Desiderio Pilla
Homework 3 Individual
CISC 684

Tic-Tac-Toe Endgame Data Set

The dataset I used for this assignment can be found at:
https://archive.ics.uci.edu/ml/datasets/Tic-Tac-Toe+Endgame

This is a binary classification problem. This database encodes the complete set of possible board configurations at the end of tic-tac-toe games, where "x" is assumed to have played first. The target concept is "win for x" (i.e., true when "x" has one of 8 possible ways to create a "three-in-a-row").

Results

Algorithm accuracy	Bagging accuracy	AdaBoost accuracy
65.34 %	96.35 %	96.45 %

Both the Bagging and AdaBoost algorithms increased the accuracy *significantly*. With the algorithms, the model was able to predict the classification in almost every case. By sampling many different subsets of the data and taking an average for its predicted classification, the model is able to produce a much higher accuracy. This raw database gives a stripped-down decision tree algorithm fit, which is also why Bagging helped increase the accuracy.

Extra Credit:

Dow Jones Index Data Set: https://archive.ics.uci.edu/ml/datasets/Dow+Jones+Index

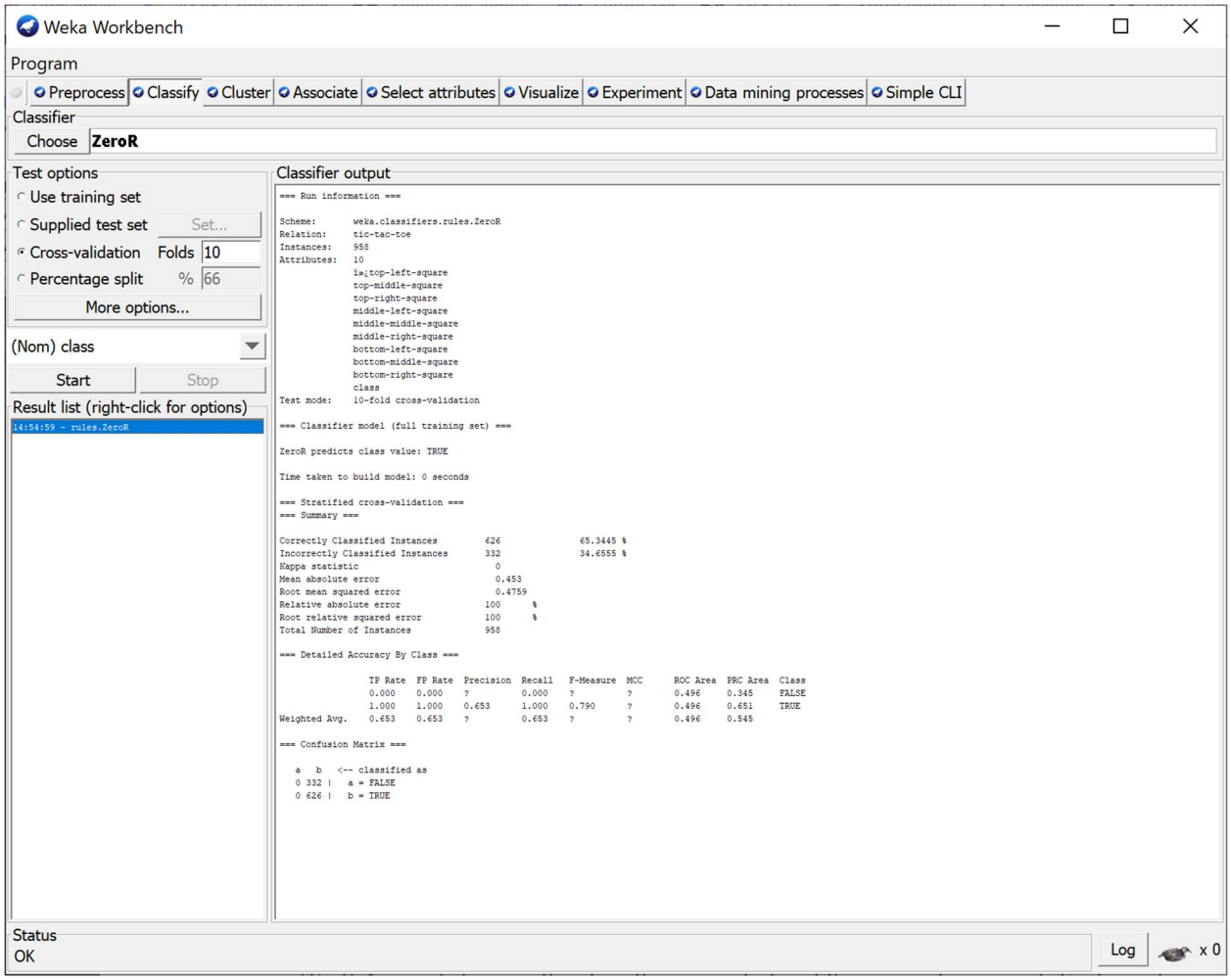
Results:

