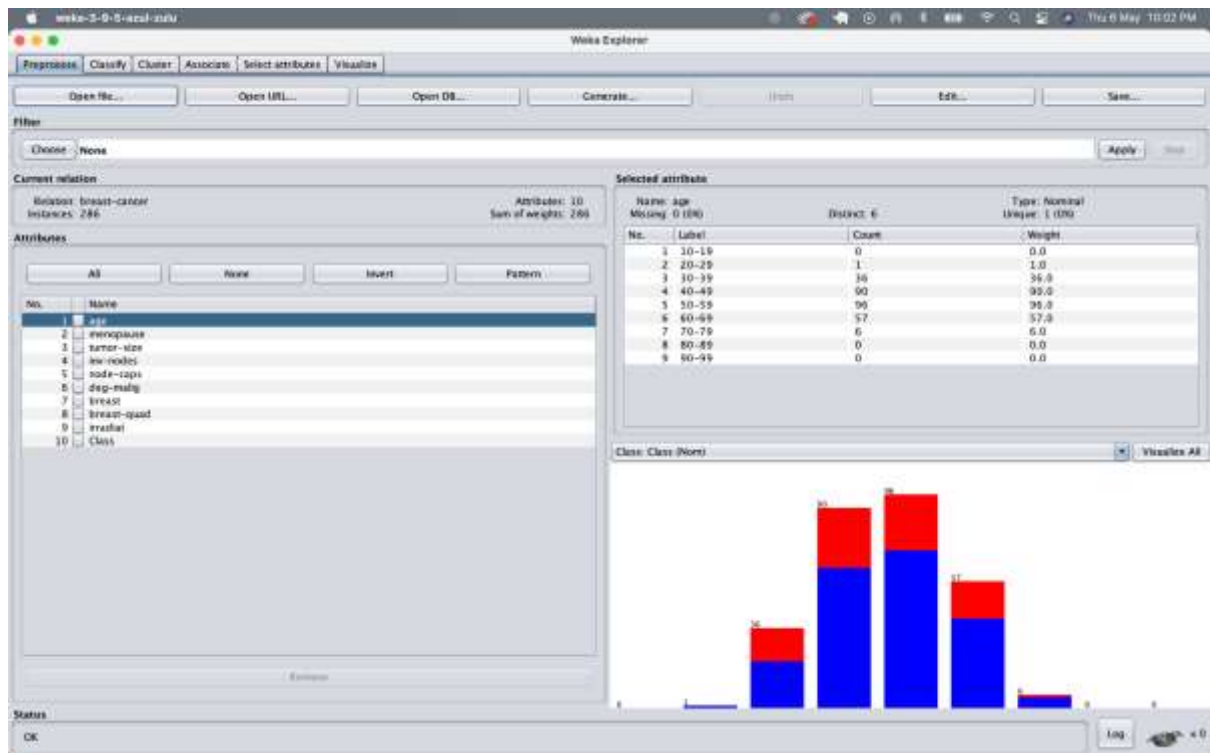


Name: Mareena Fernandes

TE IT	Roll number : 8669
Expt. number : 4	Date of implementation: 10/05/2021
Aim : To analyze and evaluate the performance of different classification algorithms using WEKA (data mining tool)	
Related Course outcome : CO3 Upon completion of this course students will be able to evaluate the performance of different data mining algorithms using latest tools	
<p>Theory : WEKA is a data mining system developed by the University of Waikato in New Zealand that implements data mining algorithms. It is a collection of machine learning algorithms for data mining tasks. The algorithms are applied directly to a dataset. WEKA implements algorithms for data preprocessing, classification, regression, clustering, association rules; it also includes a visualization tools.</p> <p>When 'WEKA GUI Chooser' window appears on the screen, we can select one of the four options. The options are 1. Simple CLI : provides a simple command line interface and allows direct execution of WEKA commands. 2. Explorer : is an environment for exploring data. 3. Experimenter : is an environment for performing experiments and conducting statistical tests between learning schemes. 4. Knowledge Flow : is a Java-Beans-based interface for setting up and running machine learning experiments.</p> <p>Classifiers in WEKA are the models for predicting nominal or numeric quantities. The learning schemes available in WEKA include decision trees, Bayes net, neural network, support vector machine and so on.</p> <p>In this experiment analysis and evaluation of three classification algorithm: Naive Bayesian algorithm, C4.5 algorithm, zero R is done. Before running the classification algorithm it is required to set test options. The selected test options were :</p> <ol style="list-style-type: none">1. Use training set : Evaluates the classifier on how well it predicts the class of the instances it was trained on.2. Cross-validation : Evaluates the classifier by cross-validation, using the number of folds (10)3. Percentage split : Evaluates the classifier on how well it predicts a certain percentage of the data, which is held out for testing. <p>In the classifier evaluation options following options are checked :</p> <ol style="list-style-type: none">1. Output model : The output is the classification model on the full training set.2. Output per-class stats : The precision/recall and true/false statistics for each class output3. Output confusion matrix : The confusion matrix of one classifier's prediction is included in the output <p>When training set is complete, the 'classifier' output area on the right panel of the classifier window is filled with text describing the results of training and testing.</p>	

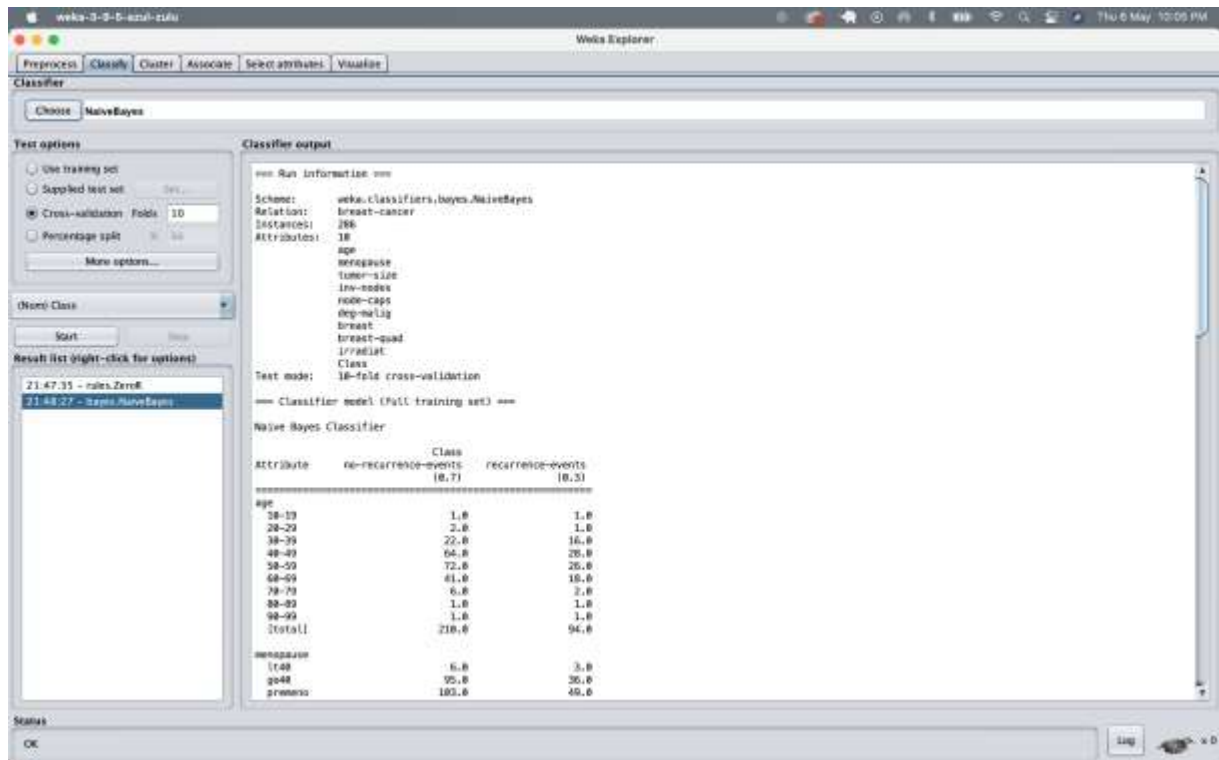
Dataset:



