**Weka Classification Results**

I used the ZeroR classifer on 24 attributes of the train-a data table with 10 folds and the patient disposition as my target attribute for my first classification run. The ZeroR method resulted in 44% correct classifications. Next I used the J48 classifier on the same 24 attributes of the train-a data table with 5 folds. My first run failed, so I increased the ram for Weka from 4 GB to 6 GB. The J48 then did have a complete run on the 24 attributes. I was unable, or unwilling to wait for a visualization of the data tree with this J48 run. I did another J48 run on just 6 attributes and was able to quickly visualize the resulting tree, figure 1. The tree though was poorly constructed and unreadable. I first J48 run with 24 attributes resulted in an increase of correct classifications to 49% with another increase to 50 % on the 6 attribute run.

**ZeroR Results**

=== Run information ===

Scheme: weka.classifiers.rules.ZeroR

Relation: train-a-weka.filters.unsupervised.attribute.NumericToNominal-R9-weka.filters.unsupervised.attribute.NumericToNominal-R1-3-weka.filters.unsupervised.attribute.Remove-R17-30,32-45,48-71-weka.filters.unsupervised.attribute.Remove-R1-3-weka.filters.unsupervised.attribute.NumericToNominal-R10-weka.filters.unsupervised.attribute.NumericToNominal-R11-23

Instances: 99999

Attributes: 24

Interval

Patient-Sex

Age

Patient-Race

Patient-Ethnicity

Patient-Disposition

Length of Stay

Admit-Type

Admit-Source

Hospital-ID

Region-ID

Principal-Dx-Code

Admit-Dx-Code

Principal-PR-Code

Cause-E-Code

Place-E-Code

Complication-Minor

Complication-Sever

Trauma-Minor

Trauma-Severe

Trauma-Severity

Nosocomial Inf

Severity

Cost Weight

Test mode: 10-fold cross-validation

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

ZeroR predicts class value: 1

Time taken to build model: 0.25 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 44001 44.0014 %

Incorrectly Classified Instances 55998 55.9986 %

Kappa statistic 0

Mean absolute error 0.0838

Root mean squared error 0.2046

Relative absolute error 100 %

Root relative squared error 100 %

**ZeroR Results**

Total Number of Instances 99999

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

1 1 0.44 1 0.611 0.5 1

0 0 0 0 0 0.5 2

0 0 0 0 0 0.5 3

0 0 0 0 0 0.498 4

0 0 0 0 0 0.499 5

0 0 0 0 0 0.5 6

0 0 0 0 0 0.5 7

0 0 0 0 0 0.5 8

0 0 0 0 0 0.5 20

0 0 0 0 0 0.454 43

0 0 0 0 0 0.496 50

0 0 0 0 0 0.495 51

0 0 0 0 0 0.498 61

0 0 0 0 0 0.5 62

0 0 0 0 0 0.491 63

0 0 0 0 0 0.5 65

0 0 0 0 0 0.25 66

Weighted Avg. 0.44 0.44 0.194 0.44 0.269 0.5

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p q <-- classified as

44001 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | a = 1

4203 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | b = 2

20963 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | c = 3

575 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | d = 4

1328 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | e = 5

21721 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | f = 6

1501 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | g = 7

220 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | h = 8

2203 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | i = 20

26 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | j = 43

218 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | k = 50

234 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | l = 51

211 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | m = 61

2408 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | n = 62

92 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | o = 63

90 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | p = 65

5 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | q = 66

**Results from first J48 run**

=== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: train-a-weka.filters.unsupervised.attribute.NumericToNominal-R9-weka.filters.unsupervised.attribute.NumericToNominal-R1-3-weka.filters.unsupervised.attribute.Remove-R17-30,32-45,48-71-weka.filters.unsupervised.attribute.Remove-R1-3-weka.filters.unsupervised.attribute.NumericToNominal-R10-weka.filters.unsupervised.attribute.NumericToNominal-R11-23