Algorithm accuracy	Bagging accuracy	AdaBoost accuracy
2.93 %	18.93 %	4.13 %

For this dataset, Bagging and AdaBoost increased the accuracy slightly. In either case, the accuracy was still far below any level of significance (much smaller than 50%). The accuracy of each individual tree alone was very poor (note the 2.93 % algorithm accuracy). By taking the average of many poor trees, the aggregated predictions (Bagging and AdaBoost) were not much better.

Weka Outputs

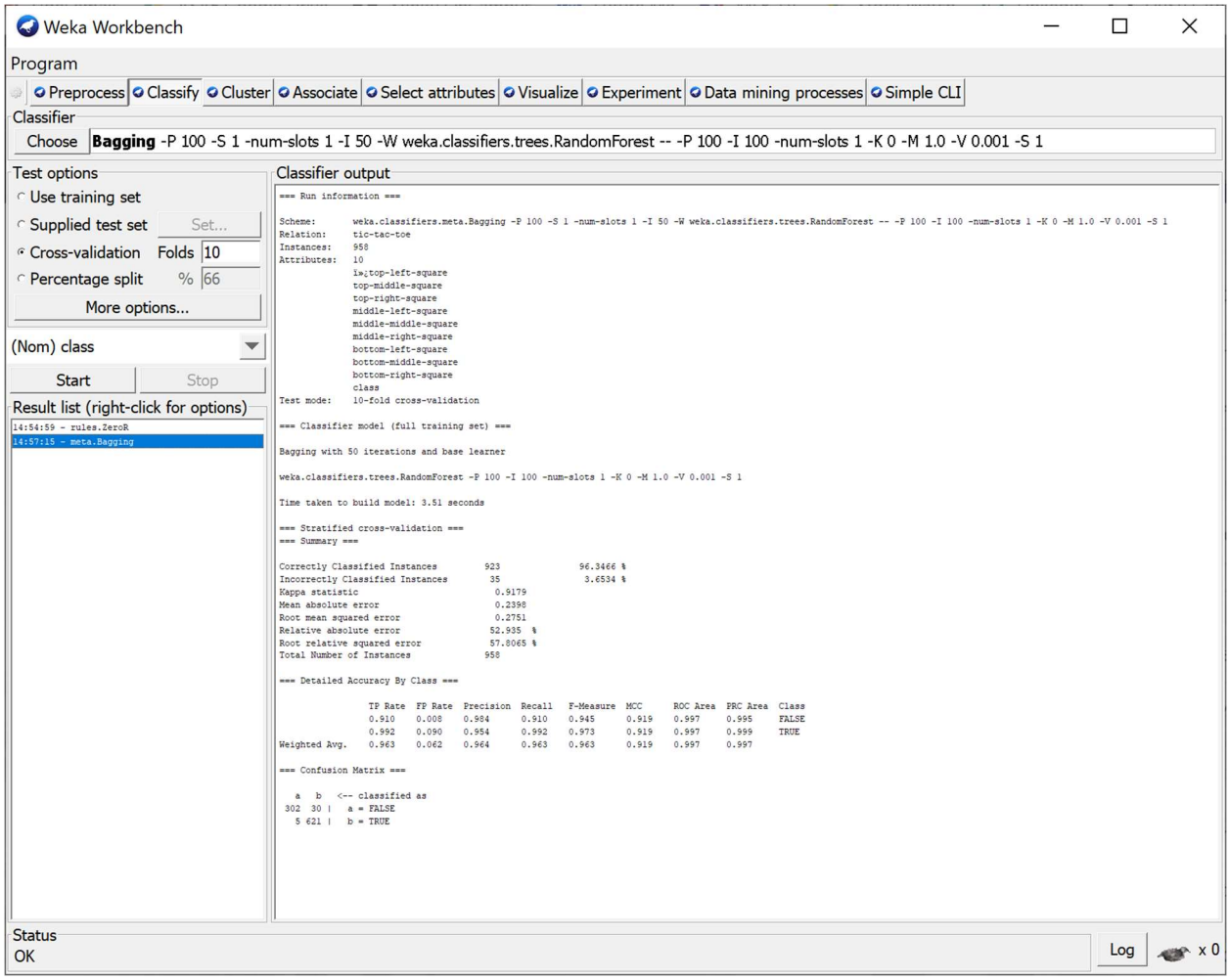
Algorithm Output (Tic-Tac-Toe)