Classifier Output:

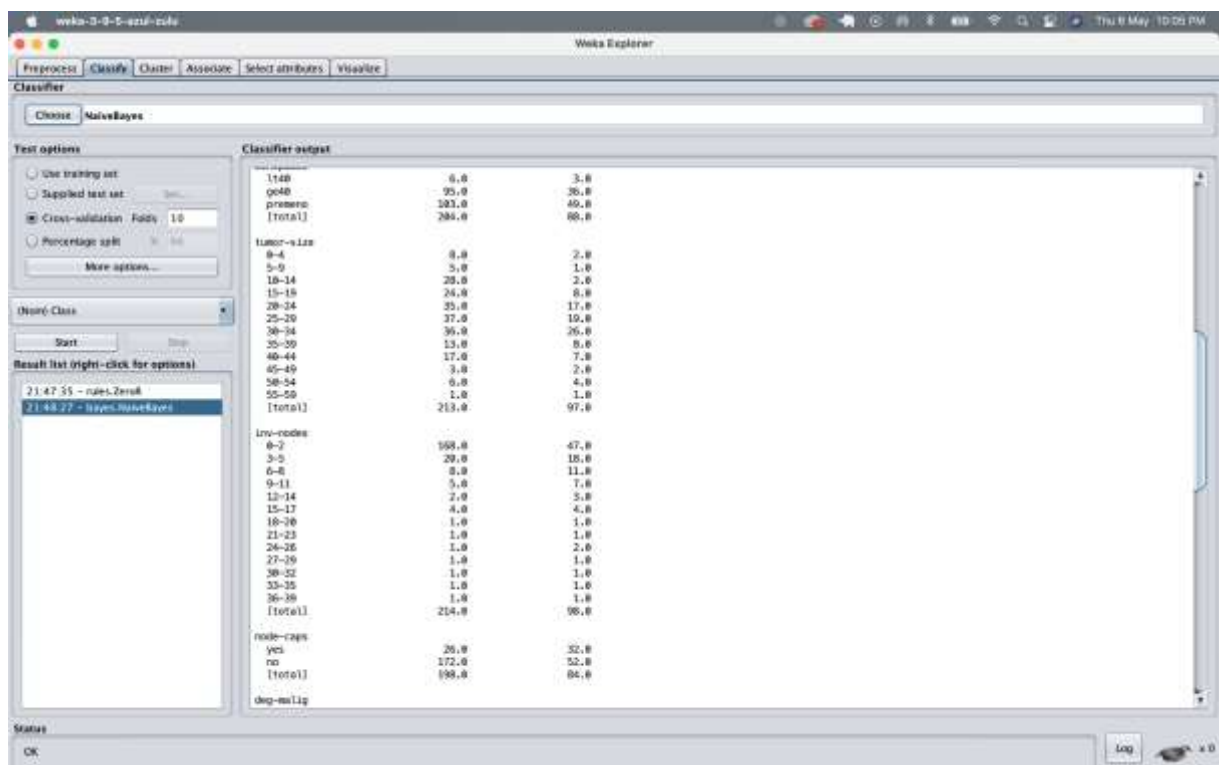
Naive Bayes

Cross Validation



The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section on the left indicates 'Cross-validation' is chosen with 10 folds. The 'Classifier output' pane on the right displays the 'Run information' and the 'Classifier model' for the 'breast-cancer' dataset. The model is a Naive Bayes Classifier. The 'Test mode' is '10-fold cross-validation'. The 'Classifier model (Full training set)' is shown as a table with three columns: 'Attribute', 'no-recurrence-events', and 'recurrence-events'. The 'no-recurrence-events' column has a value of 10.7, and the 'recurrence-events' column has a value of 0.3. The table lists the following attributes and their corresponding values:

Attribute	no-recurrence-events	recurrence-events
age		
18-19	1.0	1.0
20-29	2.0	1.0
30-39	22.0	16.0
40-49	64.0	26.0
50-59	72.0	26.0
60-69	61.0	18.0
70-79	6.0	2.0
80-89	1.0	1.0
90-99	1.0	1.0
[total]	218.0	94.0
menopause		
1-40	6.0	3.0
40-49	95.0	36.0
premeno	183.0	45.0



The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section on the left indicates 'Cross-validation' is chosen with 10 folds. The 'Classifier output' pane on the right displays the 'Run information' and the 'Classifier model' for the 'breast-cancer' dataset. The model is a Naive Bayes Classifier. The 'Test mode' is '10-fold cross-validation'. The 'Classifier model (Full training set)' is shown as a table with three columns: 'Attribute', 'no-recurrence-events', and 'recurrence-events'. The 'no-recurrence-events' column has a value of 10.7, and the 'recurrence-events' column has a value of 0.3. The table lists the following attributes and their corresponding values:

Attribute	no-recurrence-events	recurrence-events
1-40	6.0	3.0
40-49	95.0	36.0
premeno	183.0	45.0
[total]	284.0	84.0
tumor-size		
0-4	0.0	2.0
5-9	5.0	1.0
10-14	20.0	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	0.0	4.0
55-59	1.0	1.0
[total]	213.0	97.0
lin-nodes		
0-2	168.0	47.0
3-5	29.0	18.0
6-8	8.0	11.0
9-11	5.0	7.0
12-14	2.0	5.0
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
36-38	1.0	1.0
[total]	214.0	98.0
node-caps		
yes	26.0	32.0
no	172.0	52.0
[total]	198.0	84.0
deg-malign		

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 %: 65
 More options...

(Next Class)

Start

Result list (right-click for options)

21.47.55 - rules.Zeroth
 21.48.27 - Local.NaiveBayes

Classifier output

Test set	TP Rate	FP Rate	Precision	Recall	F-Measure	RCC	ROC Area	PRC Area	Class
1	68.0	13.0							
2	100.0	29.0							
3	41.0	46.0							
[total]	284.0	88.0							
breast									
left	184.0	58.0							
right	90.0	37.0							
[total]	280.0	87.0							
breast-quad									
left_low	72.0	27.0							
left_high	76.0	36.0							
right_low	21.0	14.0							
right_high	19.0	7.0							
central	18.0	5.0							
[total]	206.0	89.0							
irradiat									
yes	30.0	32.0							
no	105.0	55.0							
[total]	280.0	87.0							

Time taken to build model: 0 seconds

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	285	71.6703 %
Incorrectly Classified Instances	81	20.3217 %
Kappa statistic	0.2857	
Mean absolute error	0.3272	
Root mean squared error	0.4534	
Relative absolute error	78.2686 %	
Root relative squared error	99.1872 %	
Total Number of Instances	286	

== Detailed Accuracy By Class ==

TP Rate	FP Rate	Precision	Recall	F-Measure	RCC	ROC Area	PRC Area	Class
0.626	0.505	0.775	0.835	0.806	0.288	0.701	0.837	no-recurrence-events
0.435	0.564	0.529	0.435	0.477	0.268	0.701	0.514	recurrence-events
Weighted avg.	0.717	0.448	0.784	0.717	0.788	0.701	0.741	

Status

OK

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 %: 65
 More options...

