

German Credit dataset

Classifying a dataset

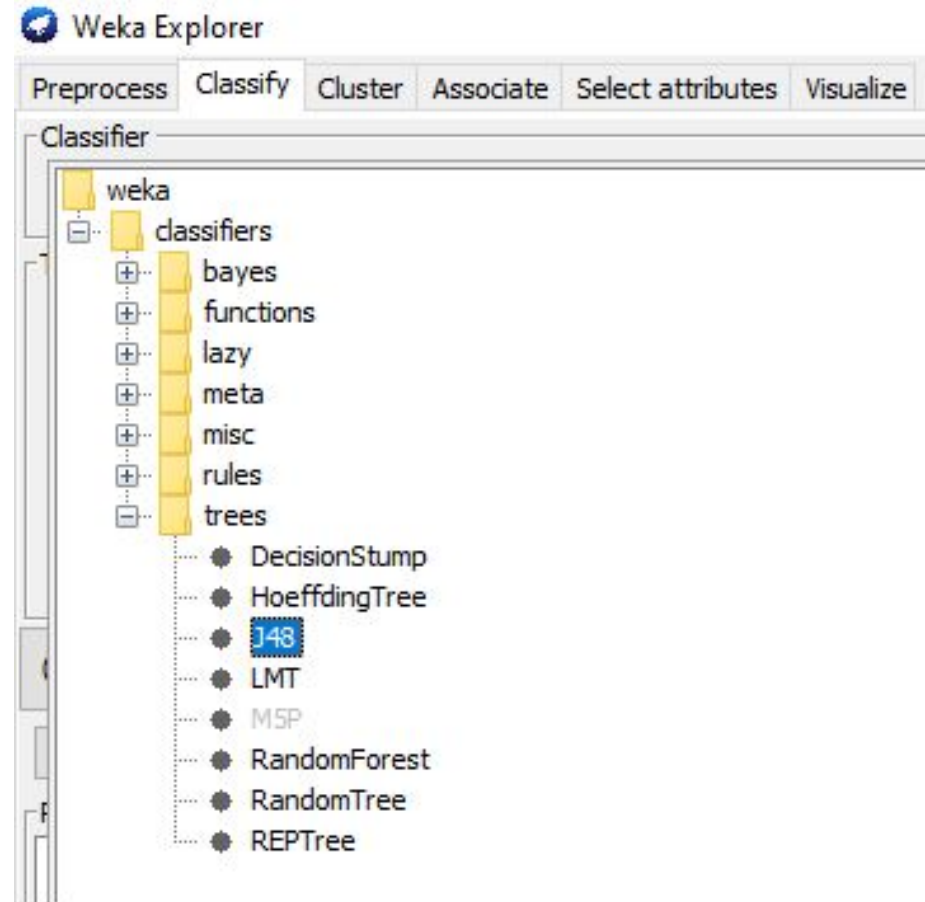
Choose a dataset with class attribute in it.

Select Classify

Choose - **Trees** and select J48

(Nom) - Class

Start the process



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)

20:22:15 - trees.J48

20:33:52 - trees.J48

Classifier output

Time taken to build model: 0.02 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances	855	85.5	%
Incorrectly Classified Instances	145	14.5	%
Kappa statistic	0.6251		
Mean absolute error	0.2312		
Root mean squared error	0.34		
Relative absolute error	55.0377	%	
Root relative squared error	74.2015	%	
Total Number of Instances	1000		

