

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)

The screenshot shows the Weka Workbench interface. The 'Program' tab is set to 'Classify'. The 'Classifier' dropdown is set to 'ZeroR'. Under 'Test options', 'Cross-validation' is selected with 'Folds' set to 10 and 'Percentage split' set to 66. The '(Nom) class' dropdown is set to 'Start'. The 'Result list' shows 'rules.ZeroR' selected. The 'Classifier output' pane displays the following information:

```
=== Run information ===
Scheme:      weka.classifiers.rules.ZeroR
Relation:    tic-tac-toe
Instances:   958
Attributes:  10
1=top-left-square
2=top-middle-square
3=top-right-square
4=middle-left-square
5=middle-middle-square
6=middle-right-square
7=bottom-left-square
8=bottom-middle-square
9=bottom-right-square
10=class
Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===
ZeroR predicts class value: TRUE

Time taken to build model: 0 seconds

=== 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

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.000    0.000    ?         0.000    ?         ?       0.496    0.345    FALSE
               1.000    1.000    0.653    1.000    0.790    ?       0.496    0.651    TRUE
Weighted Avg.   0.653    0.653    ?         0.653    ?         ?       0.496    0.545

=== Confusion Matrix ===
  a  b  <-- classified as
 0 332 | a = FALSE
 0 626 | b = TRUE
```

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

```
=== 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)

Weka Workbench

Program

Preprocess Classify Cluster Associate Select attributes Visualize Experiment Data mining processes Simple CLI

Classifier

Choose **Bagging** -P 100 -S 1 -num-slots 1 -I 50 -W weka.classifiers.trees.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) class

Start Stop

Result list (right-click for options)

14:54:59 - rules.ZeroR

14:57:15 - meta.Bagging

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 50 -W weka.classifiers.trees.RandomForest -- -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    tic-tac-toe
Instances:    958
Attributes:   10
1=Top-left-square
2=Top-middle-square
3=Top-right-square
4=Middle-left-square
5=Middle-middle-square
6=Middle-right-square
7=Bottom-left-square
8=Bottom-middle-square
9=Bottom-right-square
10=class
Test mode:    10-fold cross-validation

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

Bagging with 50 iterations and base learner

weka.classifiers.trees.RandomForest -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1

Time taken to build model: 3.51 seconds

=== 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

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0.910    0.008    0.984    0.910    0.945    0.919    0.997    0.995    FALSE
               0.992    0.090    0.954    0.992    0.973    0.919    0.997    0.999    TRUE
Weighted Avg.   0.963    0.062    0.964    0.963    0.963    0.919    0.997    0.997

=== Confusion Matrix ===
  a b  <-- classified as
 302 30 | a = FALSE
 5 621 | b = TRUE
```

Status OK

Log x 0

```
=== 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)

Weka Workbench

Program

Preprocess Classify Cluster Associate Select attributes Visualize Experiment Data mining processes Simple CLI

Classifier

Choose **AdaBoostM1** -P 100000 -S 1 -I 50 -W weka.classifiers.trees.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) class

Start Stop

Result list (right-click for options)

- 14:54:59 - rules.ZeroR
- 14:57:15 - meta.Bagging
- 15:00:46 - meta.AdaBoostM1

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.meta.AdaBoostM1 -P 100000 -S 1 -I 50 -W weka.classifiers.trees.RandomForest -- -P 100 -I 100 -num-slots 1 -K 0 -M 1.0 -V 0.001 -S 1
Relation:    tic-tac-toe
Instances:   958
Attributes:  10
1=Top-left-square
2=Top-middle-square
3=Top-right-square
4=Middle-left-square
5=Middle-middle-square
6=Middle-right-square
7=Bottom-left-square
8=Bottom-middle-square
9=Bottom-right-square
class

Test mode:   10-fold cross-validation

=== Classifier model (full training set) ===
AdaBoostM1: No boosting possible, one classifier used!
RandomForest

Bagging with 100 iterations and base learner
weka.classifiers.trees.RandomTree -K 0 -M 1.0 -V 0.001 -S -1155869325 -do-not-check-capabilities

Time taken to build model: 0.19 seconds

=== 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

=== Detailed Accuracy By Class ===
               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               ----  -
0.904    0.003    0.993    0.904    0.946    0.922    0.998    0.996    FALSE
0.997    0.096    0.951    0.997    0.973    0.922    0.998    0.999    TRUE
Weighted Avg.    0.965    0.064    0.966    0.965    0.964    0.922    0.998    0.998

=== Confusion Matrix ===
      a   b   <-- classified as
300  32 | a = FALSE
 2 624 | b = TRUE
```

Status

OK

Log x 0

```
=== 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)

Weka Workbench

Program: **Preprocess** **Classify** **Cluster** **Associate** **Select attributes** **Visualize** **Experiment** **Data mining processes** **Simple CLI**

Classifier: Choose **AdaBoostM1 -P 100000 -S 1 -I 50 -W weka.classifiers.trees.DecisionStump**

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

(Nom) percent_return_next_d...
Start Stop

Result list (right-click for options)
14:05:31 - rules.ZeroR
14:06:03 - meta.Bagging
14:06:35 - meta.AdaBoostM1

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.rules.ZeroR
Relation:    dow_jones_index.data
Instances:   750
Attributes:  14
    1>quarter
    stock
    date
    open
    high
    low
    close
    volume
    percent_change_price
    next_weeks_open
    next_weeks_close
    percent_change_next_weeks_price
    days_to_next_dividend
    percent_return_next_dividend
Test mode:   10-fold cross-validation

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

ZeroR predicts class value: 0.58

Time taken to build model: 0 seconds

=== 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

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0.000	0.000	?	0.000	?	?	0.413	0.016	0.07
	0.000	0.000	?	0.000	?	?	0.248	0.007	0.08
	0.000	0.000	?	0.000	?	?	0.248	0.007	0.09
	0.000	0.000	?	0.000	?	?	0.049	0.001	0.1
	0.000	0.000	?	0.000	?	?	0.099	0.003	0.16
	0.000	0.000	?	0.000	?	?	0.399	0.011	0.17
	0.000	0.000	?	0.000	?	?	0.418	0.015	0.18
	0.000	0.000	?	0.000	?	?	0.449	0.012	0.19
	0.000	0.000	?	0.000	?	?	0.149	0.004	0.2
	0.000	0.000	?	0.000	?	?	0.049	0.001	0.27

Status: OK Log x 0

