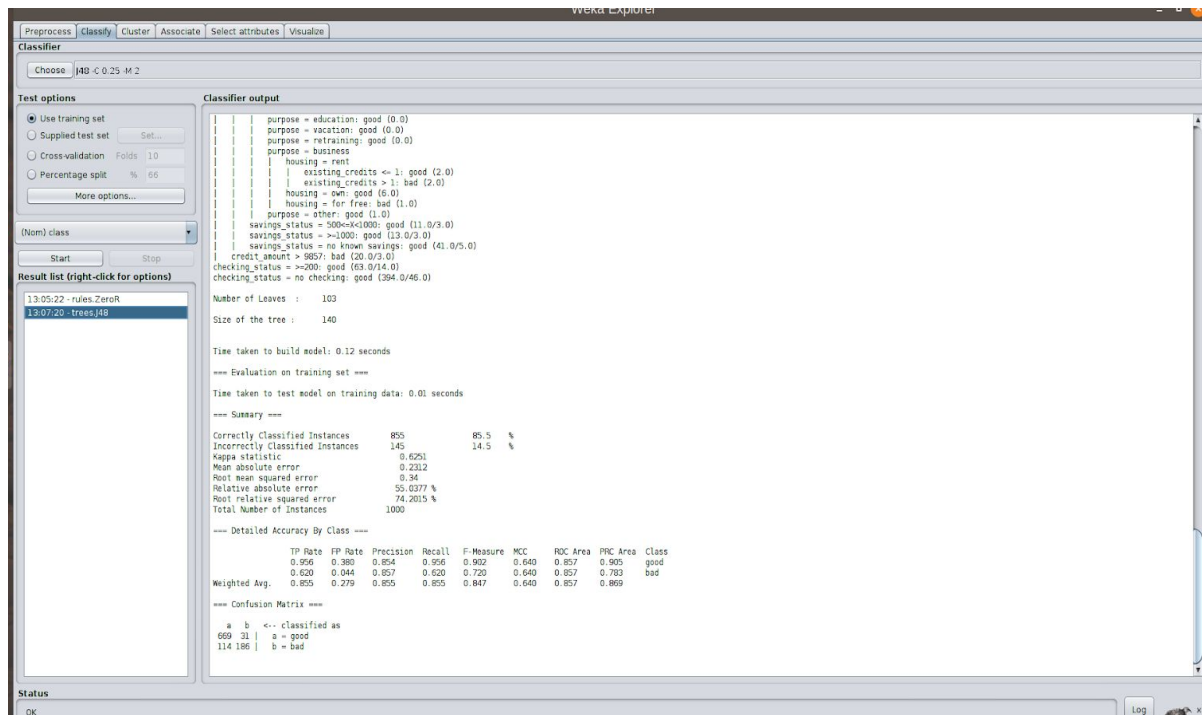


Exploring German credit dataset using Decision trees

- 1) The class is classified using decision trees using the following steps :
 - a) Open weka and load the dataset
 - b) Select class attribute and click on the classify button
 - c) Select training set option and then J-48 classifier
 - d) Click on the start button to get classifier output



- 2) Size of the tree : 140

Number of leaves : 103

Time taken to build the model : 0.01 sec.

Default percentage split : 66%

Default cross validation folds : 10

Accuracy : 72.6471%.

3) Percentage split : 70%

Default cross validation folds : 10

Accuracy : 73.66%

Classifier
Choose J48 C 0.25 M 2

Test options
☐ Use training set
☐ Supplied test set
☐ Cross-validation Folds 10
☒ Percentage split % 70
More options...

(Nom) class
Start Stop

Result list (right-click for options)
13:05:22 - rules.ZeroR
13:07:20 - trees.J48
13:15:37 - trees.J48
13:16:17 - trees.J48

Classifier output
| | | purpose = education: good (0.0)
| | | purpose = vacation: good (0.0)
| | | purpose = retraining: good (0.0)
| | | purpose = business
| | | | housing = rent
| | | | existing_credits <= 1: good (2.0)
| | | | existing_credits > 1: bad (2.0)
| | | | housing = own: good (6.0)
| | | | housing = for free: bad (1.0)
| | | | purpose = other: good (1.0)
| | | savings_status = 500-<x<1000: good (11.0/3.0)
| | | savings_status = >=1000: good (13.0/3.0)
| | | savings_status = no known savings: good (41.0/5.0)
| | credit_amount > 9857: bad (20.0/3.0)
| checking_status = >=200: good (63.0/14.0)
| checking_status = no checking: good (394.0/46.0)
Number of Leaves : 103
Size of the tree : 140
Time taken to build model: 0.03 seconds
==== Evaluation on test split ====
Time taken to test model on test split: 0 seconds
==== Summary ====
Correctly Classified Instances 221 73.6667 %
Incorrectly Classified Instances 79 26.3333 %
Kappa statistic 0.2579
Mean absolute error 0.323
Root mean squared error 0.47
Relative absolute error 78.2126 %
Root relative squared error 105.9524 %
Total Number of Instances 300
==== Detailed Accuracy By Class ====
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
Weighted Avg. 0.737 0.501 0.716 0.737 0.722 0.263 0.636 0.424 bad
==== Confusion Matrix ====
a b <- classified as
102 29 | a = good
50 29 | b = bad

