



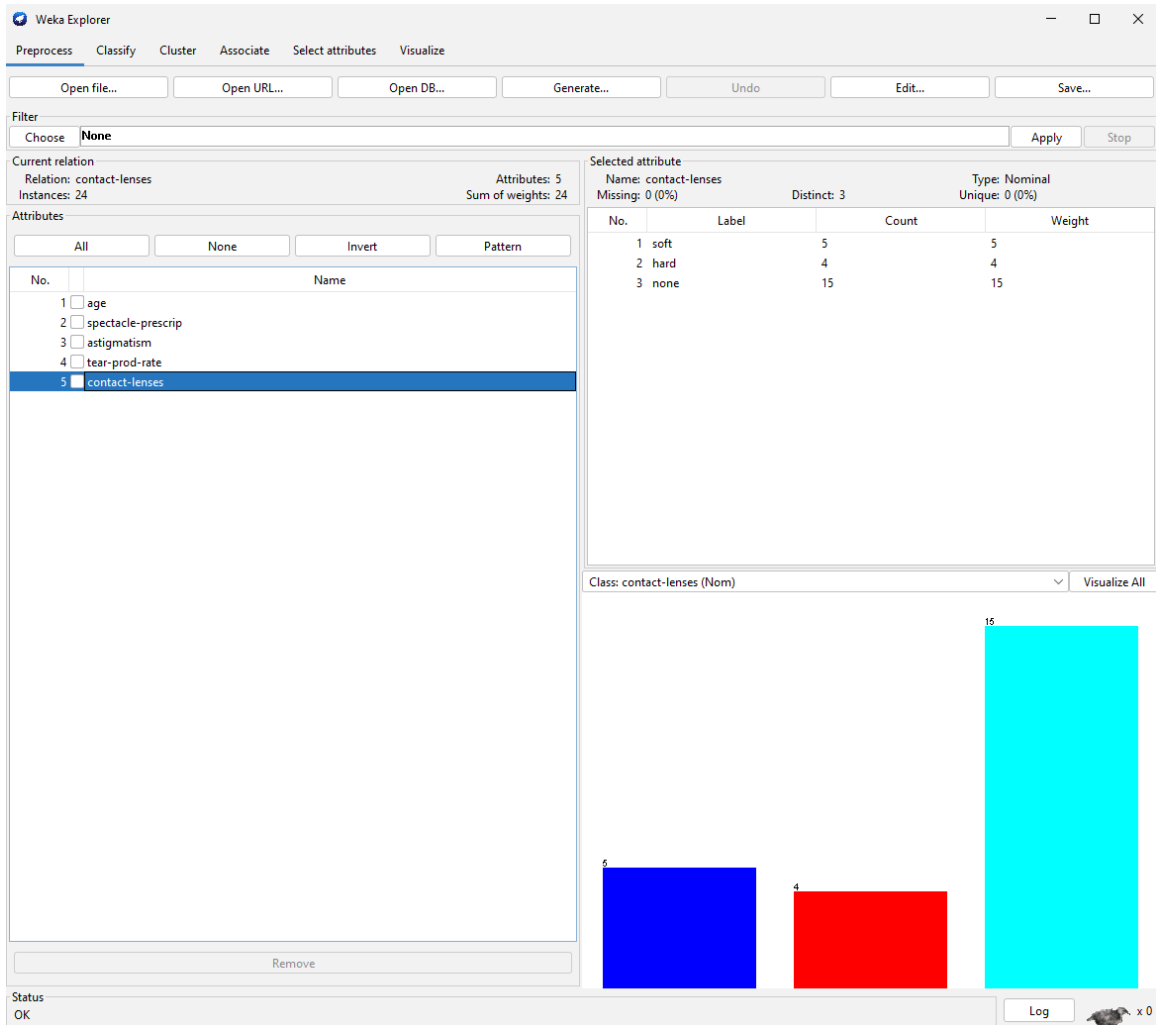
Assignment 2

DATA CLASSIFICATION USING WEKA

Matthew Acs | COP 4703 | 4/9/2022

DATASET

In this assignment, three classifiers were used to classify the contact-lenses data set using Weka. The three classifiers that were used were ZeroR, J48, and the Naive Bayes classifiers. The screenshot below shows the data set on the Weka Explorer.



ZEROR

A screenshot of the results and the text output from the ZeroR classifier are shown below. It correctly classified 62.5% of the test data.