```
=== 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)

Weka Workbench

Program: ☒ Preprocess ☒ **Classify** ☐ Cluster ☐ Associate ☐ Select attributes ☐ Visualize ☐ Experiment ☐ Data mining processes ☐ Simple CLI

Classifier: Choose **AdaBoostM1 -P 100000 -S 1 -I 50 -W weka.classifiers.trees.DecisionStump**

Test options:
☐ Use training set
☐ Supplied test set
☒ Cross-validation Folds **10**
☐ Percentage split % **66**

(Nom) percent_return_next_d...

Result list (right-click for options):
14:05:31 - rules.ZeroR
14:06:03 - meta.Bagging
14:06:35 - meta.AdaBoostM1

Classifier output

```
=== Run information ===
Scheme:      weka.classifiers.meta.Bagging -P 100 -S 1 -num-slots 1 -I 50 -W weka.classifiers.trees.REPTree -- -M 2 -V 0.001 -N 3 -S 1 -L -1 -I
Relation:    dow_jones_index.data
Instances:   750
Attributes:  14
    1>quarter
    stock
    date
    open
    high
    low
    close
    volume
    percent_change_price
    next_weeks_open
    next_weeks_close
    percent_change_next_weeks_price
    days_to_next_dividend
    percent_return_next_dividend
Test mode:   10-fold cross-validation

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

Bagging with 50 iterations and base learner
weka.classifiers.trees.REPTree -M 2 -V 0.001 -N 3 -S 1 -L -1 -I 0.0

Time taken to build model: 0.74 seconds

=== 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

=== Detailed Accuracy By Class ===

```

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
1.000	0.015	0.560	1.000	0.718	0.743	0.990	0.521	0.07	0.07
0.000	0.000	?	0.000	?	?	0.993	0.170	0.08	0.08
0.000	0.000	?	0.000	?	?	0.982	0.163	0.09	0.09
0.000	0.000	?	0.000	?	?	0.487	0.001	0.1	0.1
0.000	0.000	?	0.000	?	?	0.972	0.063	0.16	0.16
0.000	0.000	?	0.000	?	?	0.968	0.150	0.17	0.17
0.846	0.019	0.440	0.846	0.579	0.601	0.983	0.366	0.18	0.18
0.000	0.000	?	0.000	?	?	0.971	0.175	0.19	0.19

Status: OK x 0

```
=== 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)

Weka Workbench

Program

Preprocess Classify Cluster Associate Select attributes Visualize Experiment Data mining processes Simple CLI

Classifier

Choose **AdaBoostM1** -P 100000 -S 1 -I 50 -W weka.classifiers.trees.DecisionStump

Test options

☐ Use training set

☐ Supplied test set Set...

☒ Cross-validation Folds 10

☐ Percentage split % 66

More options...

(Nom) percent_return_next_d...

Start Stop

Result list (right-click for options)

- 14:05:31 - rules.ZeroR
- 14:06:03 - meta.Bagging
- 14:06:35 - meta.AdaBoostM1

Classifier output

high
low
close
volume
percent_change_price
next_weeks_open
next_weeks_close
percent_change_next_weeks_price
days_to_next_dividend
percent_return_next_dividend

Test mode: 10-fold cross-validation

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

AdaBoostM1: No boosting possible, one classifier used!
Decision Stump

Classifications

next_weeks_open <= 43.629999999999995 : 0.07
next_weeks_open > 43.629999999999995 : 0.58
next_weeks_open is missing : 0.58

Class distributions

next_weeks_open <= 43.629999999999995

0.07	0.08	0.09	0.1	0.16	0.17	0.18	0.19	0.2	0.27	0.28	0.29	0.31	0.32	0.33	0.34	0.35	0.36	0
0.039886039886039885	0.014245014245014245	0.014245014245014245	0.002849002849002849	0.0	0.014245014245014245	0.031339031339031												

next_weeks_open > 43.629999999999995

0.07	0.08	0.09	0.1	0.16	0.17	0.18	0.19	0.2	0.27	0.28	0.29	0.31	0.32	0.33	0.34	0.35	0.36	0
0.0	0.0	0.0	0.0	0.005012531328320802	0.007518796992481203	0.005012531328320802	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0

next_weeks_open is missing

0.07	0.08	0.09	0.1	0.16	0.17	0.18	0.19	0.2	0.27	0.28	0.29	0.31	0.32	0.33	0.34	0.35	0.36	0
0.018666666666666668	0.006666666666666667	0.006666666666666667	0.006666666666666667	0.0013333333333333333	0.0026666666666666666	0.010666666666666666												0

Time taken to build model: 0.02 seconds

=== 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	

=== Detailed Accuracy By Class ===

Status
OK

Log x 0

```
=== 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
```