|  |
| --- |
| April 17,2019 - Team 03 (Mackenzie O., Madison R., Madeline S.) |

|  |
| --- |
| Executive Summary |

# Analysis of Competitive EBAY Auctions

This paper analyzes specific attributes of auctions in order to classify them as competitive or noncompetitive.  A competitive auction is defined as an auction with at least two bids placed on the item auctioned. The dataset includes attributes that describe the item (Category), rate the seller (sellerRating), outline the auction terms (currency, Duration, endDay, OpenPrice), and provide the price at which the auction closed (ClosePrice).

CART Decision Tree (Appendix B)

The prediction algorithm generated results with an 87.35% accuracy for training and 85.13% accuracy for testing sets. They have 92% and 91.5% precision respectively. In Appendix B, you can see a pictorial representation of the decision tree model. The model predicted that Open Price, followed by Close Price, are by far the most important predictors in determining whether an auction is competitive or not. Looking at the rules, it alternated between OpenPrice and ClosePrice in decreasing ranges to narrow down to a pure tree until it had to use other attributes.

C 4/5.0 Decision Tree (Appendix C)

The prediction algorithm generated results with an 88.82% accuracy for training and 88.56% accuracy for testing sets. In Appendix C you can see a pictorial representation of the decision tree model. The model predicted that Open Price, followed by Close Price, are by far the most important predictors in determining whether an auction is competitive or not. Looking at the rules, it alternated between OpenPrice and ClosePrice in ranges to narrow down to a pure tree until it had to use other attributes.

Bayes Net (Appendix D)

The model had performance metrics of 70.59% accuracy, 69.35% recall, 27.90% FP rate, 75.16% precision.  The accuracy, recall, and precision are decent, but the FP rate is relatively high.  An ROC curve was produced to further evaluate the performance of the Bayesian Network.  The area under the curve is a measure of the predictive accuracy of the model. In this case, the area under the curve appears to be about 0.75, which is relatively accurate.

Neural Net (Appendix E)

Neural networks predict a continuous or categorical target based on one or more predictors by finding unknown and possibly complex patterns in the data. We predicted that the training dataset will be competitive 76.84% of the time and the testing set 77.9%. The models have 71.2 and 75.6% precision in their prediction respectively. We also know that close price is the most important variable in making a product competitive.

Apriori Modelling (Appendix F)

The network automatically generated five association rules, all of which with the same end goal: to see how competitive the products are. We can determine if each rule is a good or poor rule based on calculating improvement/lift. According to the generated rules, the two best predictors are endDay= “Monday” and currency=”US” based on their support, confidence and lift.

## Conclusions

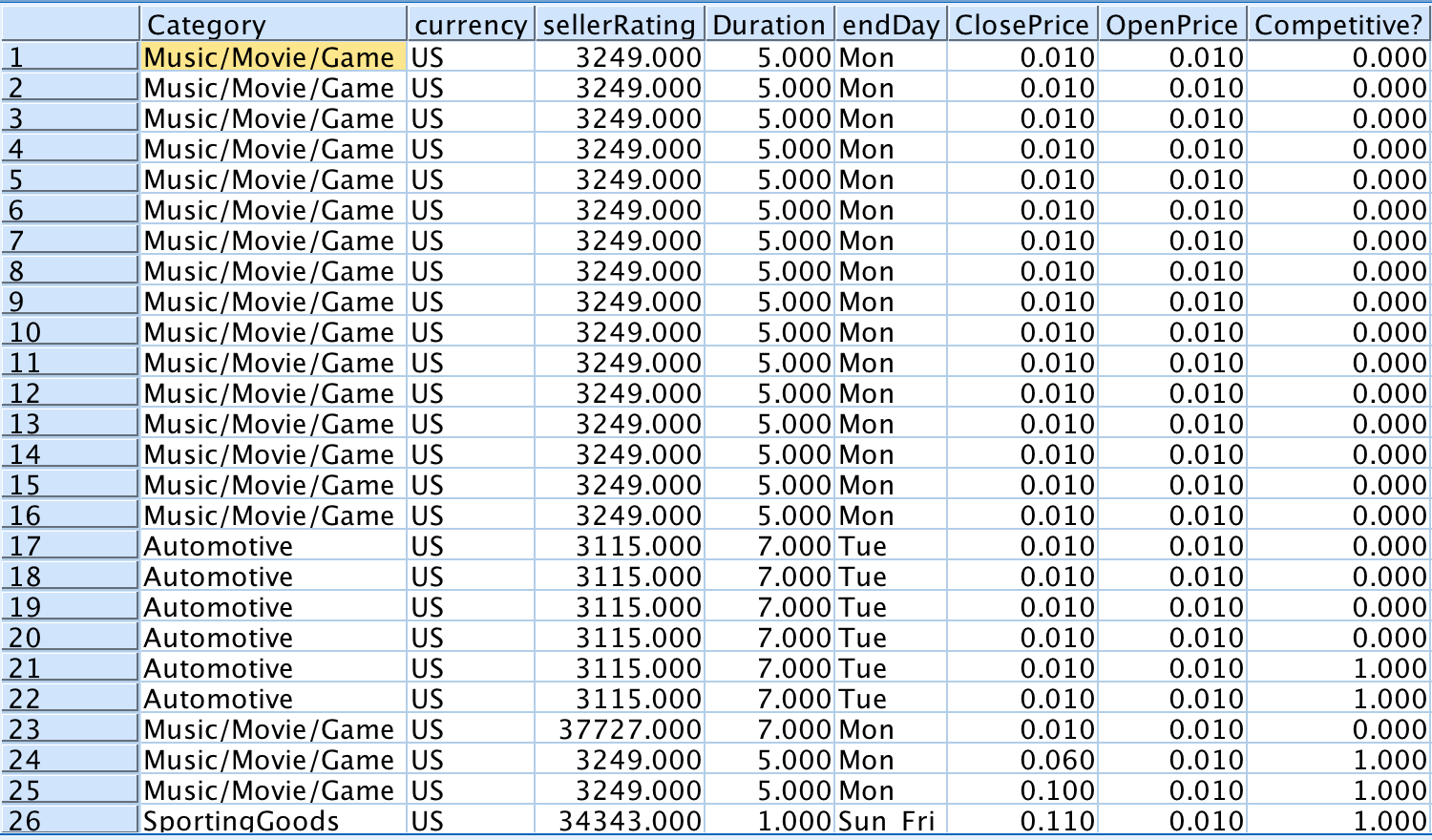
Looking at all of the models, it seems that Open and Close Prices are important factors in whether an auction is competitive or not. The apriori results suggest that there is a correlation in competitive auctions closing on Mondays and in US dollars. The C4/5.0 decision Tree was the most accurate at 88.82% and 88.56% accurate for training and testing. Overall, both decision tree models did better than the Bayesian or Neural Network models in terms of recall, accuracy, specificity, and precision.