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

Correctly Classified Instances         626               65.3445 %
Incorrectly Classified Instances       332               34.6555 %
Kappa statistic                          0
Mean absolute error                      0.453
Root mean squared error                  0.4759
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              958
```

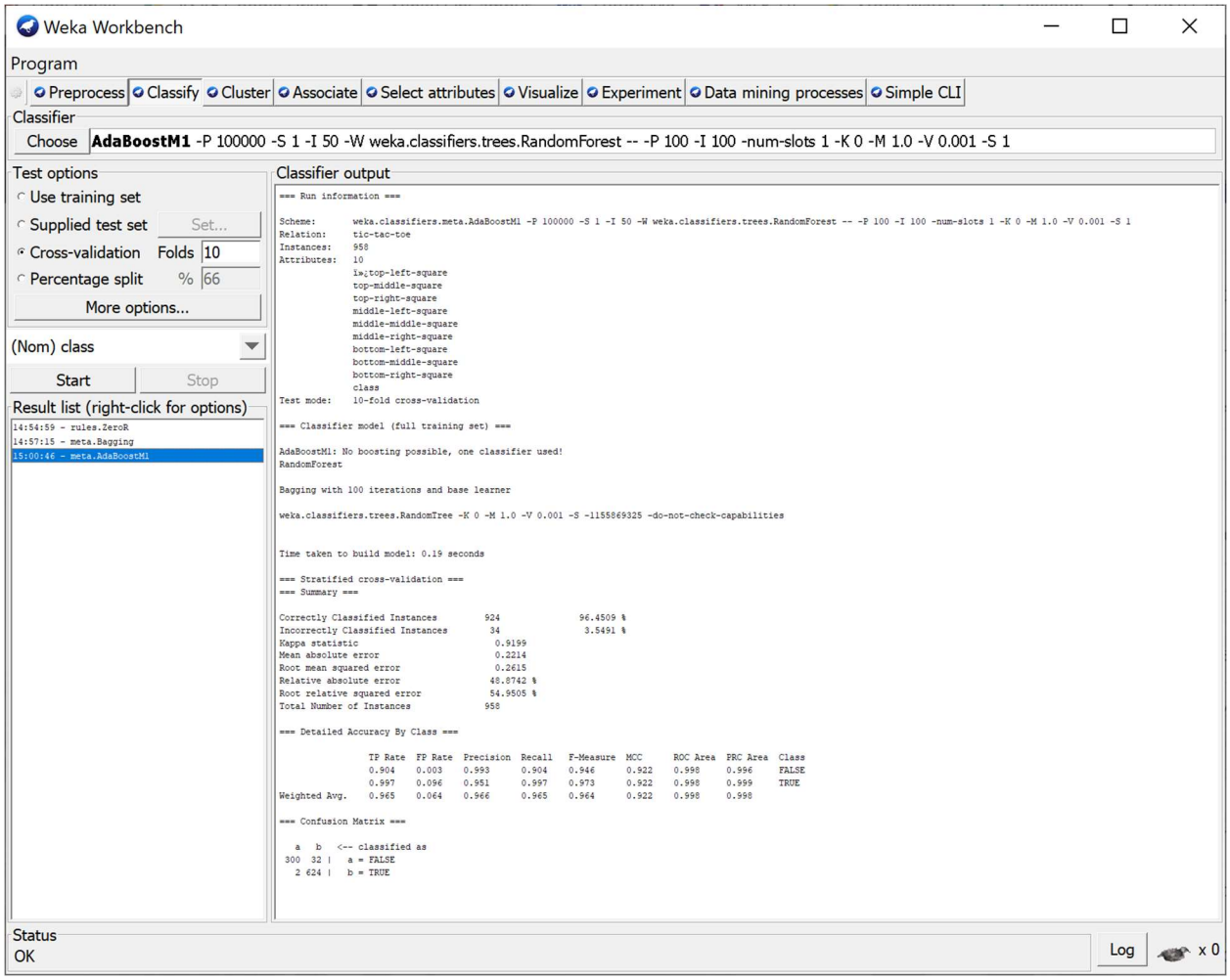
Bagging Output (Tic-Tac-Toe)




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

Correctly Classified Instances         923               96.3466 %
Incorrectly Classified Instances        35                3.6534 %
Kappa statistic                          0.9179
Mean absolute error                      0.2398
Root mean squared error                  0.2751
Relative absolute error                 52.935  %
Root relative squared error             57.8065 %
Total Number of Instances              958
```