Status
OK Log

4) Percentage split : 75%

Default cross validation folds : 10

Accuracy : 76%

Classifier
Choose J48 C 0.25 M 2

Test options
☐ Use training set
☐ Supplied test set
☐ Cross-validation Folds 10
☒ Percentage split % 75
More options...

(Nom) class
Start Stop

Result list (right-click for options)
13:05:22 - rules.ZeroR
13:07:20 - trees.J48
13:15:37 - trees.J48
13:16:27 - trees.J48
13:16:35 - trees.J48

Classifier output
| | | purpose = education: good (0.0)
| | | purpose = vacation: good (0.0)
| | | purpose = retraining: good (0.0)
| | | purpose = business
| | | | housing = rent
| | | | existing_credits <= 1: good (2.0)
| | | | existing_credits > 1: bad (2.0)
| | | | housing = own: good (6.0)
| | | | housing = for free: bad (1.0)
| | | | purpose = other: good (1.0)
| | | savings_status = 500-<x<1000: good (11.0/3.0)
| | | savings_status = >=1000: good (13.0/3.0)
| | | savings_status = no known savings: good (41.0/5.0)
| | credit_amount > 9857: bad (20.0/3.0)
| checking_status = >=200: good (63.0/14.0)
| checking_status = no checking: good (394.0/46.0)
Number of Leaves : 103
Size of the tree : 140
Time taken to build model: 0.01 seconds
==== Evaluation on test split ====
Time taken to test model on test split: 0 seconds
==== Summary ====
Correctly Classified Instances 190 76 %
Incorrectly Classified Instances 60 24 %
Kappa statistic 0.3032
Mean absolute error 0.3073
Root mean squared error 0.4365
Relative absolute error 74.6888 %
Root relative squared error 98.4212 %
Total Number of Instances 250
==== Detailed Accuracy By Class ====
TP Rate FP Rate Precision Recall F-Measure MCC ROC Area PRC Area Class
Weighted Avg. 0.760 0.465 0.742 0.760 0.747 0.330 0.673 0.478 bad
==== Confusion Matrix ====
a b <- classified as
163 21 | a = good
39 27 | b = bad

5) Percentage split : 80%

Default cross validation folds : 10

Accuracy : 77%

The screenshot shows the Orange3 Classifier window. The 'Test options' section has 'Percentage split' selected with a value of 80. The 'Classifier output' section displays the following information:

- Number of Leaves : 103
- Size of the tree : 140
- Time taken to build model: 0.02 seconds
- Time taken to test model on test split: 0 seconds
- === Evaluation on test split ===
- === Summary ===
- Correctly Classified Instances: 154 (77 %)
- Incorrectly Classified Instances: 46 (23 %)
- Kappa statistic: 0.3867
- Mean absolute error: 0.2947
- Root mean squared error: 0.4439
- Relative absolute error: 72.2746 %
- Root relative squared error: 100.8586 %
- Total Number of Instances: 200
- === Detailed Accuracy By Class ===
- Confusion Matrix: a = good, b = bad

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.852	0.471	0.540	0.847	0.387	0.691	0.487	good	
	0.529	0.148	0.551	0.529	0.540	0.387	0.691	0.487	bad

6) Percentage split : 85%

Default cross validation folds : 10

Accuracy : 72%

The screenshot shows the Orange3 Classifier window. The 'Test options' section has 'Percentage split' selected with a value of 85. The 'Classifier output' section displays the following information:

- Number of Leaves : 103
- Size of the tree : 140
- Time taken to build model: 0.01 seconds
- Time taken to test model on test split: 0 seconds
- === Evaluation on test split ===
- === Summary ===
- Correctly Classified Instances: 108 (72 %)
- Incorrectly Classified Instances: 42 (28 %)
- Kappa statistic: 0.2182
- Mean absolute error: 0.3225
- Root mean squared error: 0.4719
- Relative absolute error: 82.3680 %
- Root relative squared error: 108.3147 %
- Total Number of Instances: 150
- === Detailed Accuracy By Class ===
- Confusion Matrix: a = good, b = bad

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.832	0.622	0.803	0.832	0.817	0.219	0.650	0.802	good
	0.378	0.168	0.424	0.378	0.400	0.219	0.650	0.392	bad