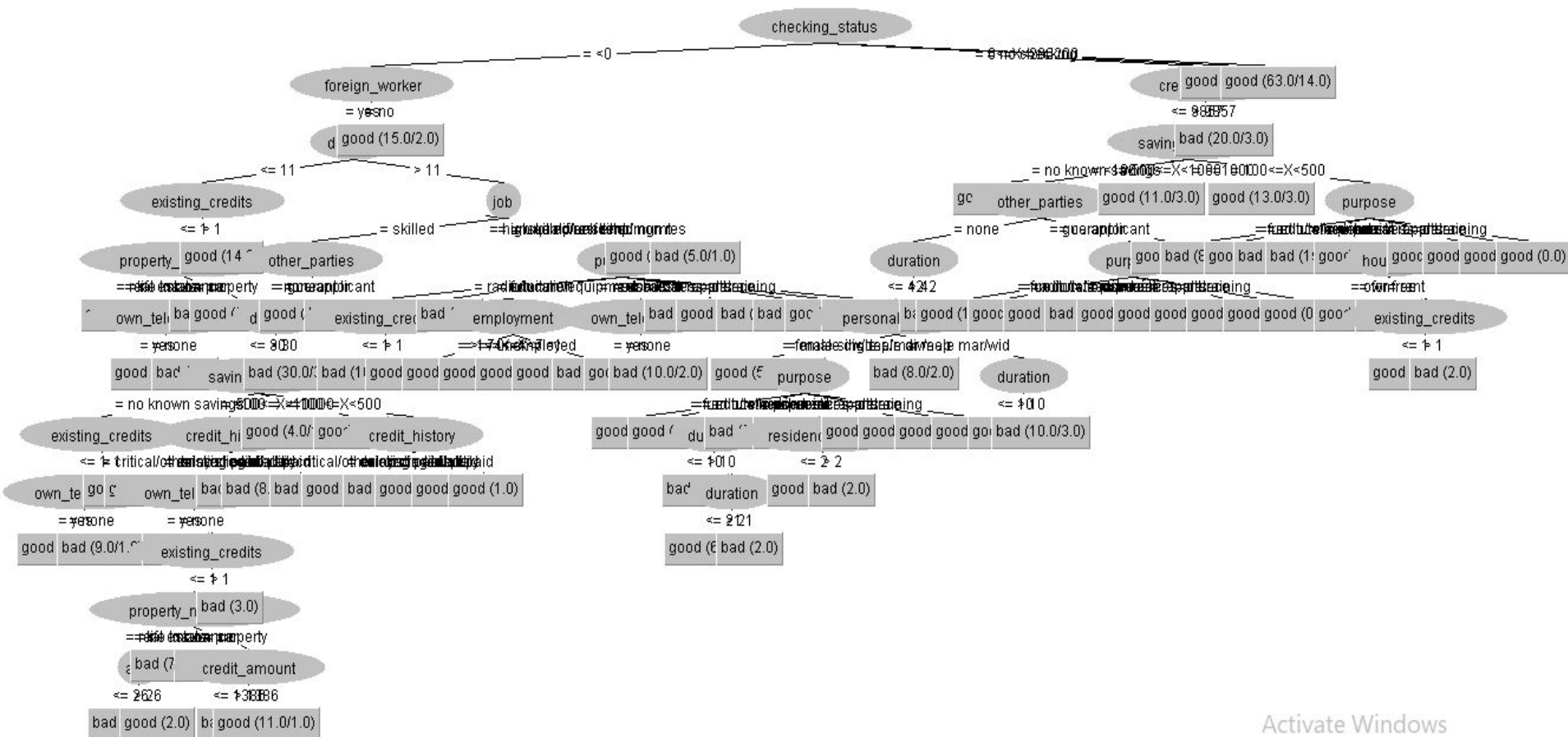
=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.956	0.380	0.854	0.956	0.902	0.640	0.857	0.905	good
	0.620	0.044	0.857	0.620	0.720	0.640	0.857	0.783	bad
Weighted Avg.	0.855	0.279	0.855	0.855	0.847	0.640	0.857	0.869	

=== Confusion Matrix ===

```
a  b  <-- classified as
669 31 |  a = good
114 186 |  b = bad
```

Visualize Tree



Activate Windows
Go to Settings to activate Windows.