.

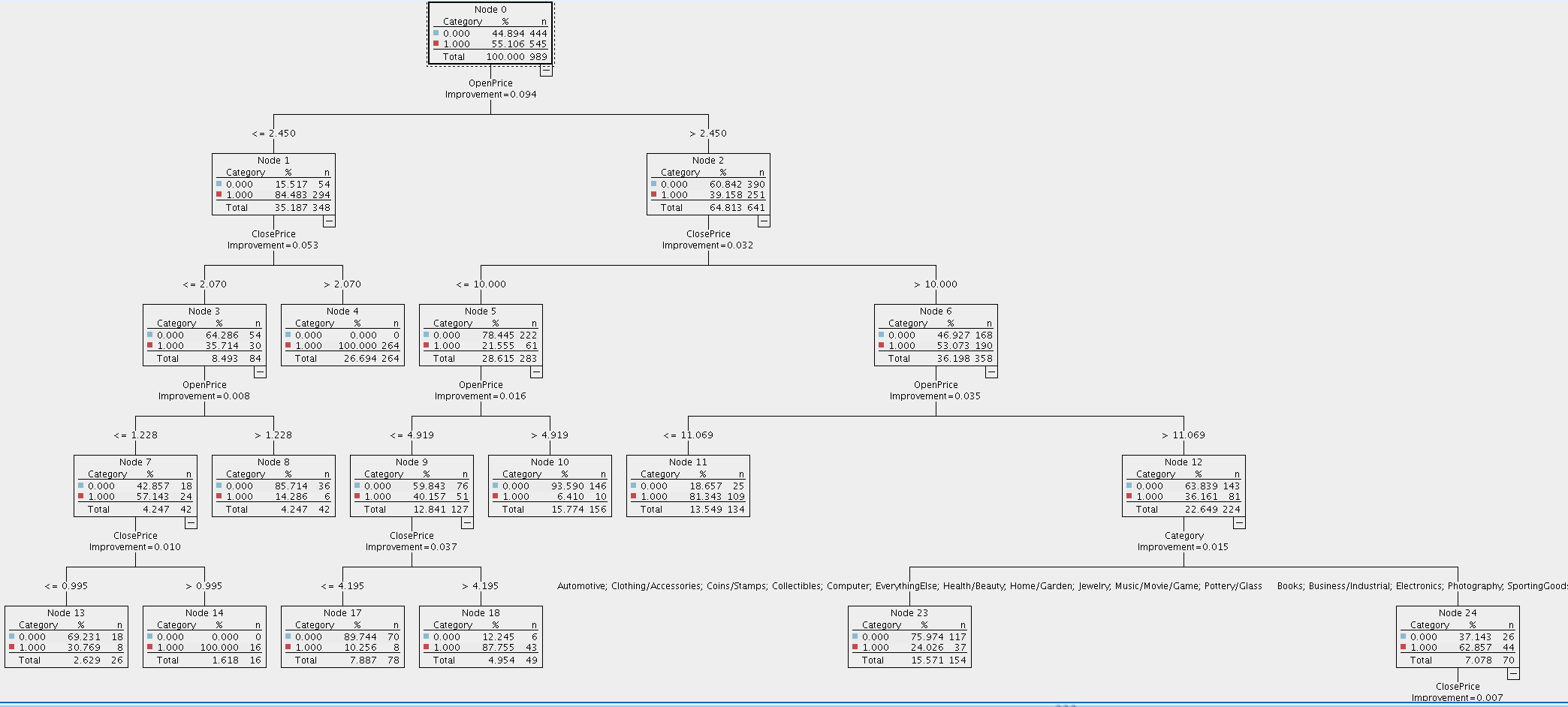
## Appendix A:

## Data Collection

Data collected for each auction includes the category of the auction, the currency, the seller’s Rating, the duration of the sale, the day of the week if ended, the price the auction closed at, the price it opened at, and whether it was competitive or not.