(Next Class)

Start

Result list (right-click for options)

21.47.55 - rules.Zeroth
 21.48.27 - Local.NaiveBayes

Classifier output

Test set	TP Rate	FP Rate	Precision	Recall	F-Measure	RCC	ROC Area	PRC Area	Class
1	72.0	27.0							
2	70.0	36.0							
3	21.0	14.0							
right_low	19.0	7.0							
central	18.0	5.0							
[total]	206.0	89.0							
irradiat									
yes	30.0	32.0							
no	105.0	55.0							
[total]	280.0	87.0							

Time taken to build model: 0 seconds

== Stratified cross-validation ==

== Summary ==

Correctly Classified Instances	285	71.6703 %
Incorrectly Classified Instances	81	20.3217 %
Kappa statistic	0.2857	
Mean absolute error	0.3272	
Root mean squared error	0.4534	
Relative absolute error	78.2686 %	
Root relative squared error	99.1872 %	
Total Number of Instances	286	

== Detailed Accuracy By Class ==

TP Rate	FP Rate	Precision	Recall	F-Measure	RCC	ROC Area	PRC Area	Class
0.626	0.505	0.775	0.835	0.806	0.288	0.701	0.837	no-recurrence-events
0.435	0.564	0.529	0.435	0.477	0.268	0.701	0.514	recurrence-events
Weighted avg.	0.717	0.448	0.784	0.717	0.788	0.701	0.741	

== Confusion Matrix ==

	0	1	← classified as
168 33	168	33	0 = no-recurrence-events
48 37	48	37	1 = recurrence-events

Status

OK

Use Training Sets:

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section has 'Use training set' selected. The 'Classifier output' section displays the following information:

Run information

Scheme: weka.classifiers.bayes.NaiveBayes
Relation: breast-cancer
Instances: 286
Attributes: 10

Test mode: evaluate on training data

Classifier model (full training set)

Naive Bayes Classifier

Attribute	no-recurrence-events (0.7)	recurrence-events (0.3)
age		
39-49	1.0	1.0
50-59	2.0	1.0
60-69	22.0	36.0
70-79	64.0	28.0
80-89	72.0	26.0
90-99	41.0	38.0
100-109	6.0	2.0
110-119	1.0	1.0
120-129	1.0	1.0
[total]	218.0	94.0
tumor-size		
1-4	6.0	3.0
5-9	35.0	36.0
10-14	183.0	49.0
[total]	224.0	88.0
lin-nodes		
0-2	308.0	47.0
3-5	29.0	16.0
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	96.0
node-caps		
no	26.0	32.0
yes	172.0	52.0
[total]	198.0	84.0
deg-malign		

The screenshot shows the Weka Explorer interface with the Naive Bayes classifier selected. The 'Test options' section has 'Use training set' selected. The 'Classifier output' section displays the following information:

Run information

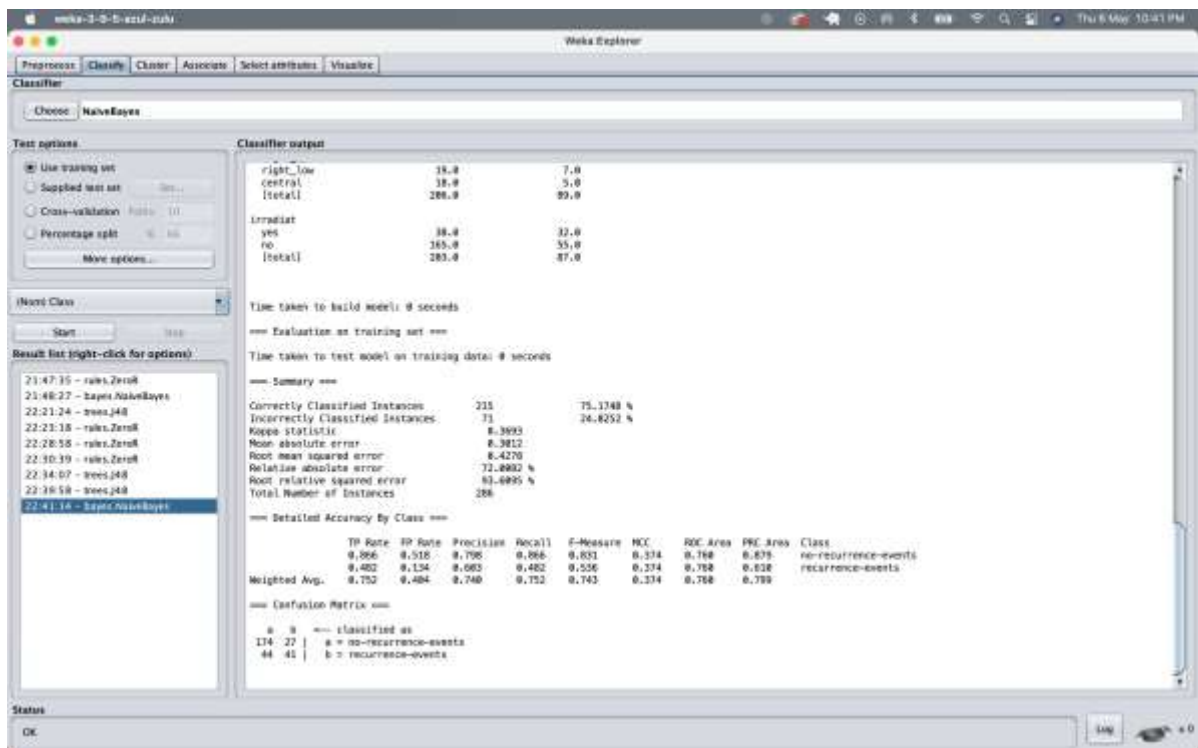
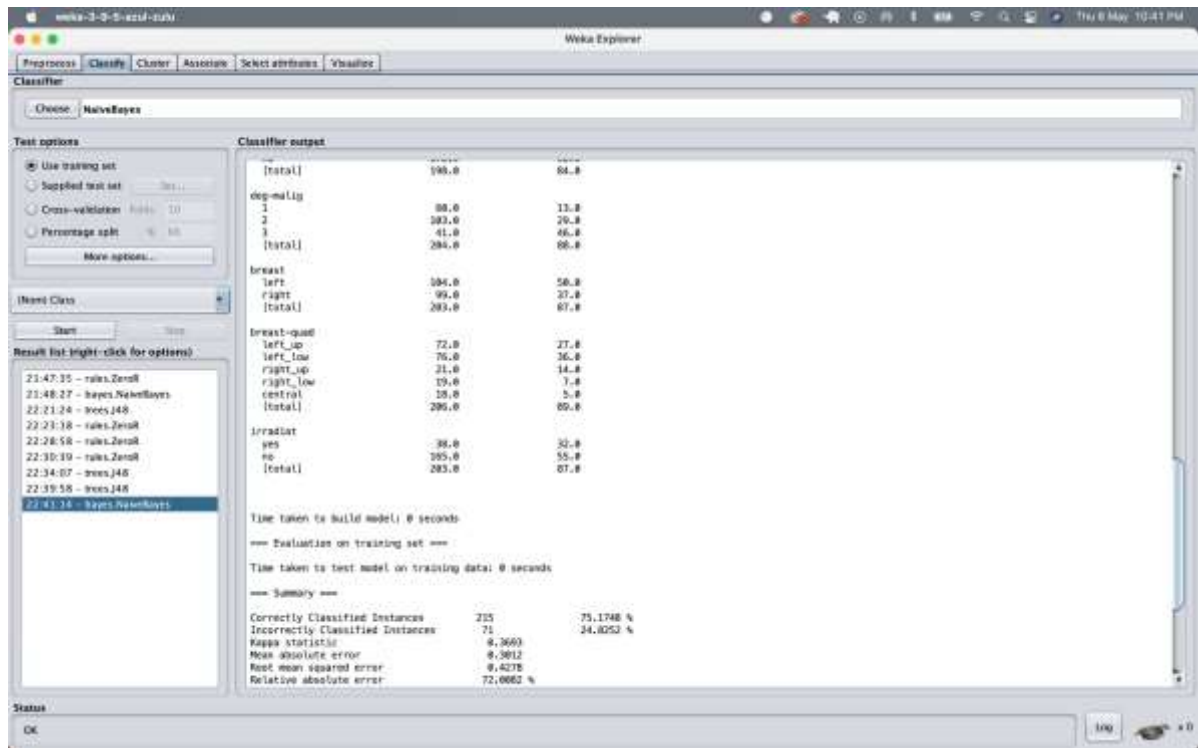
Scheme: weka.classifiers.bayes.NaiveBayes
Relation: breast-cancer
Instances: 286
Attributes: 10