Screenshot of Results:

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The 'Classifier' dropdown is set to 'NaiveBayes'. The 'Test options' section shows 'Cross-validation' selected with 'Folds' set to 10. The 'Result list' on the left shows three entries: '13:24:02 - rules.ZeroR' (selected), '13:25:08 - trees.J48', and '13:26:49 - bayes.NaiveBayes'. The 'Classifier output' pane displays the following text:

```

=== Run information ===

Scheme:      weka.classifiers.rules.ZeroR
Relation:    contact-lenses
Instances:   24
Attributes:  5
    age
    spectacle-prescrip
    astigmatism
    tear-prod-rate
    contact-lenses

Test mode:   10-fold cross-validation

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

ZeroR predicts class value: none

Time taken to build model: 0 seconds

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

Correctly Classified Instances      15           62.5  %
Incorrectly Classified Instances    9           37.5  %
Kappa statistic                     0
Mean absolute error                 0.3778
Root mean squared error             0.4367
Relative absolute error             100  %
Root relative squared error         100  %
Total Number of Instances          24

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
          0.000    0.000    ?          0.000    ?          ?       0.216    0.187    soft
          0.000    0.000    ?          0.000    ?          ?       0.200    0.167    hard
          1.000    1.000    0.625     1.000    0.769     ?       0.152    0.529    none
Weighted Avg.   0.625    0.625    ?          0.625    ?          ?       0.173    0.398

=== Confusion Matrix ===

  a  b  c  <-- classified as
  0  0  5 | a = soft
  0  0  4 | b = hard
  0  0 15 | c = none
  
```

The status bar at the bottom shows 'Status OK' and a 'Log' button.

Text Output of Results:

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR

Relation: contact-lenses

Instances: 24

Attributes: 5

age

spectacle-prescrip

astigmatism

tear-prod-rate

contact-lenses

Test mode: 10-fold cross-validation

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

ZeroR predicts class value: none

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	15	62.5 %
--------------------------------	----	--------

Incorrectly Classified Instances	9	37.5 %
----------------------------------	---	--------

Kappa statistic	0
-----------------	---

Mean absolute error	0.3778
---------------------	--------

Root mean squared error	0.4367
-------------------------	--------

Relative absolute error	100 %
-------------------------	-------

Root relative squared error	100 %
-----------------------------	-------

Total Number of Instances	24
---------------------------	----

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.216	0.187	soft
	0.000	0.000	?	0.000	?	?	0.200	0.167	hard
	1.000	1.000	0.625	1.000	0.769	?	0.152	0.529	none
Weighted Avg.	0.625	0.625	?	0.625	?	?	0.173	0.398	

=== Confusion Matrix ===

a b c <-- classified as

o o 5 | a = soft

o o 4 | b = hard

o o 15 | c = none

J48

A screenshot of the results and the text output from the J48 classifier are shown below. It correctly classified 83.33% of the test data.

Screenshot of Results:

Weka Explorer

Preprocess **Classify** Cluster Associate Select attributes Visualize

Classifier: Choose **NaiveBayes**

Test options

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

More options...

(Nom) contact-lenses

Start Stop

Result list (right-click for options)

- 13:24:02 - rules.ZeroR
- 13:25:08 - trees.J48**
- 13:26:49 - bayes.NaiveBayes

Classifier output

```

contact-lenses
Test mode: 10-fold cross-validation

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

J48 pruned tree
-----
tear-prod-rate = reduced: none (12.0)
tear-prod-rate = normal
| astigmatism = no: soft (6.0/1.0)
| astigmatism = yes
| | spectacle-prescrip = myope: hard (3.0)
| | spectacle-prescrip = hypermetrope: none (3.0/1.0)

Number of Leaves : 4
Size of the tree : 7

Time taken to build model: 0.01 seconds

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

Correctly Classified Instances      20          83.3333 %
Incorrectly Classified Instances    4           16.6667 %
Kappa statistic                    0.71
Mean absolute error                 0.15
Root mean squared error             0.3249
Relative absolute error             39.7059 %
Root relative squared error         74.3898 %
Total Number of Instances          24

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
          1.000    0.053    0.833     1.000    0.909     0.889    0.947    0.833    soft
          0.750    0.100    0.600     0.750    0.667     0.596    0.813    0.592    hard
          0.800    0.111    0.923     0.800    0.857     0.669    0.811    0.865    none
Weighted Avg.   0.833    0.097    0.851     0.833    0.836     0.703    0.840    0.813

=== Confusion Matrix ===

  a  b  c  <-- classified as
  5  0  0 | a = soft
  0  3  1 | b = hard
  1  2 12 | c = none

```

Status OK

Log x 0

Text Output of Results:

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: contact-lenses

Instances: 24

Attributes: 5

age

spectacle-prescrip

astigmatism

tear-prod-rate

contact-lenses

Test mode: 10-fold cross-validation

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

J48 pruned tree

tear-prod-rate = reduced: none (12.0)

tear-prod-rate = normal

| astigmatism = no: soft (6.0/1.0)

| astigmatism = yes

| | spectacle-prescrip = myope: hard (3.0)

| | spectacle-prescrip = hypermetrope: none (3.0/1.0)

Number of Leaves : 4

Size of the tree : 7

Time taken to build model: 0.01 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	20	83.3333 %
Incorrectly Classified Instances	4	16.6667 %
Kappa statistic	0.71	
Mean absolute error	0.15	
Root mean squared error	0.3249	
Relative absolute error	39.7059 %	
Root relative squared error	74.3898 %	
Total Number of Instances	24	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.053	0.833	1.000	0.909	0.889	0.947	0.833	soft
	0.750	0.100	0.600	0.750	0.667	0.596	0.813	0.592	hard
	0.800	0.111	0.923	0.800	0.857	0.669	0.811	0.865	none
Weighted Avg.	0.833	0.097	0.851	0.833	0.836	0.703	0.840	0.813	

=== Confusion Matrix ===

a b c <-- classified as

5 0 0 | a = soft

0 3 1 | b = hard

1 2 12 | c = none

NAIVE BAYES

A screenshot of the results and the text output from the Naive Bayes classifier are shown below. It correctly classified 70.833% of the test data.

