

J48 TREE

The screenshot shows the Weka Explorer interface with the J48 Tree classifier selected. The classifier output pane displays the following information:

Number of Leaves : 3
Size of the tree : 5
Time taken to build model: 0.06 seconds

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

==== Summary ====

Correctly Classified Instances	1599	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0		
Root mean squared error	0		
Relative absolute error	0	%	
Root relative squared error	0	%	
Total Number of Instances	1599		

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	good
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	excellent
1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000	poor
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

==== Confusion Matrix ====

a	b	c	<-- classified as
1319	0	0	a = good
0	217	0	b = excellent
0	0	63	c = poor

NAÏVE BAYES

The screenshot shows the Weka Explorer interface with the Naïve Bayes classifier selected. The classifier output pane displays the following information:

weight sum
precision
1319 217 63
1 1 1

Time taken to build model: 0.03 seconds

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

==== Summary ====

Correctly Classified Instances	1531	95.7473 %
Incorrectly Classified Instances	68	4.2527 %
Kappa statistic	0.8585	
Mean absolute error	0.0400	
Root mean squared error	0.1565	
Relative absolute error	20.3953 %	
Root relative squared error	49.5104 %	
Total Number of Instances	1599	

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.973	0.118	0.975	0.973	0.974	0.853	0.962	0.975	good	
0.935	0.010	0.935	0.935	0.935	0.925	0.983	0.967	excellent	
0.698	0.014	0.677	0.698	0.688	0.675	0.954	0.525	poor	
Weighted Avg.	0.957	0.099	0.958	0.957	0.958	0.856	0.964	0.956	

==== Confusion Matrix ====

a	b	c	<-- classified as
1284	14	21	a = good
14	203	0	b = excellent
19	0	44	c = poor

RANDOM FOREST

Weka Explorer

Preprocess Classify Cluster Associate Select attributes Visualize

Classifier

Choose RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Test options

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

(Nom) quality category

Start Stop

Result list (right-click for options)

- 11:54:58 - trees.J48
- 11:56:26 - bayes.NaiveBayes
- 11:56:45 - trees.RandomForest

Classifier output

Bagging with 100 iterations and base learner

weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S 1 -do-not-check-capabilities

Time taken to build model: 0.78 seconds

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

Correctly Classified Instances	1599	100	%
Incorrectly Classified Instances	0	0	%
Kappa statistic	1		
Mean absolute error	0.0154		
Root mean squared error	0.0424		
Relative absolute error	7.675 %		
Root relative squared error	13.4036 %		
Total Number of Instances	1599		

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	good
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	excellent
	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	poor
Weighted Avg.	1.000	0.000	1.000	1.000	1.000	1.000	1.000	1.000	

=== Confusion Matrix ===

a	b	c	<-- classified as
1319	0	0	a = good
0	217	0	b = excellent
0	0	63	c = poor

Status

OK Log x 0

1. What has been done in the past week?
 - a. The first thing I did was create a categorical attribute on the dataset using the numeric quality inputs. I used an IF statement in Excel to split the numerical quality values into three categories: *poor*, *good*, and *excellent*. Then, I imported the new dataset into Weka, and applied classification to the data using the three different classification methods.
2. You worked individually or together?
 - a. We primarily worked individually, using our group chat to communicate and break up tasks.
3. Do you have any intermediate results to show?
 - a. Using the Visualizing tab, I can see that the “excellent” wines tend to have a lower volatile acidity level, whereas lower quality wines tend to have higher values for volatile acidity. I can also see that “excellent” wines tend to have less sulphates than lower quality wines. Also, low quality wines tend not to have an alcohol level above 12%.
4. What are the challenges you faced in this activity?
 - a. Separating wine quality into categories was a little bit of a challenge. I’m sure there is a way to do it in Weka, but I couldn’t figure out how. In any case, it was easy enough to do using Excel.
5. What do you expect to do in the next week?
 - a. Next week, I imagine we’ll be looking at ways to better visualize these data.