Test mode: evaluate on training data

Classifier model (full training set)

Naive Bayes Classifier

Attribute	no-recurrence-events (0.7)	recurrence-events (0.3)
age		
39-49	1.0	1.0
50-59	2.0	1.0
60-69	22.0	36.0
70-79	64.0	28.0
80-89	72.0	26.0
90-99	41.0	38.0
100-109	6.0	2.0
110-119	1.0	1.0
120-129	1.0	1.0
[total]	218.0	94.0
tumor-size		
1-4	6.0	3.0
5-9	35.0	36.0
10-14	183.0	49.0
[total]	224.0	88.0
lin-nodes		
0-2	308.0	47.0
3-5	29.0	16.0
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	96.0
node-caps		
no	26.0	32.0
yes	172.0	52.0
[total]	198.0	84.0
deg-malign		



Percentage Split:

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section on the left indicates a 'Percentage split' of 66%. The 'Classifier output' section on the right displays the following information:

Run information

- Scheme: weka.classifiers.bayes.NaiveBayes
- Relation: breast-cancer
- Instances: 286
- Attributes: 14

Test model: split 66.0% train, remainder test

Classifier model (full training set) is:

Naive Bayes Classifier

Attribute	no-recurrence-events (0, 7)	recurrence-events (8, 3)
age		
18-29	1.0	1.0
30-39	2.0	1.0
40-49	22.0	16.0
50-59	64.0	26.0
60-69	72.0	18.0
70-79	6.0	2.0
80-89	1.0	1.0
90-99	1.0	1.0
[total]	218.0	94.0
menopause		
140	6.0	3.0
ge40	95.0	36.0
[total]	101.0	49.0

The screenshot shows the Weka Explorer interface with the NaiveBayes classifier selected. The 'Test options' section on the left indicates a 'Percentage split' of 60%. The 'Classifier output' section on the right displays the following information:

Run information

- Scheme: weka.classifiers.bayes.NaiveBayes
- Relation: breast-cancer
- Instances: 286
- Attributes: 14

Test model: split 60.0% train, remainder test

Classifier model (full training set) is:

Naive Bayes Classifier

Attribute	no-recurrence-events (0, 7)	recurrence-events (8, 3)
age		
18-29	1.0	1.0
30-39	2.0	1.0
40-49	22.0	16.0
50-59	64.0	26.0
60-69	72.0	18.0
70-79	6.0	2.0
80-89	1.0	1.0
90-99	1.0	1.0
[total]	218.0	94.0
menopause		
140	6.0	3.0
ge40	95.0	36.0
[total]	101.0	49.0

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation

☒ Percentage split % 66

More options...

Default Class

Start

Result list (right-click for options)

21:47:35 - rules.ZeroR

21:48:27 - bayes.NaiveBayes

22:21:24 - trees.J48

22:23:18 - rules.ZeroR

22:28:58 - rules.ZeroR

22:30:39 - rules.ZeroR

22:34:07 - trees.J48

22:39:58 - trees.J48

22:41:14 - bayes.NaiveBayes

22:43:49 - bayes.NaiveBayes

Classifier output

	no-recurrence-events	recurrence-events
node-capac		
yes	26.0	32.0
no	172.0	52.0
[total]	198.0	84.0
deg-malign		
0	69.0	13.0
2	185.0	79.0
3	41.0	40.0
[total]	285.0	83.0
breast		
left	184.0	58.0
right	99.0	37.0
[total]	283.0	87.0
breast-quad		
left_low	72.0	27.0
left_low	76.0	36.0
right_low	21.0	14.0
right_low	19.0	7.0
central	18.0	5.0
[total]	205.0	89.0
irradiat		
yes	38.0	32.0
no	165.0	55.0
[total]	283.0	87.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	69	71.134 %
Incorrectly Classified Instances	28	28.866 %

Status

OK

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose NaiveBayes

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation

☒ Percentage split % 66

More options...

Default Class

Start

Result list (right-click for options)

21:47:35 - rules.ZeroR

21:48:27 - bayes.NaiveBayes

22:21:24 - trees.J48

22:23:18 - rules.ZeroR

22:28:58 - rules.ZeroR

22:30:39 - rules.ZeroR

22:34:07 - trees.J48

22:39:58 - trees.J48

22:41:14 - bayes.NaiveBayes

22:43:49 - bayes.NaiveBayes

Classifier output

	no-recurrence-events	recurrence-events
right_low	19.0	7.0
central	18.0	5.0
[total]	286.0	89.0
irradiat		
yes	38.0	32.0
no	165.0	55.0
[total]	283.0	87.0

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances	69	71.134 %
Incorrectly Classified Instances	28	28.866 %
Kappa statistic	0.3274	
Mean absolute error	0.2431	
Root mean squared error	0.4925	
Relative absolute error	79.9872 %	
Root relative squared error	100.9522 %	
Total Number of Instances	97	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	PRC Area	Class
Weighted Avg.	0.711	0.288	0.701	0.711	0.703	0.676	0.711	