Screenshot of Results:

Classifier
Choose **NaiveBayes**

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

(Nom) contact-lenses
 Start Stop

Result list (right-click for options)
 13:24:02 - rules.ZeroR
 13:25:08 - trees.J48
 13:26:49 - bayes.NaiveBayes

Classifier output

Attribute	soft	hard	none
presbyopic	2.0	2.0	7.0
[total]	8.0	7.0	18.0
spectacle-prescrip	3.0	4.0	8.0
myope	3.0	4.0	8.0
hypermetrope	4.0	2.0	9.0
[total]	7.0	6.0	17.0
astigmatism	6.0	1.0	8.0
no	6.0	1.0	8.0
yes	1.0	5.0	9.0
[total]	7.0	6.0	17.0
tear-prod-rate	1.0	1.0	13.0
reduced	1.0	1.0	13.0
normal	6.0	5.0	4.0
[total]	7.0	6.0	17.0

Time taken to build model: 0 seconds

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

Metric	Value	Percentage
Correctly Classified Instances	17	70.8333 %
Incorrectly Classified Instances	7	29.1667 %
Kappa statistic	0.4391	
Mean absolute error	0.2545	
Root mean squared error	0.3326	
Relative absolute error	67.3578 %	
Root relative squared error	76.1544 %	
Total Number of Instances	24	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.800	0.053	0.800	0.800	0.800	0.800	0.747	0.947	0.710	soft
0.250	0.100	0.333	0.250	0.286	0.169	0.925	0.692	0.692	hard
0.800	0.444	0.750	0.800	0.774	0.365	0.830	0.930	0.930	none
Weighted Avg.	0.708	0.305	0.691	0.708	0.698	0.412	0.870	0.844	

=== Confusion Matrix ===

a	b	c	<-- classified as
4	0	1	a = soft
0	1	3	b = hard
1	2	12	c = none

Status: OK

Log x 0

Text Output of Results:

=== Run information ===

Scheme: weka.classifiers.bayes.NaiveBayes

Relation: contact-lenses

Instances: 24

Attributes: 5

age

spectacle-prescrip

astigmatism

tear-prod-rate

contact-lenses

Test mode: 10-fold cross-validation

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

Naive Bayes Classifier

Class

Attribute soft hard none

(0.22) (0.19) (0.59)

=====

age

young 3.0 3.0 5.0

pre-presbyopic 3.0 2.0 6.0

presbyopic 2.0 2.0 7.0

[total] 8.0 7.0 18.0

spectacle-prescrip

myope 3.0 4.0 8.0

hypermetrope 4.0 2.0 9.0

[total]	7.0	6.0	17.0
---------	-----	-----	------

astigmatism

no	6.0	1.0	8.0
----	-----	-----	-----

yes	1.0	5.0	9.0
-----	-----	-----	-----

[total]	7.0	6.0	17.0
---------	-----	-----	------

tear-prod-rate

reduced	1.0	1.0	13.0
---------	-----	-----	------

normal	6.0	5.0	4.0
--------	-----	-----	-----

[total]	7.0	6.0	17.0
---------	-----	-----	------

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	17	70.8333 %
--------------------------------	----	-----------

Incorrectly Classified Instances	7	29.1667 %
----------------------------------	---	-----------

Kappa statistic	0.4381
-----------------	--------

Mean absolute error	0.2545
---------------------	--------

Root mean squared error	0.3326
-------------------------	--------

Relative absolute error	67.3578 %
-------------------------	-----------

Root relative squared error	76.1544 %
-----------------------------	-----------

Total Number of Instances	24
---------------------------	----

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.800	0.053	0.800	0.800	0.800	0.747	0.947	0.710	soft
	0.250	0.100	0.333	0.250	0.286	0.169	0.925	0.692	hard
	0.800	0.444	0.750	0.800	0.774	0.365	0.830	0.930	none
Weighted Avg.	0.708	0.305	0.691	0.708	0.698	0.412	0.870	0.844	

=== Confusion Matrix ===

a b c <-- classified as

4 0 1 | a = soft

0 1 3 | b = hard

1 2 12 | c = none

CONCLUSIONS

Clearly, the results show that the J48 classifier was the most accurate in classifying the contact-lenses data set.