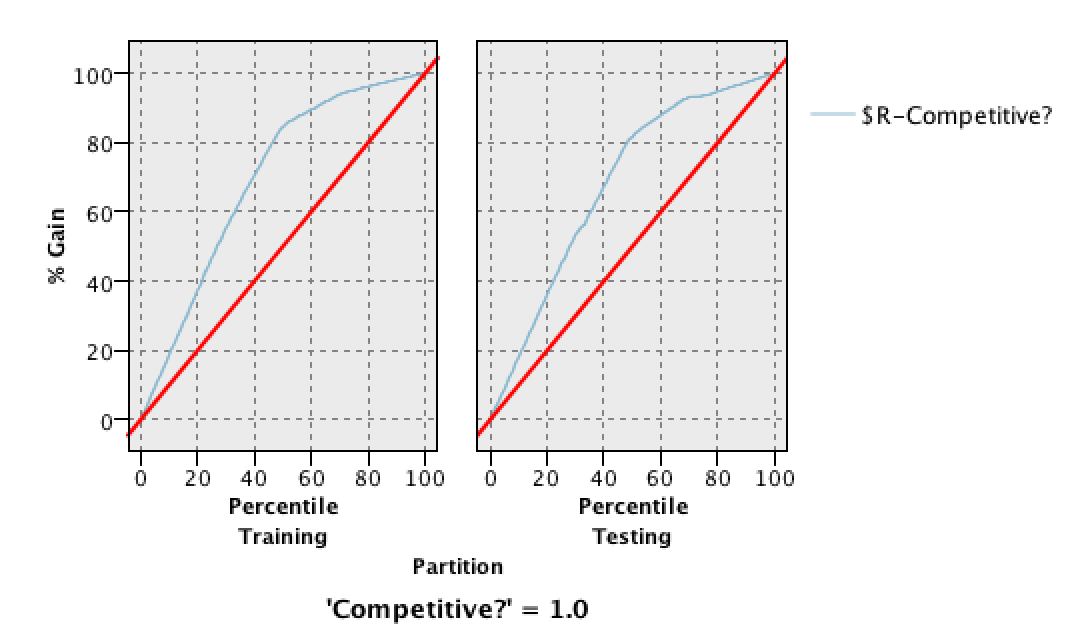


## Appendix B:

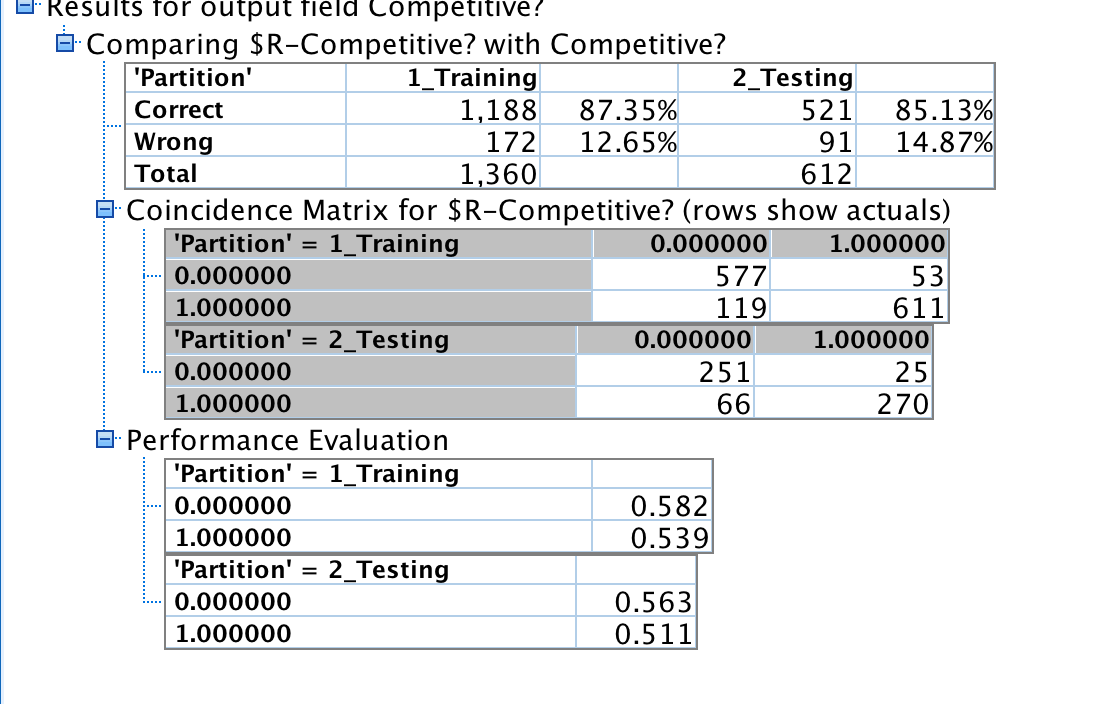
CART Decision Tree

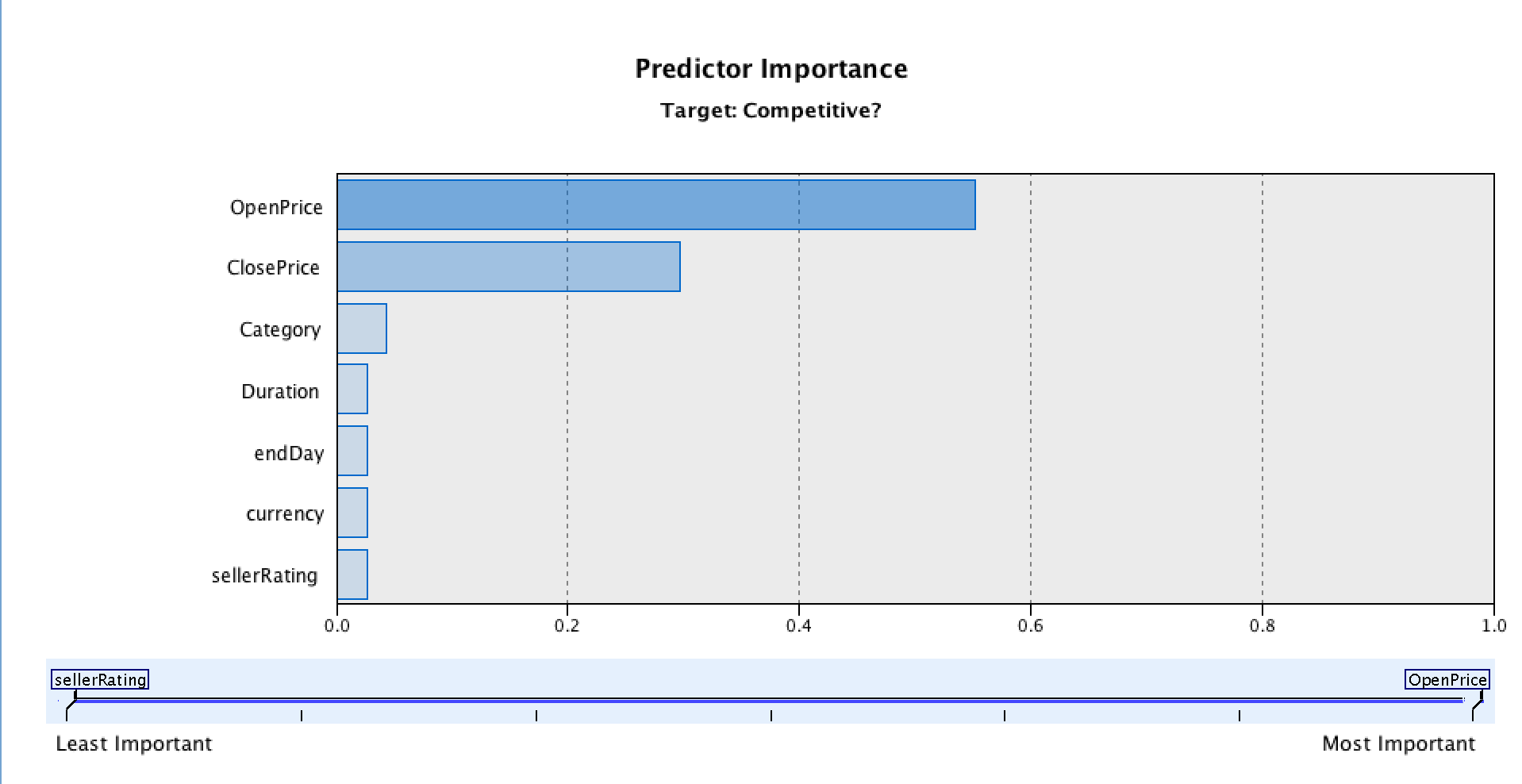
Examples of Some Rules: The root is Open Price. It seems that the model alternates between ranges of OpenPrice and Close Price attributes to gain better results until it is too narrow and must resort to other attributes to make decisions for a pure tree.

|  |  |  |  |
| --- | --- | --- | --- |
| **Antecedent** | **Consequence** | **Support** | **Confidence** |
| If (Open Price)>2.450 | Then Competitive=0 | 390/989 | 390/641 |
| If (Open Price)<=2.450 | Then Competitive=1 | 294/989 | 294/348 |
| If (Open Price<=2.450) ^(Close Price<=2.070) | Then Competitive=0 | 84/989 | 54/84 |
| If (Open Price<=2.450) ^(Close Price>2.070) | Then Competitive=1 | 264/989 | 100% |
| If (Open Price<=2.450) ^(Close Price<=2.070)^(OpenPrice<=1.228) | Then Competitive=1 | 42/989 | 24/42 |
| If (Open Price<=2.450) ^(Close Price<=2.070)^(OpenPrice>1.228) | Then Competitive=0 | 42/989 | 36/42 |
| If (Open Price<=2.450) ^(Close Price<=2.070)^(OpenPrice<=1.228)^(ClosePrice<=0.995) | Then Competitive=0 | 26/989 | 18/26 |
| If (Open Price<=2.450) ^(Close Price<=2.070)^(OpenPrice<=1.228)^(ClosePrice>0.995) | Then Competitive=1 | 16/989 | 100% |

Gain Chart: At about 50% of the sample, we have around 85% gain for training and testing.