=== Confusion Matrix ===

a b -- classified as

53 11 a = no-recurrence-events

17 16 b = recurrence-events

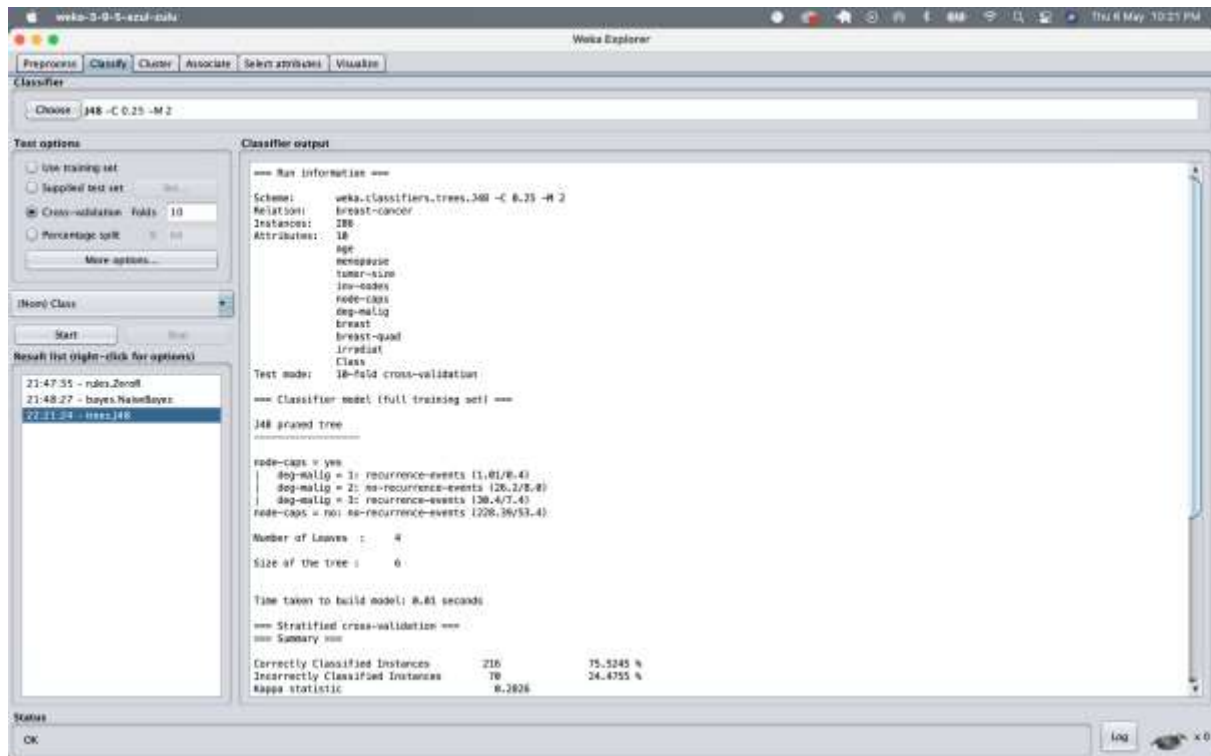
Status

OK

Classifier Output:

Decision Tree

Cross Validation



This screenshot shows the Weka Explorer interface with the Classifier tab selected. The classifier chosen is J48 -C 0.25 -M 2. The Test options are set to Cross-validation, Folds: 10. The Classifier output pane displays the following information:

```
=== Run information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:  breast-cancer
Instances: 286
Attributes:
  age
  menopause
  tumor-size
  inv-nodes
  node-caps
  deg-malign
  breast
  breast-quad
  irradiat
  Class

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===
J48 pruned tree

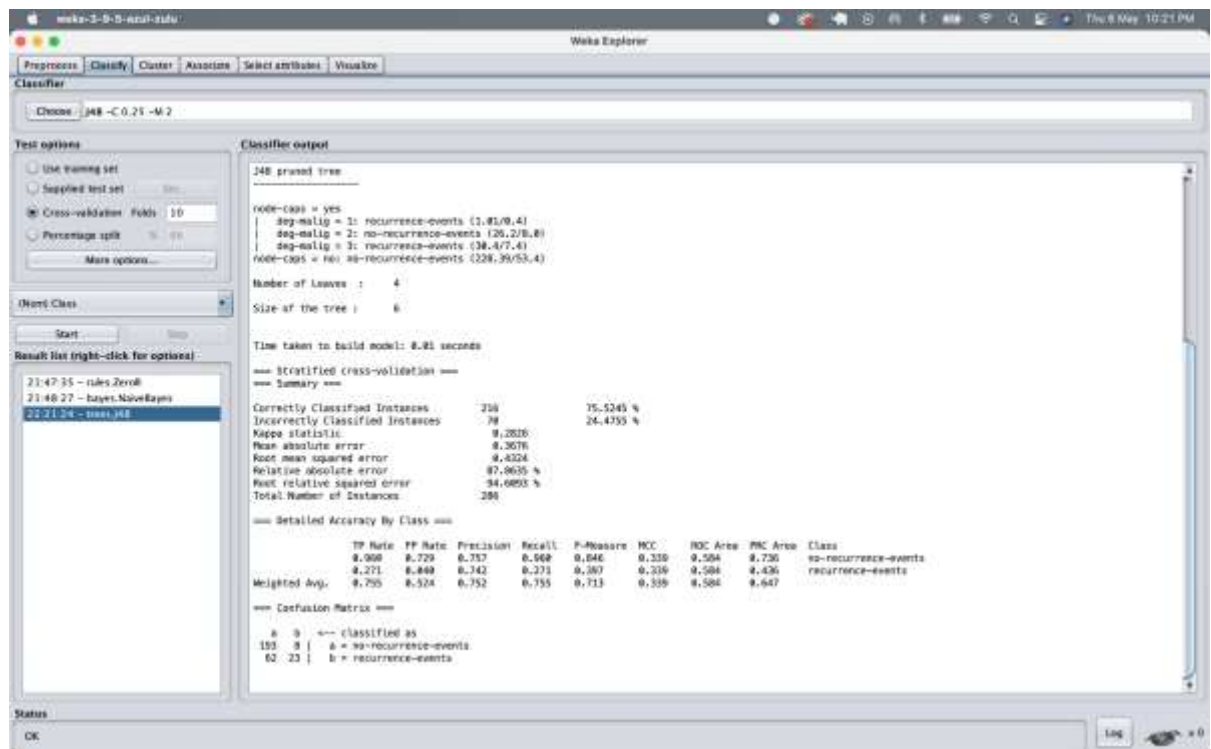
node-caps = yes
| deg-malign = 1: recurrence-events (1,01/0,4)
| deg-malign = 2: no-recurrence-events (26,2/8,0)
| deg-malign = 3: recurrence-events (38,4/7,4)
node-caps = no: no-recurrence-events (228,39/53,4)

Number of Leaves :    4
Size of the tree :    6

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances    216      75.5245 %
Incorrectly Classified Instances    70      24.4755 %
Kappa statistic      0.2826
```