No of leaves : 98

Size of the tree 135

Time taken to build model 0.02sec

Time taken to test model on training data: 0.02 seconds

Confusion matrix

a	b	classified as
---	---	---------------

669	31	a = good
-----	----	----------

114	186	b = bad
-----	-----	---------

Correctly classified instance 85.5%

Incorrectly classified instance 14.5%

Changes in percentage split to find the best accuracy

Percentage split - 65%

Test options

☐ Use training set

☐ Supplied test set

Set...

☐ Cross-validation Folds

10

☒ Percentage split %

65

More options...

(Nom) class

▼

Start

Stop

Result list (right-click for options)

20:22:15 - trees.J48

20:33:52 - trees.J48

22:03:23 - trees.J48

22:04:55 - trees.J48

Classifier output

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	257	73.4286 %
Incorrectly Classified Instances	93	26.5714 %
Kappa statistic	0.2971	
Mean absolute error	0.3305	
Root mean squared error	0.4721	
Relative absolute error	79.6609 %	
Root relative squared error	106.3358 %	
Total Number of Instances	350	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.833	0.543	0.811	0.833	0.822	0.298	0.629	0.795	good
	0.457	0.167	0.494	0.457	0.475	0.298	0.629	0.379	bad
Weighted Avg.	0.734	0.444	0.728	0.734	0.731	0.298	0.629	0.686	

=== Confusion Matrix ===

a

b

<-- classified as

215

43

|

a = good

50

42

|

b = bad

Percentage split - 67%

Test options

☐ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds

10

☒ Percentage split

%

97

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

20:22:15 - trees.J48

20:33:52 - trees.J48

22:03:23 - trees.J48

22:04:55 - trees.J48

22:08:27 - trees.J48

22:08:33 - trees.J48

22:08:40 - trees.J48

22:08:47 - trees.J48

22:08:52 - trees.J48

22:08:58 - trees.J48

22:09:04 - trees.J48

22:09:13 - trees.J48

22:09:18 - trees.J48

22:09:24 - trees.J48

22:09:29 - trees.J48

22:09:35 - trees.J48

22:09:40 - trees.J48

Classifier output

Time taken to build model: 0.05 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0 seconds

=== Summary ===

Correctly Classified Instances 23 76.6667 %

Incorrectly Classified Instances 7 23.3333 %

Kappa statistic 0.4615

Mean absolute error 0.3113

Root mean squared error 0.4544

Relative absolute error 71.8682 %

Root relative squared error 96.1535 %

Total Number of Instances 30

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.850	0.400	0.810	0.850	0.829	0.463	0.688	0.753	good
	0.600	0.150	0.667	0.600	0.632	0.463	0.688	0.658	bad
Weighted Avg.	0.767	0.317	0.762	0.767	0.763	0.463	0.688	0.721	

=== Confusion Matrix ===

a b <-- classified as

17 3 | a = good

4 6 | b = bad

Percentage split - 79%

Test options

☐ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds 10

☒ Percentage split

% 79

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

22:09:13 - trees.J48
22:09:18 - trees.J48
22:09:24 - trees.J48
22:09:29 - trees.J48
22:09:35 - trees.J48
22:09:40 - trees.J48
22:12:32 - trees.J48
22:12:37 - trees.J48
22:12:47 - trees.J48
22:12:54 - trees.J48
22:12:59 - trees.J48
22:13:03 - trees.J48
22:13:07 - trees.J48
22:13:11 - trees.J48
22:13:17 - trees.J48
22:13:21 - trees.J48
22:13:25 - trees.J48
22:13:29 - trees.J48
22:13:35 - trees.J48
22:13:41 - trees.J48

Classifier output

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	164	78.0952 %
Incorrectly Classified Instances	46	21.9048 %
Kappa statistic	0.4196	
Mean absolute error	0.2853	
Root mean squared error	0.4331	
Relative absolute error	69.8259 %	
Root relative squared error	98.3263 %	
Total Number of Instances	210	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.859	0.444	0.848	0.859	0.854	0.420	0.711	0.840	good
	0.556	0.141	0.577	0.556	0.566	0.420	0.711	0.515	bad
Weighted Avg.	0.781	0.366	0.778	0.781	0.780	0.420	0.711	0.756	

=== Confusion Matrix ===

```
a  b  <-- classified as
134 22 |  a = good
 24 30 |  b = bad
```


Percentage split - 79%

Correctly classified instances 78.0952%

Changes in cross validation to find the best accuracy

Cross validation folds 8

☐ Supplied test set Set...
☒ Cross-validation Folds
☐ Percentage split %
 More options...