|  |  |  |
| --- | --- | --- |
|  | Training: | Testing: |
| Accuracy: TP+TN)/(TP+TN+FP+FN) | (611+577)/(577+53+119+611)=87.3% | (270+251)/(251+25+66+270)=85.13% |
| Specificity: FP/(TN+FP) | 53/(577+53)=8.3% | (25/(25+251)=9% |
| Recall: TP/(TP+FN) | 611/(611+119)=83.69% | 270/(270+66)=80.35% |
| Precision: TP/(TP+FP) | 611/(611+53)=92% | 270/(270+25)=91.5% |



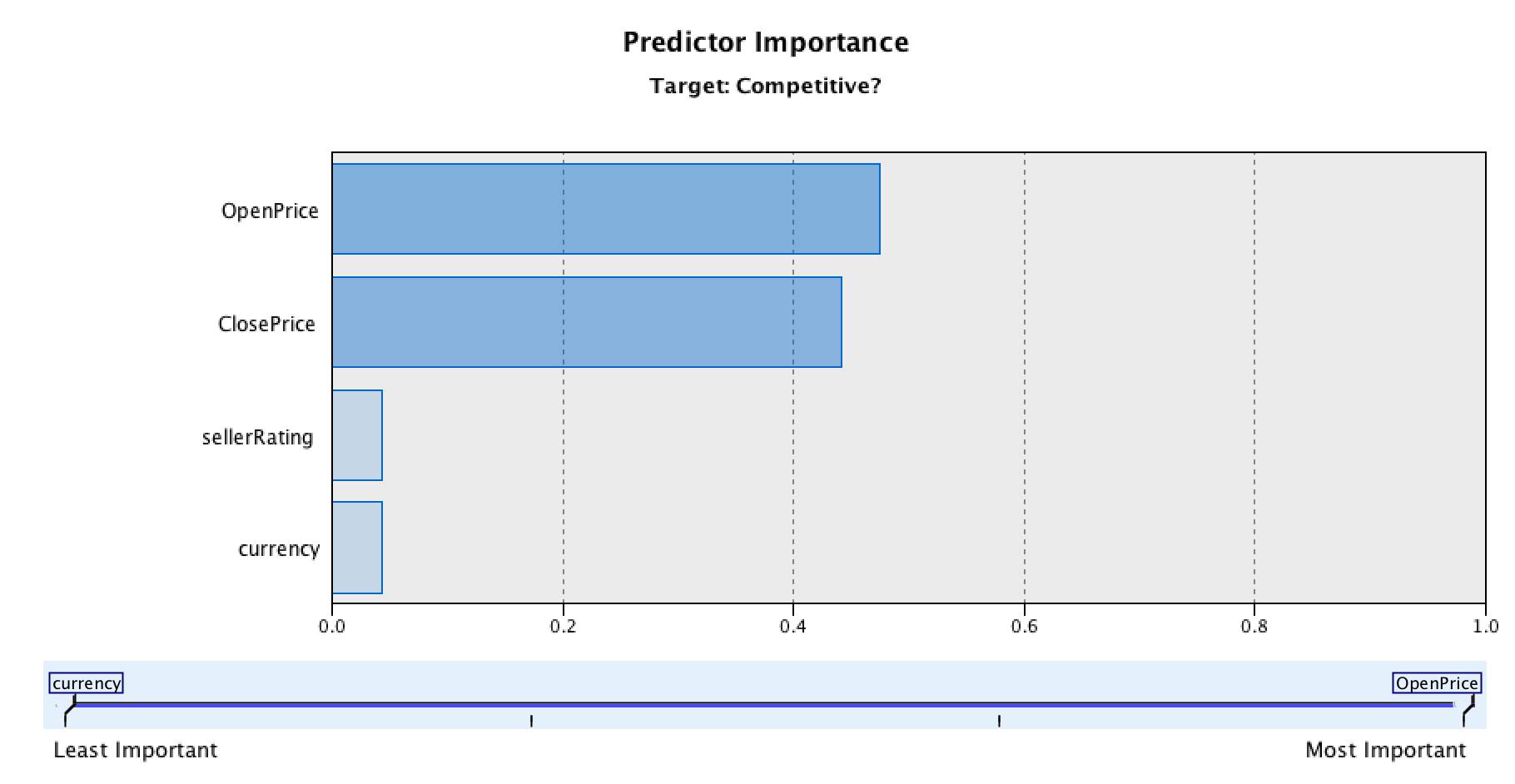
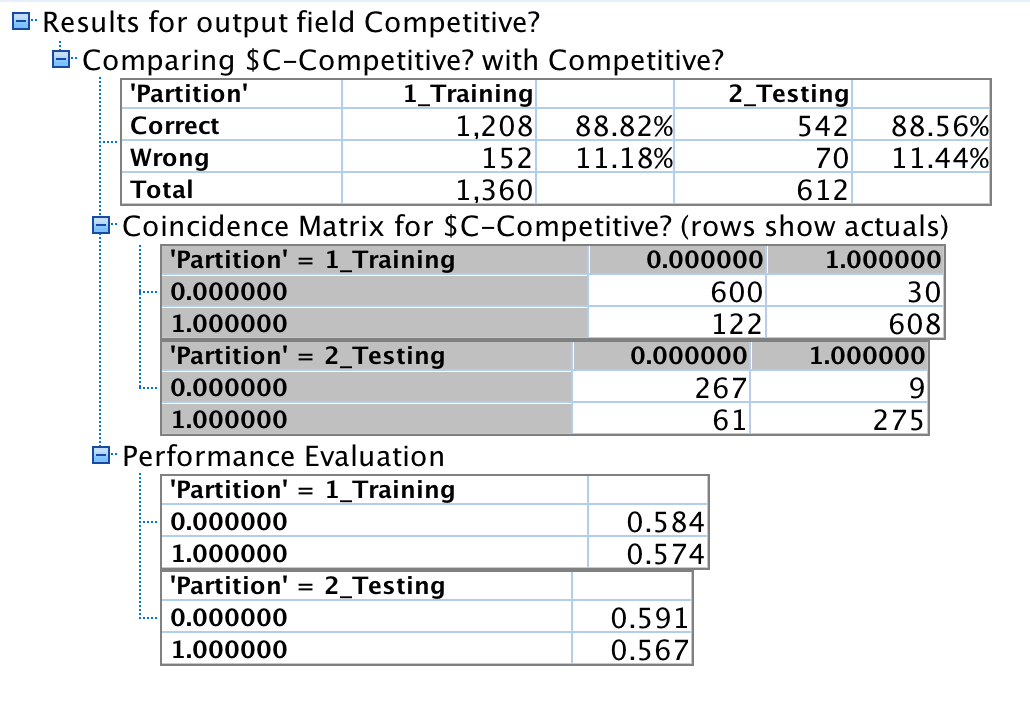
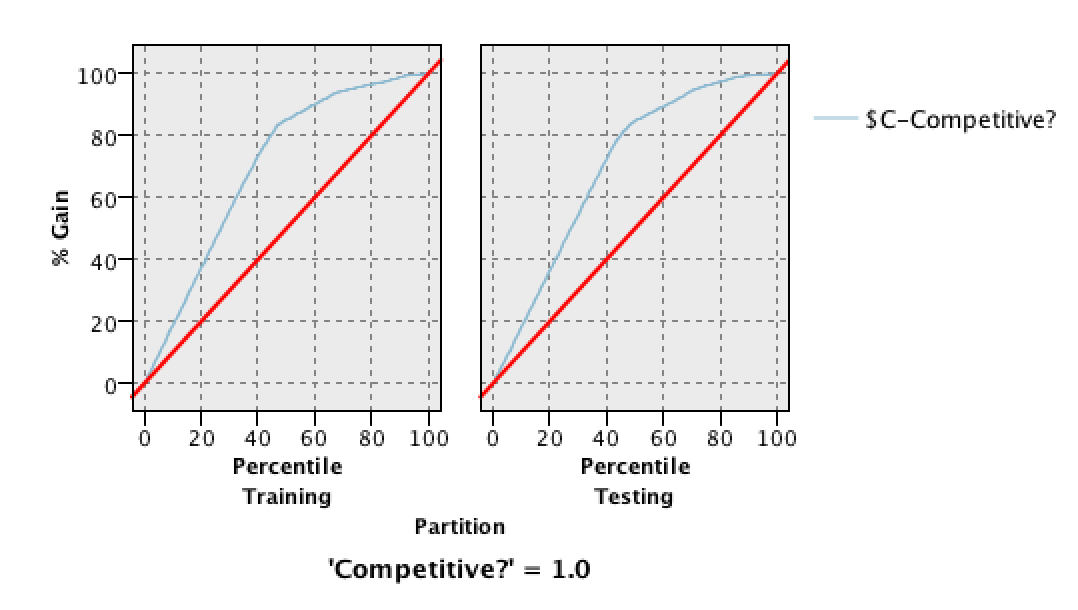