Instances: 99999

Attributes: 24

Interval

Patient-Sex

Age

Patient-Race

Patient-Ethnicity

Patient-Disposition

Length of Stay

Admit-Type

Admit-Source

Hospital-ID

Region-ID

Principal-Dx-Code

Admit-Dx-Code

Principal-PR-Code

Cause-E-Code

Place-E-Code

Complication-Minor

Complication-Sever

Trauma-Minor

Trauma-Severe

Trauma-Severity

Nosocomial Inf

Severity

Cost Weight

Test mode: 5-fold cross-validation

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

J48 pruned tree

Number of Leaves : 307112

Size of the tree : 309598

Time taken to build model: 184.75 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 49146 49.1465 %

Incorrectly Classified Instances 50853 50.8535 %

Kappa statistic 0.203

Mean absolute error 0.0756

Root mean squared error 0.2067

Relative absolute error 90.225 %

Root relative squared error 100.9909 %

Total Number of Instances 99999

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.829 0.549 0.542 0.829 0.656 0.664 1

0.022 0.006 0.133 0.022 0.038 0.608 2

0.415 0.147 0.429 0.415 0.422 0.663 3

0.016 0.001 0.102 0.016 0.027 0.553 4

0.011 0.002 0.075 0.011 0.02 0.551 5

0.172 0.089 0.349 0.172 0.23 0.579 6

0.003 0 0.106 0.003 0.006 0.658 7

0.014 0 0.079 0.014 0.023 0.604 8

0.009 0.002 0.081 0.009 0.016 0.572 20

0 0 0 0 0 0.552 43

0 0 0 0 0 0.499 50

0 0 0 0 0 0.546 51

0.019 0 0.148 0.019 0.034 0.549 61

0.047 0.003 0.259 0.047 0.079 0.633 62

0.011 0 0.083 0.011 0.019 0.457 63

0 0 0 0 0 0.537 65

0 0 0 0 0 0.519 66

Weighted Avg. 0.491 0.292 0.422 0.491 0.431 0.637

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p q <-- classified as

36464 207 4023 31 52 3007 18 15 59 0 5 8 11 97 2 2 0 | a = 1

3306 92 452 4 6 317 2 1 8 0 0 1 1 13 0 0 0 | b = 2

9163 153 8694 22 59 2640 10 11 87 0 1 10 5 104 4 0 0 | c = 3

384 4 117 9 1 56 0 0 0 0 0 1 0 3 0 0 0 | d = 4

755 12 385 3 15 141 3 1 2 0 1 0 0 10 0 0 0 | e = 5

13205 159 4384 12 53 3727 8 2 62 0 3 7 6 87 4 2 0 | f = 6

1341 4 89 0 2 58 5 0 0 0 2 0 0 0 0 0 0 | g = 7

124 1 61 0 1 28 0 3 2 0 0 0 0 0 0 0 0 | h = 8

1083 26 813 4 5 244 0 3 20 0 0 1 0 4 0 0 0 | i = 20

20 0 5 1 0 0 0 0 0 0 0 0 0 0 0 0 0 | j = 43

116 4 66 0 0 29 1 0 1 0 0 0 0 1 0 0 0 | k = 50

87 3 90 1 1 48 0 2 1 0 0 0 0 1 0 0 0 | l = 51

112 2 69 0 0 24 0 0 0 0 0 0 4 0 0 0 0 | m = 61

947 24 981 1 4 329 0 0 5 1 2 1 0 112 1 0 0 | n = 62

40 0 34 0 0 17 0 0 0 0 0 0 0 0 1 0 0 | o = 63

65 1 14 0 2 8 0 0 0 0 0 0 0 0 0 0 0 | p = 65

3 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | q = 66

**Results from second J48 run**

== Run information ===

Scheme: weka.classifiers.trees.J48 -C 0.25 -M 2

Relation: train-a-weka.filters.unsupervised.attribute.NumericToNominal-R9-weka.filters.unsupervised.attribute.NumericToNominal-R1-3-weka.filters.unsupervised.attribute.Remove-R17-30,32-45,48-71-weka.filters.unsupervised.attribute.Remove-R1-3-weka.filters.unsupervised.attribute.NumericToNominal-R10-weka.filters.unsupervised.attribute.NumericToNominal-R11-23-weka.filters.unsupervised.attribute.Remove-R1,8-24