This screenshot shows the same Weka Explorer interface, but with more detailed output displayed in the Classifier output pane:

```
J48 pruned tree

node-caps = yes
| deg-malign = 1: recurrence-events (1,01/0,4)
| deg-malign = 2: no-recurrence-events (26,2/8,0)
| deg-malign = 3: recurrence-events (38,4/7,4)
node-caps = no: no-recurrence-events (228,39/53,4)

Number of Leaves :    4
Size of the tree :    6

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances    216      75.5245 %
Incorrectly Classified Instances    70      24.4755 %
Kappa statistic      0.2826
Mean absolute error      0.3076
Root mean squared error      0.4324
Relative absolute error      87.0635 %
Relative squared error      94.0693 %
Total Number of Instances      286

=== Detailed Accuracy By Class ===


|               | TP Rate | FP Rate | Precision | Recall | F-Measure | MCC   | ROC Area | PRC Area | Class                |
|---------------|---------|---------|-----------|--------|-----------|-------|----------|----------|----------------------|
| Weighted Avg. | 0.908   | 0.729   | 0.757     | 0.669  | 0.646     | 0.339 | 0.504    | 0.730    | no-recurrence-events |
|               | 0.271   | 0.840   | 0.742     | 0.271  | 0.387     | 0.216 | 0.504    | 0.436    | recurrence-events    |



=== Confusion Matrix ===
      0   1  <-- classified as
153   0 | 0 = no-recurrence-events
 62  23 | 1 = recurrence-events
```

Use Training Set:

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' panel on the left has 'Use training set' checked. The 'Classifier output' panel on the right displays the following information:

```
=== Raw information ===
Scheme:   weka.classifiers.trees.J48 -C 0.25 -M 2
Relation:  breast-cancer
Instances: 286
Attributes:
  age
  menopause
  tumor-size
  int-nodes
  node-caps
  deg-malign
  breast
  breast-quad
  irradiat
  Class

Test mode:  evaluate on training data

=== Classifier model (full training set) ===
J48 pruned tree

node-caps = yes
| deg-malign = 1: recurrence-events (1,01/0,4)
| deg-malign = 2: no-recurrence-events (16,2/0,0)
| deg-malign = 3: recurrence-events (130,4/7,4)
node-caps = no: no-recurrence-events (128,36/53,4)

Number of leaves :    4
Size of the tree :    6

Time taken to build model: 0 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0 seconds

=== Summary ===
```

The 'Result list (right-click for options)' on the left shows a list of classifiers with their accuracy on the training set. The J48 classifier is highlighted with an accuracy of 22.34.02.

Classifier	Accuracy
21:47:35 - rules.Zeroth	21.47.35
21:48:27 - bayes.NaiveBayes	21.48.27
22:21:24 - trees.J48	22.21.24
22:23:18 - rules.Zeroth	22.23.18
22:28:58 - rules.Zeroth	22.28.58
22:30:39 - rules.Zeroth	22.30.39
22:34:02 - trees.J48	22.34.02

The screenshot shows the Weka Explorer interface with the J48 classifier selected. The 'Test options' panel on the left has 'Cross-validation' selected. The 'Classifier output' panel on the right displays the following information:

```
node-caps = yes
| deg-malign = 1: recurrence-events (1,01/0,4)
| deg-malign = 2: no-recurrence-events (16,2/0,0)
| deg-malign = 3: recurrence-events (130,4/7,4)
node-caps = no: no-recurrence-events (128,36/53,4)

Number of leaves :    4
Size of the tree :    6

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      237      75.8741 %
Incorrectly Classified Instances    89      24.1259 %
Kappa statistic                     0.2808
Mean absolute error                 0.3858
Root mean squared error             0.4269
Relative absolute error              87.4492 %
Root relative squared error         92.4617 %
Total number of instances           286

=== Detailed Accuracy By Class ===

      TP Rate  FP Rate  Precision  Recall  F-Measure  ROC  ROC Area  PRC Area  Class
0.271  0.835  0.767  0.271  0.486  0.252  0.639  0.461  no-recurrence-events
0.759  0.523  0.760  0.759  0.716  0.252  0.639  0.676  recurrence-events

Weighted Avg.  0.759  0.523  0.760  0.759  0.716  0.252  0.639  0.676

=== Confusion Matrix ===

  A B  == classified as
104 7 |  A = no-recurrence-events
 62 23 |  B = recurrence-events
```

The 'Result list (right-click for options)' on the left shows a list of classifiers with their accuracy on the test set. The J48 classifier is highlighted with an accuracy of 22.34.02.

Classifier	Accuracy
21:47:35 - rules.Zeroth	21.47.35
21:48:27 - bayes.NaiveBayes	21.48.27
22:21:24 - trees.J48	22.21.24
22:23:18 - rules.Zeroth	22.23.18
22:28:58 - rules.Zeroth	22.28.58
22:30:39 - rules.Zeroth	22.30.39
22:34:02 - trees.J48	22.34.02

Percentage Split:

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: J48 -C 0.25 -M 2

Test options:

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

More options...

(New) Class: +

Start Stop

Result list (right-click for options):

- 21:47:35 - rules.ZeroR
- 21:48:27 - bayes.NaiveBayes
- 22:21:24 - trees.J48
- 22:23:18 - rules.ZeroR
- 22:28:58 - rules.ZeroR
- 22:30:39 - rules.ZeroR
- 22:34:07 - trees.J48
- 22:35:18 - trees.J48

Classifier output

Run information ==

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2
Relation: breast-cancer
Instances: 288
Attributes: 18
age
menopause
tumor-size
lax-nodes
node-caps
deg-malign
breast
breast-quad
irradiat
Class
Test model: split 66.7% train, remainder test

Classifier model (full training set) ==

J48 pruned tree

```
node-caps = yes
| deg-malign = 1: recurrence-events (1.01/0.4)
| deg-malign = 2: no-recurrence-events (26.2/0.0)
| deg-malign = 3: recurrence-events (38.4/7.4)
node-caps = no: no-recurrence-events (228.39/53.4)
```

Number of Leaves : 4
Size of the tree : 6

Time taken to build model: 0 seconds

Evaluation on test split ==

Time taken to test model on test split: 0 seconds

Summary ==

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose: J48 -C 0.25 -M 2

Test options:

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

More options...

(New) Class: +

Start Stop

Result list (right-click for options):

- 21:47:35 - rules.ZeroR
- 21:48:27 - bayes.NaiveBayes
- 22:21:24 - trees.J48
- 22:23:18 - rules.ZeroR
- 22:28:58 - rules.ZeroR
- 22:30:39 - rules.ZeroR
- 22:34:07 - trees.J48
- 22:35:18 - trees.J48

Classifier output

```
node-caps = yes
| deg-malign = 1: recurrence-events (1.01/0.4)
| deg-malign = 2: no-recurrence-events (26.2/0.0)
| deg-malign = 3: recurrence-events (38.4/7.4)
node-caps = no: no-recurrence-events (228.39/53.4)
```

Number of Leaves : 4
Size of the tree : 6

Time taken to build model: 0 seconds

Evaluation on test split ==

Time taken to test model on test split: 0 seconds

Summary ==

Correctly Classified Instances	88	69.8412 %
Incorrectly Classified Instances	31	31.0588 %
Kappa statistic	0.7003	
Mean absolute error	0.3656	
Root mean squared error	0.6079	
Relative absolute error	92.4094 %	
Root relative squared error	182.0049 %	
Total Number of Instances	97	

Detailed Accuracy By Class ==

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	AUC	ROC Area	PRC Area	Class
Weighted Avg.	0.688	0.582	0.657	0.688	0.658	0.217	0.683	0.610		no-recurrence-events
										recurrence-events