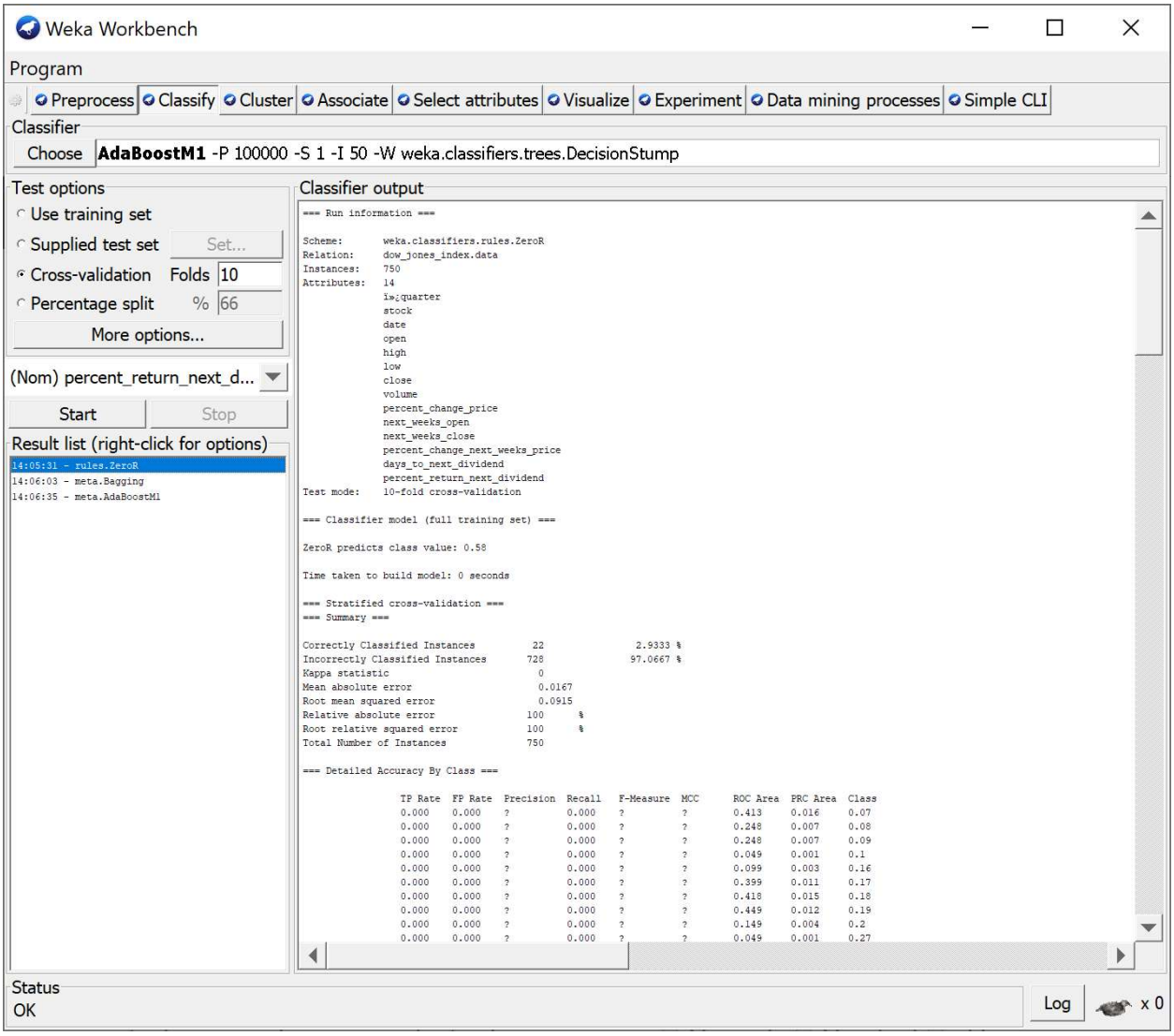
AdaBoost Output (Tic-Tac-Toe)