## Appendix C:

## C 4/5.0 Decision Tree

Samples of Decision Tree Rules: Much like the CART tree, this tree uses OpenPrice as root and alternates between Open and Close price for most of its decisions.

|  |  |  |  |
| --- | --- | --- | --- |
| **Antecedent** | **Consequence** | **Support** | **Confidence** |
| If (Open Price>1.230 | Then Competitive=0 | 1541/1972 | 865/1541 |
| If (Open Price<=1.230 | Then Competitive=1 | 390/1972 | 390/431 |
| If (Open Price)>1.230^(Close Price<=1.160) | Then Competitive=0 | 59/1972 | 41/59 |
| If (Open Price>1.230) ^(Close Price>1.160) | Then Competitive=1 | 372/1972 | 100% |
| If (Open Price)>1.230^(Close Price<=1.160)^ (Close Price<=0.035) | Then Competitive=0 | 23/1972 | 21/23 |
| If (Open Price)>1.230^(Close Price<=1.160)^(ClosePrice>0.035) | Then Competitive=0 | 36/1972 | 20/36 |
| If (Open Price)>1.230^(Close Price<=1.160)^(ClosePrice>0.035)^(OpenPrice<=0.617) | Then Competitive=1 | 11/1972 | 100% |
| If (Open Price)>1.230^(Close Price<=1.160)^(ClosePrice>0.035)^(OpenPrice>0.617) | Then Competitive=0 | 25/1972 | 20/25 |