(Nom) class ▾

Start

Stop

Result list (right-click for options)

22:20:31 - trees.J48
 22:20:35 - trees.J48
 22:20:40 - trees.J48
 22:20:45 - trees.J48
 22:20:51 - trees.J48
 22:20:57 - trees.J48
 22:21:02 - trees.J48
 22:21:07 - trees.J48
 22:21:18 - trees.J48
 22:21:23 - trees.J48
 22:21:28 - trees.J48
 22:21:33 - trees.J48
 22:21:37 - trees.J48
 22:21:42 - trees.J48
 22:21:48 - trees.J48
 22:21:54 - trees.J48
 22:21:59 - trees.J48
 22:22:03 - trees.J48
 22:22:10 - trees.J48
 22:22:15 - trees.J48
 22:22:38 - trees.J48

Size of the tree : 135

Time taken to build model: 0.03 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	726	72.6	%
Incorrectly Classified Instances	274	27.4	%
Kappa statistic	0.2996		
Mean absolute error	0.3319		
Root mean squared error	0.4692		
Relative absolute error	78.9988	%	
Root relative squared error	102.3972	%	
Total Number of Instances	1000		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.856	0.577	0.776	0.856	0.814	0.305	0.663	0.765	good
	0.423	0.144	0.557	0.423	0.481	0.305	0.663	0.469	bad
Weighted Avg.	0.726	0.447	0.710	0.726	0.714	0.305	0.663	0.676	

=== Confusion Matrix ===

```

  a  b  <-- classified as
599 101 |  a = good
173 127 |  b = bad

```

NaiveBayes

Choose **NaiveBayes**

Test options

☒ Use training set

☐ Supplied test set

Set...

☐ Cross-validation

Folds

8

☐ Percentage split

%

54

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

22:20:13 - trees.J48
22:20:51 - trees.J48
22:20:57 - trees.J48
22:21:02 - trees.J48
22:21:07 - trees.J48
22:21:18 - trees.J48
22:21:23 - trees.J48
22:21:28 - trees.J48
22:21:33 - trees.J48
22:21:37 - trees.J48
22:21:42 - trees.J48
22:21:48 - trees.J48
22:21:54 - trees.J48
22:21:59 - trees.J48
22:22:03 - trees.J48
22:22:10 - trees.J48
22:22:15 - trees.J48
22:22:38 - trees.J48
22:22:43 - trees.J48
22:25:46 - bayes.NaiveBayes
22:26:20 - trees.J48
22:26:38 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0.01 seconds

=== Evaluation on training set ===

Time taken to test model on training data: 0.02 seconds

=== Summary ===

Correctly Classified Instances	772	77.2	%
Incorrectly Classified Instances	228	22.8	%
Kappa statistic	0.43		
Mean absolute error	0.2822		
Root mean squared error	0.4077		
Relative absolute error	67.1611	%	
Root relative squared error	88.9725	%	
Total Number of Instances	1000		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.873	0.463	0.815	0.873	0.843	0.433	0.809	0.906	good
	0.537	0.127	0.644	0.537	0.585	0.433	0.809	0.609	bad
Weighted Avg.	0.772	0.362	0.763	0.772	0.766	0.433	0.809	0.817	

=== Confusion Matrix ===

```
a  b  <-- classified as
611 89 |  a = good
139 161 |  b = bad
```

Percentage split 62%

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

☐ Cross-validation Folds

☒ Percentage split %

(Nom) class

Result list (right-click for options)