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

Correctly Classified Instances         924               96.4509 %
Incorrectly Classified Instances        34                3.5491 %
Kappa statistic                          0.9199
Mean absolute error                      0.2214
Root mean squared error                  0.2615
Relative absolute error                 48.8742 %
Root relative squared error             54.9505 %
Total Number of Instances              958
```

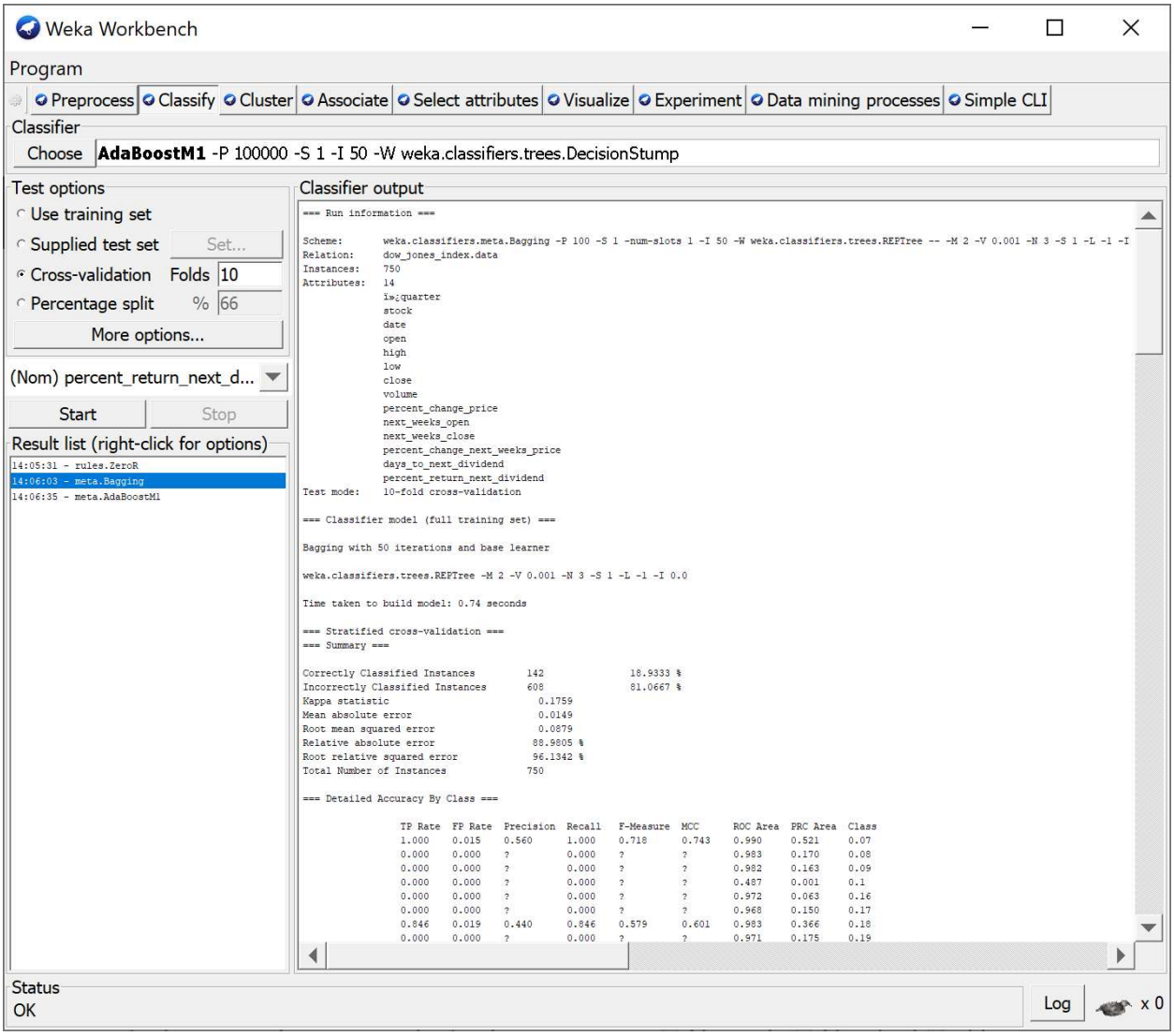
Algorithm Output (Dow Jones Index)




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

Correctly Classified Instances          22               2.9333 %
Incorrectly Classified Instances       728              97.0667 %
Kappa statistic                          0
Mean absolute error                      0.0167
Root mean squared error                  0.0915
Relative absolute error                100      %
Root relative squared error            100      %
Total Number of Instances              750
```