Instances: 99999

Attributes: 6

Patient-Sex

Age

Patient-Race

Patient-Ethnicity

Patient-Disposition

Length of Stay

Test mode: 3-fold cross-validation

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

J48 pruned tree

Number of Leaves : 1174

Size of the tree : 2347

Time taken to build model: 45.89 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances 50143 50.1435 %

Incorrectly Classified Instances 49856 49.8565 %

Kappa statistic 0.2166

Mean absolute error 0.0759

Root mean squared error 0.1977

Relative absolute error 90.5484 %

Root relative squared error 96.6032 %

Total Number of Instances 99999

=== Detailed Accuracy By Class ===

TP Rate FP Rate Precision Recall F-Measure ROC Area Class

0.843 0.532 0.554 0.843 0.669 0.715 1

0.003 0.001 0.09 0.003 0.006 0.611 2

0.516 0.188 0.421 0.516 0.464 0.731 3

0 0 0 0 0 0.592 4

0.002 0.001 0.032 0.002 0.003 0.525 5

0.101 0.061 0.315 0.101 0.153 0.591 6

0 0 0 0 0 0.744 7

0 0 0 0 0 0.533 8

0.001 0.001 0.051 0.001 0.003 0.658 20

0 0 0 0 0 0.576 43

0 0 0 0 0 0.621 50

0 0 0 0 0 0.608 51

0.005 0 0.143 0.005 0.009 0.614 61

0.002 0.001 0.052 0.002 0.005 0.625 62

0 0 0 0 0 0.516 63

0 0 0 0 0 0.582 65

0 0 0 0 0 0.354 66

Weighted Avg. 0.501 0.287 0.407 0.501 0.425 0.679

=== Confusion Matrix ===

a b c d e f g h i j k l m n o p q <-- classified as

37099 33 4749 1 13 2052 1 0 17 0 0 1 3 32 0 0 0 | a = 1

3417 12 557 0 3 209 0 0 1 0 0 0 0 4 0 0 0 | b = 2

8200 32 10817 3 18 1837 0 0 23 0 0 0 1 32 0 0 0 | c = 3

350 0 194 0 0 30 0 0 0 0 0 0 0 1 0 0 0 | d = 4

783 2 417 0 2 118 0 0 4 0 0 0 0 2 0 0 0 | e = 5

13087 37 6328 2 17 2203 1 1 8 0 0 0 2 35 0 0 0 | f = 6

1416 1 52 0 0 30 0 0 2 0 0 0 0 0 0 0 0 | g = 7

130 0 66 0 1 23 0 0 0 0 0 0 0 0 0 0 0 | h = 8

954 6 1077 0 6 155 0 0 3 0 0 0 0 2 0 0 0 | i = 20

20 0 4 0 0 2 0 0 0 0 0 0 0 0 0 0 0 | j = 43

84 1 112 0 0 21 0 0 0 0 0 0 0 0 0 0 0 | k = 50

75 1 139 0 1 18 0 0 0 0 0 0 0 0 0 0 0 | l = 51

106 0 87 0 0 16 0 0 0 0 0 0 1 1 0 0 0 | m = 61

1068 8 1044 0 2 279 0 0 1 0 0 0 0 6 0 0 0 | n = 62

40 0 46 0 0 5 0 0 0 0 0 0 0 1 0 0 0 | o = 63

75 0 12 0 0 3 0 0 0 0 0 0 0 0 0 0 0 | p = 65

3 0 2 0 0 0 0 0 0 0 0 0 0 0 0 0 0 | q = 66

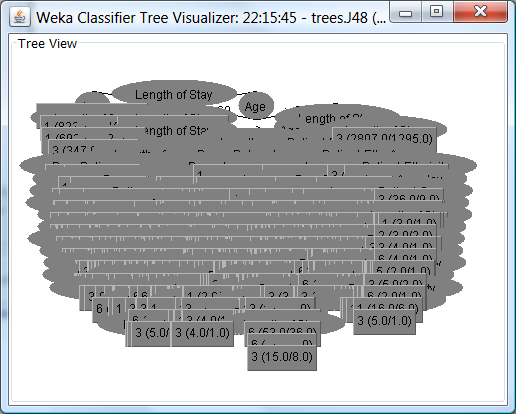


Figure 1. Visualization of classification tree for the second J48 run, which is worthless.