23:17:33 - bayes.NaiveBayes
23:17:37 - bayes.NaiveBayes
23:17:41 - bayes.NaiveBayes
23:17:44 - bayes.NaiveBayes
23:17:47 - bayes.NaiveBayes
23:17:50 - bayes.NaiveBayes
23:17:54 - bayes.NaiveBayes
23:17:58 - bayes.NaiveBayes
23:18:01 - bayes.NaiveBayes
23:18:05 - bayes.NaiveBayes
23:18:09 - bayes.NaiveBayes
23:18:13 - bayes.NaiveBayes
23:18:18 - bayes.NaiveBayes
23:18:23 - bayes.NaiveBayes
23:18:29 - bayes.NaiveBayes
23:18:37 - bayes.NaiveBayes
23:18:43 - bayes.NaiveBayes
23:19:01 - bayes.NaiveBayes
23:19:10 - bayes.NaiveBayes
23:19:21 - bayes.NaiveBayes
23:19:25 - bayes.NaiveBayes

Classifier output

Time taken to build model: 0 seconds

=== Evaluation on test split ===

Time taken to test model on test split: 0.01 seconds

=== Summary ===

Correctly Classified Instances	295	77.6316 %
Incorrectly Classified Instances	85	22.3684 %
Kappa statistic	0.4214	
Mean absolute error	0.2817	
Root mean squared error	0.3969	
Relative absolute error	67.7259 %	
Root relative squared error	89.6334 %	
Total Number of Instances	380	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.844	0.418	0.853	0.844	0.848	0.421	0.803	0.922	good
	0.582	0.156	0.564	0.582	0.573	0.421	0.803	0.601	bad
Weighted Avg.	0.776	0.351	0.779	0.776	0.777	0.421	0.803	0.839	

=== Confusion Matrix ===

a	b	<-- classified as
238	44	a = good
41	57	b = bad

CROSS VALIDATION FOLDS 8

Choose **NaiveBayes**

Test options

☐ Use training set

☐ Supplied test set

Set...

☒ Cross-validation

Folds

8

☐ Percentage split

%

62

More options...

(Nom) class

Start

Stop

Result list (right-click for options)

23:18:13 - bayes.NaiveBayes
23:18:18 - bayes.NaiveBayes
23:18:23 - bayes.NaiveBayes
23:18:29 - bayes.NaiveBayes
23:18:37 - bayes.NaiveBayes
23:18:43 - bayes.NaiveBayes
23:19:01 - bayes.NaiveBayes
23:19:10 - bayes.NaiveBayes
23:19:21 - bayes.NaiveBayes
23:19:25 - bayes.NaiveBayes
23:19:31 - bayes.NaiveBayes
23:21:52 - bayes.NaiveBayes
23:21:57 - bayes.NaiveBayes
23:22:03 - bayes.NaiveBayes
23:22:08 - bayes.NaiveBayes
23:22:12 - bayes.NaiveBayes
23:22:16 - bayes.NaiveBayes
23:22:20 - bayes.NaiveBayes
23:22:24 - bayes.NaiveBayes
23:22:29 - bayes.NaiveBayes
23:22:34 - bayes.NaiveBayes

Classifier output

```
no          34.0      5.0
[total]     702.0     302.0
```

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

```
Correctly Classified Instances      759           75.9   %
Incorrectly Classified Instances    241           24.1   %
Kappa statistic                     0.3957
Mean absolute error                  0.2937
Root mean squared error              0.4205
Relative absolute error              69.8897 %
Root relative squared error          91.7625 %
Total Number of Instances          1000
```

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.866	0.490	0.805	0.866	0.834	0.399	0.789	0.893	good
	0.510	0.134	0.619	0.510	0.559	0.399	0.789	0.575	bad
Weighted Avg.	0.759	0.383	0.749	0.759	0.752	0.399	0.789	0.798	

=== Confusion Matrix ===

```
  a   b  <-- classified as
606  94 |   a = good
147 153 |   b = bad
```