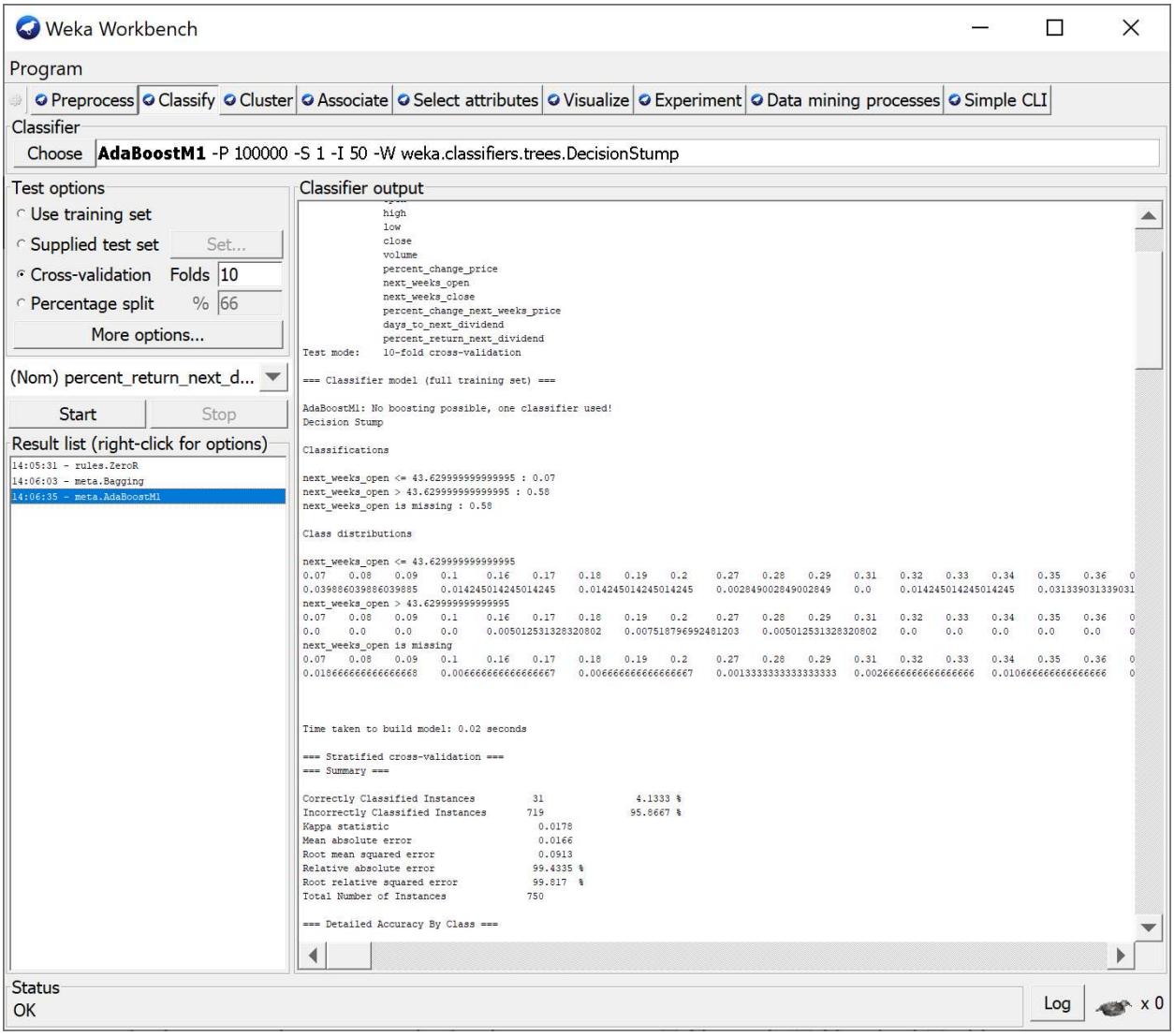
Bagging Output (Dow Jones Index)




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

Correctly Classified Instances         142               18.9333 %
Incorrectly Classified Instances       608               81.0667 %
Kappa statistic                          0.1759
Mean absolute error                      0.0149
Root mean squared error                  0.0879
Relative absolute error                 88.9805 %
Root relative squared error             96.1342 %
Total Number of Instances              750
```

AdaBoost Output (Dow Jones Index)




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

Correctly Classified Instances          31               4.1333 %
Incorrectly Classified Instances       719              95.8667 %
Kappa statistic                          0.0178
Mean absolute error                      0.0166
Root mean squared error                  0.0913
Relative absolute error                 99.4335 %
Root relative squared error             99.817  %
Total Number of Instances              750
```