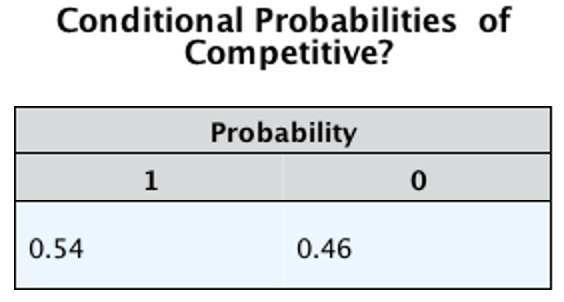
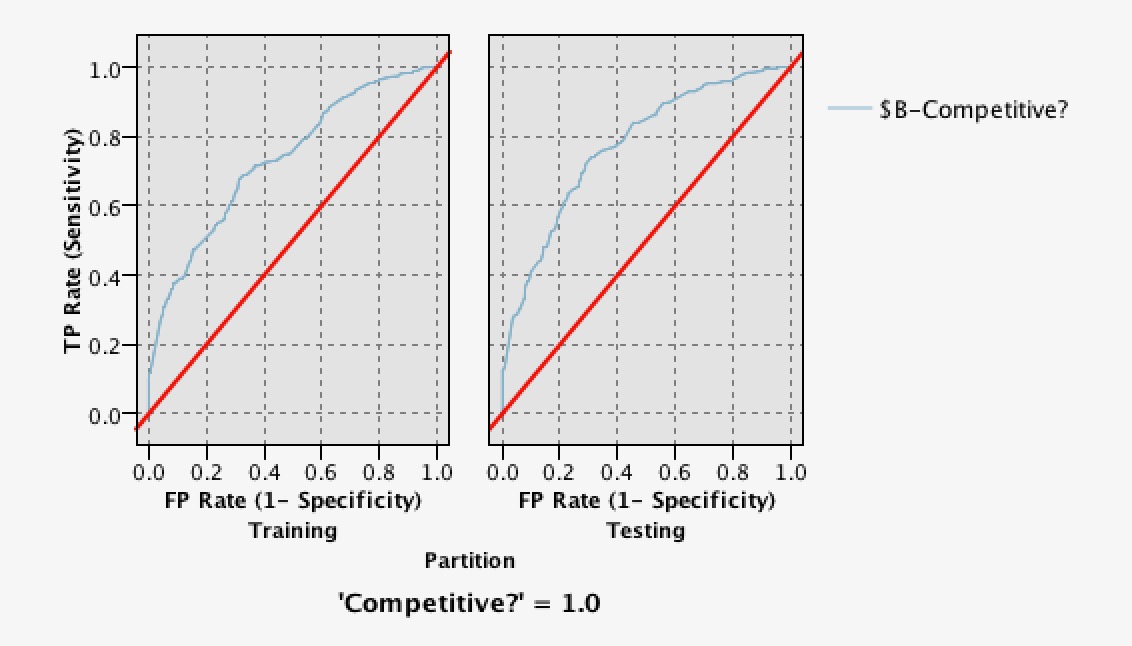


|  |  |  |
| --- | --- | --- |
|  | Training: | Testing: |
| Accuracy: (TP+TN)/(TP+TN+FP+FN) | (608+600)/(608+600+122+30)=88.82% | (267+275)/(267+9+61+275)=88.56% |
| Specificity: FP/(TN+FP) | 30/(30+600)=4.76% | 9/(267+9)=3.26% |
| Recall: TP/(TP+FN) | 608/(608+122)=83.28% | 275/(275+61)=81.84% |
| Precision: TP/(TP+FP) | 608/(608+30)=95.29% | 275/(275+9)=96.8% |

## Appendix D:

## https://lh3.googleusercontent.com/WSZA11w-G_zP1atvnxSgYIChswrRznOrx5ruxifDzVr97KKJ8xonpFwV8e9RdP7aFT-K5OhTI3fT6JDVgiO1raEHbEzWqZqTcsCeYY12Ff3_6eEKwtuJrTQE1Pd4SoG1Jp5I4CcZBayes Net