Confusion Matrix ==

```
a b -- classified as
56 0 | a = no-recurrence-events
23 18 | b = recurrence-events
```

Classifier Output:

Zero R

Cross Validation

The screenshot shows the Weka Explorer interface with the ZeroR classifier selected. The 'Test options' panel on the left has 'Cross-validation' set to 'Folds: 10'. The 'Classifier output' panel on the right displays the following information:

Run information

- Schema: weka.classifiers.rules.ZeroR
- Relation: breast-cancer
- Instances: 286
- Attributes: 10 (age, menopause, tumor-size, lvm-nodes, node-caps, deg-malign, breast, breast-quad, irradiat, Class)

Test mode: 10-fold cross-validation

Classifier model (full training set): ZeroR predicts class value: no-recurrence-events

Time taken to build model: 0 seconds

Stratified cross-validation summary:

Metric	Value	Percentage
Correctly Classified Instances	261	78.2797 %
Incorrectly Classified Instances	85	29.7283 %
Kappa statistic	0	
Mean absolute error	0.4184	
Root mean squared error	0.4571	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	286	

Detailed Accuracy By Class:

TP Rate	PP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	1.000	0.703	1.000	0.825	?	0.483	0.685	no-recurrence-events
0.000	0.000	?	0.000	?	?	0.483	0.298	recurrence-events
Weighted Avg.	0.703	0.703	?	0.703	?	?	0.575	

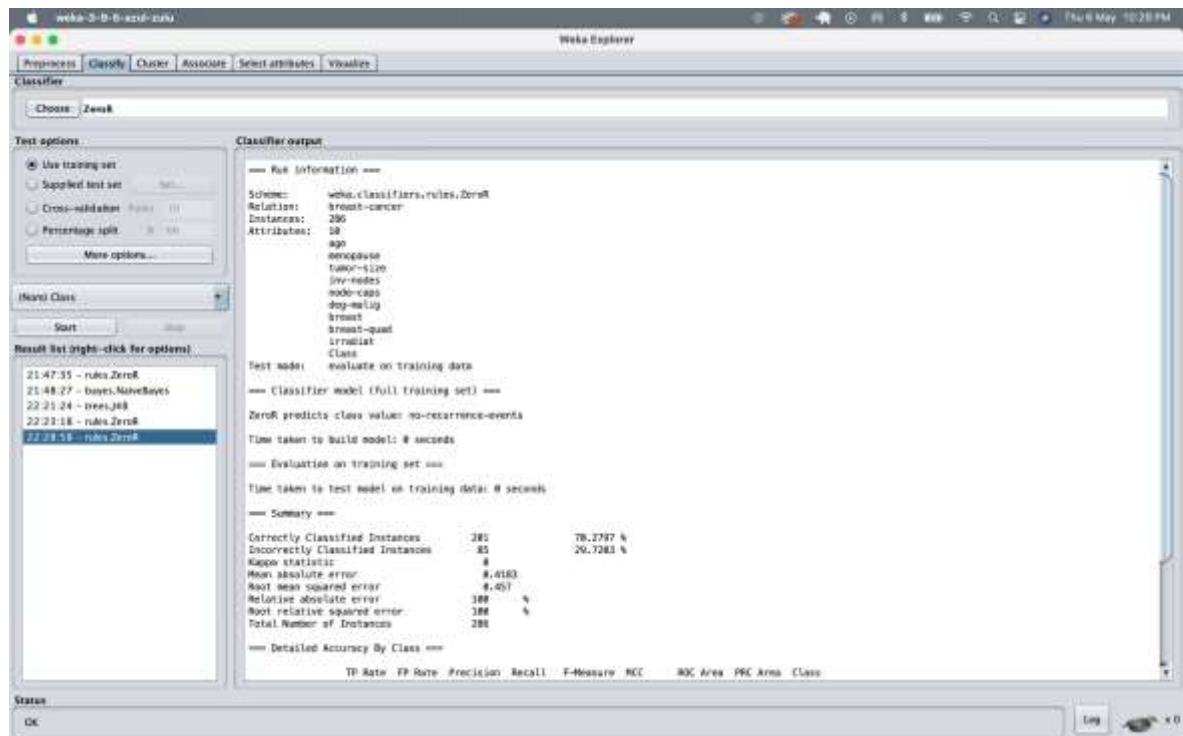
This screenshot is similar to the first one but includes a confusion matrix at the bottom of the 'Classifier output' panel.

Confusion Matrix:

```
# 0 <- classified as
# 1 <- no-recurrence-events
# 2 <- recurrence-events
# 3 <- recurrence-events
```

	0	1	2
0	261	0	0
1	0	85	0

Use training Set:



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

☒ Use training set
☐ Supplied test set
☐ Cross-validation Folds: 10
☐ Percentage split %: 65
More options...

(New) Class

Start Stop

Result list (right-click for options)

- 21:47:35 - rules.ZeroR
- 21:48:27 - bayes.NaiveBayes
- 22:21:24 - trees.J48
- 22:21:38 - rules.ZeroR
- 22:28:58 - rules.ZeroR

Classifier output

Rule information

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

Test model: evaluate on training data

Classifier model (full training set)

ZeroR predicts class value: no-recurrence-events

Time taken to build model: 0 seconds

Evaluation on training set

Time taken to test model on training data: 0 seconds

Summary

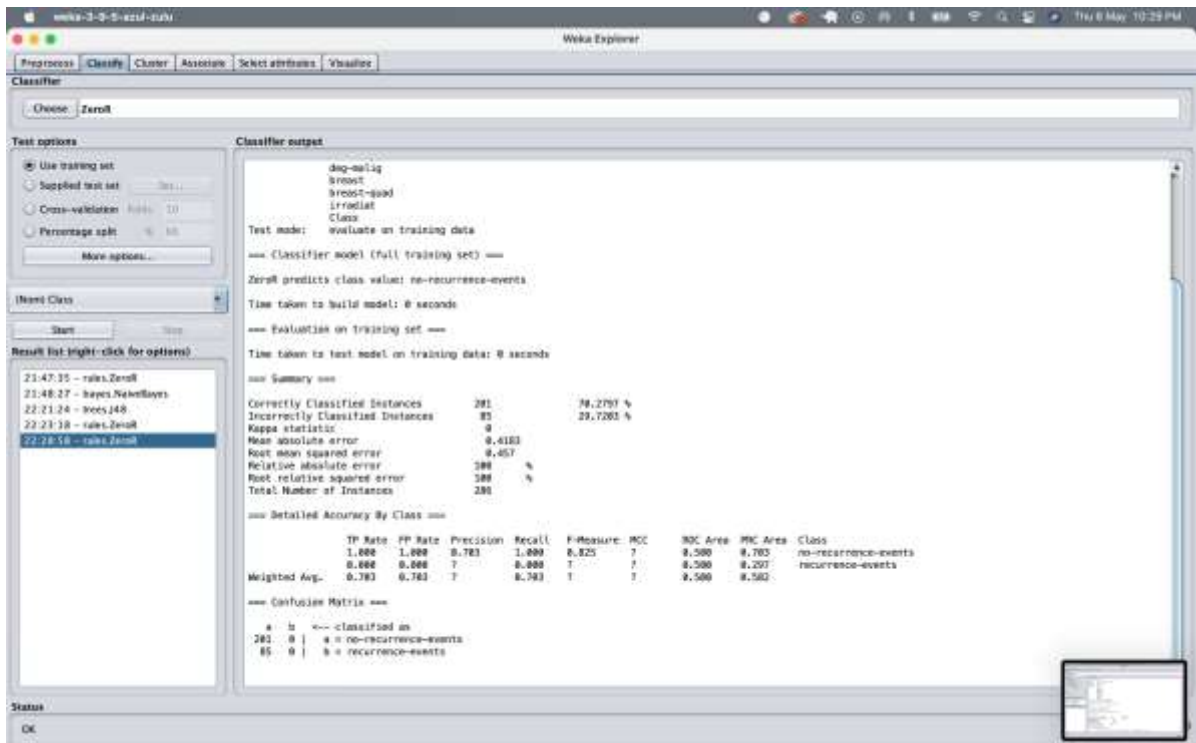
Metric	Value	Percentage
Correctly Classified Instances	281	78.2797 %
Incorrectly Classified Instances	5	20.7203 %
Kappa statistic	0	
Mean absolute error	0.0183	
Root mean squared error	0.427	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	286	

Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	0.925	1	1	0.500	0.703	no-recurrence-events
0.000	0.000	1	0.000	1	1	0.500	0.297	recurrence-events
Weighted Avg.	0.783	0.783	1	0.783	1	1	0.500	0.502

Status

OK



Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose ZeroR

Test options

☒ Use training set
☐ Supplied test set
☐ Cross-validation Folds: 10
☐ Percentage split %: 65
More options...

(New) Class

Start Stop

Result list (right-click for options)

- 21:47:35 - rules.ZeroR
- 21:48:27 - bayes.NaiveBayes
- 22:21:24 - trees.J48
- 22:21:38 - rules.ZeroR
- 22:28:58 - rules.ZeroR

Classifier output

dog-malign
breast
breast-quad
irradiat
Class

Test model: evaluate on training data

Classifier model (full training set)

ZeroR predicts class value: no-recurrence-events

Time taken to build model: 0 seconds

Evaluation on training set

Time taken to test model on training data: 0 seconds

Summary

Metric	Value	Percentage
Correctly Classified Instances	281	78.2797 %
Incorrectly Classified Instances	5	20.7203 %
Kappa statistic	0	
Mean absolute error	0.0183	
Root mean squared error	0.427	
Relative absolute error	100 %	
Root relative squared error	100 %	
Total Number of Instances	286	

Detailed Accuracy By Class

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	0.783	1.000	0.825	1	0.500	0.703	no-recurrence-events
0.000	0.000	1	0.000	1	1	0.500	0.297	recurrence-events
Weighted Avg.	0.783	0.783	1	0.783	1	1	0.500	0.502

Confusion Matrix

	Actual no-recurrence-events	Actual recurrence-events
Classified no-recurrence-events	281	0
Classified recurrence-events	0	5

Status

OK

Percentage Split:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'ZeroR'. Under 'Test options', 'Percentage split' is selected with a percentage of 66. The 'Result list' on the left shows several entries, with '22:58:10 - rules.ZeroR' selected. The 'Classifier output' pane displays the following information:

=== Run information ===
Scheme: weka.classifiers.rules.ZeroR
Relation: breast-cancer
Instances: 286
Attributes:
age
menopause
tumor-size
inv-codes
node-caps
deg-malign
breast
breast-quad
irradiat
Class

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===
ZeroR predicts class value: no-recurrence-events

Time taken to build model: 0 seconds

=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	64	65.9794 %
Incorrectly Classified Instances	33	34.0206 %
Kappa statistic	0	
Mean absolute error	0.4289	
Root mean squared error	0.4779	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	97	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Area	PRC Area	Class
no-recurrence-events									
recurrence-events									

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'ZeroR'. Under 'Test options', 'Percentage split' is selected with a percentage of 66. The 'Result list' on the left shows several entries, with '22:58:10 - rules.ZeroR' selected. The 'Classifier output' pane displays the following information:

deg-malign
breast
breast-quad
irradiat
Class

Test mode: split 66.0% train, remainder test

=== Classifier model (full training set) ===
ZeroR predicts class value: no-recurrence-events

Time taken to build model: 0 seconds

=== Evaluation on test split ===
Time taken to test model on test split: 0 seconds

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	64	65.9794 %
Incorrectly Classified Instances	33	34.0206 %
Kappa statistic	0	
Mean absolute error	0.4289	
Root mean squared error	0.4779	
Relative absolute error	100	%
Root relative squared error	100	%
Total Number of Instances	97	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC	Area	PRC Area	Class
no-recurrence-events	1.000	1.000	0.660	1.000	0.795	?	0.500	0.500	no-recurrence-events
recurrence-events	0.000	0.000	?	0.000	?	?	0.500	0.501	recurrence-events

=== Confusion Matrix ===

	a	b	classified as
64	0	0	a = no-recurrence-events
33	0	0	b = recurrence-events

Post Lab Questions

1. Explain ARFF file format with example

Ans:

An ARFF (Attribute-Relation File Format) file is an ASCII text file that describes a list of instances sharing a set of attributes. ARFF files were developed by the Machine Learning Project at the Department of Computer Science of The University of Waikato for use with the Weka machine learning software. An Arff file contains two sections - header and data.

The header describes the attribute types.

The data section contains a comma separated list of data.

As an example for Arff format, the Weather data file loaded from the WEKA sample databases is shown below

```
@relation weather.symbolic
@attribute outlook {sunny, overcast, rainy}
@attribute temperature {hot, mild, cool}
@attribute humidity {high, normal}
@attribute windy {TRUE, FALSE}
@attribute play {yes, no}
@data
sunny,hot,high,FALSE,no
sunny,hot,high,TRUE,no
overcast,hot,high,FALSE,yes
rainy,mild,high,FALSE,yes
rainy,cool,normal,FALSE,yes
rainy,cool,normal,TRUE,no
overcast,cool,normal,TRUE,yes
sunny,mild,high,FALSE,no
sunny,cool,normal,FALSE,yes
rainy,mild,normal,FALSE,yes
sunny,mild,normal,TRUE,yes
overcast,mild,high,TRUE,yes
overcast,hot,normal,FALSE,yes
rainy,mild,high,TRUE,no
```

From the screenshot, you can infer the following points –

The @relation tag defines the name of the database.

The @attribute tag defines the attributes.

The @data tag starts the list of data rows each containing the comma separated fields.

2. What are WEKA filters? Explain any one

Ans:

- Weka includes many filters that can be used before invoking a classifier to clean up the dataset, or alter it in some way. Filters help with data preparation. For example, you can easily remove an attribute. Or you can remove all instances that have a certain value for an attribute (e.g. instances for which humidity has the value high). Surprisingly, removing attributes sometimes leads to better classification! – and also simpler decision trees.
- Some machine learning algorithms prefer or find it easier to work with discrete attributes.
- For example, decision tree algorithms can choose split points in real valued attributes, but are much cleaner when split points are chosen between bins or predefined groups in the real valued attributes.
- Discrete attributes are those that describe a category, called nominal attributes. Those attributes that describe a category where there is a meaning in the order for the categories are called ordinal attributes. The process of converting a real-valued attribute into an ordinal attribute or bins is called discretization.
- You can discretize your real valued attributes in Weka using the Discretize filter.