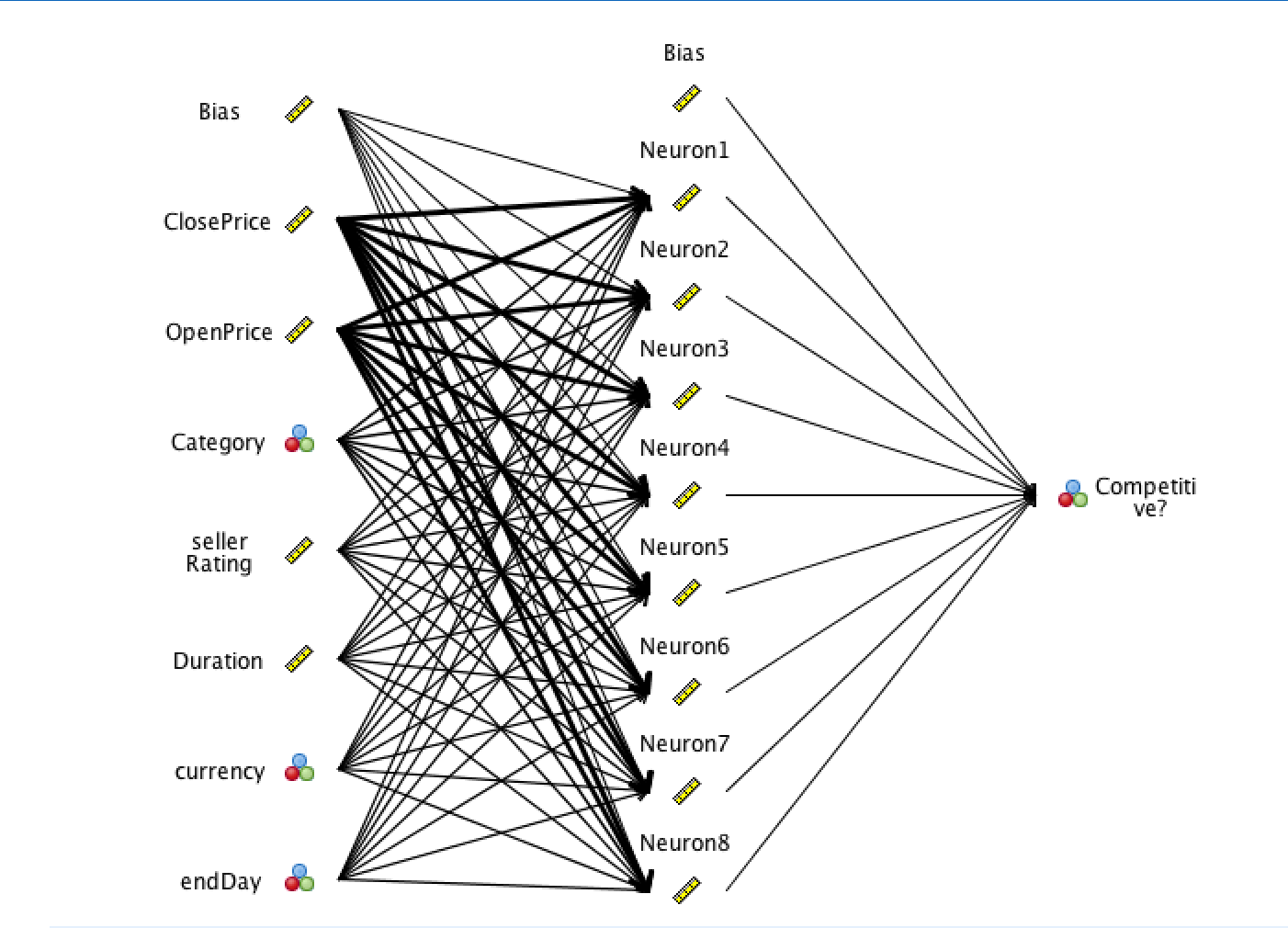
ROC Curve:



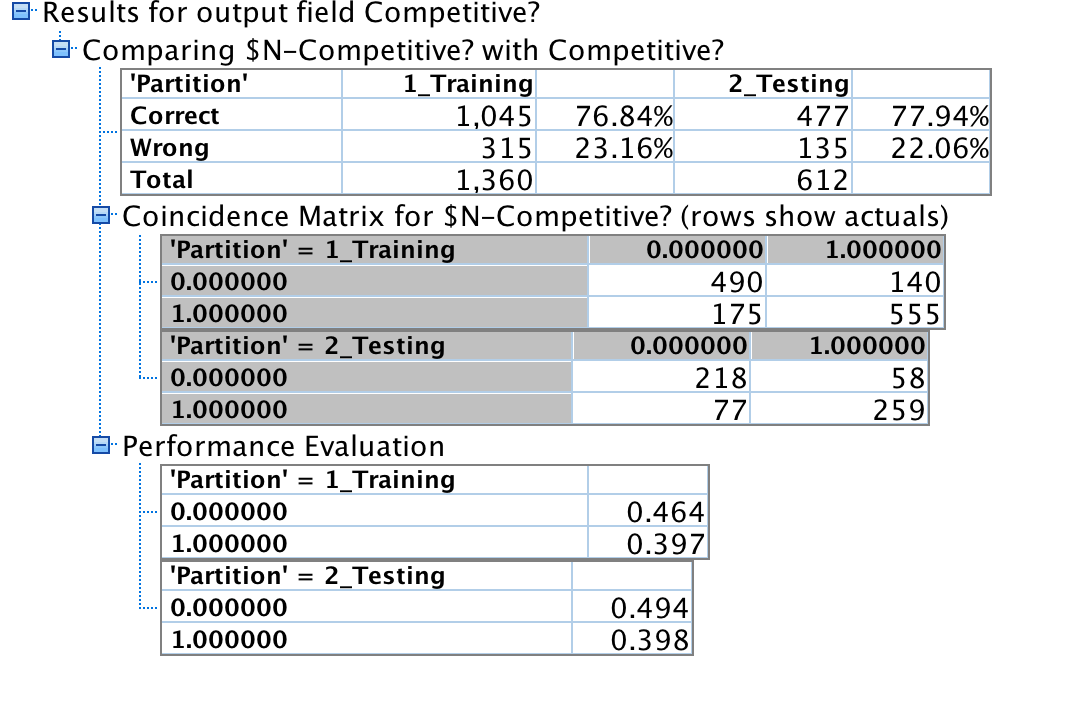
|  |  |  |
| --- | --- | --- |
|  | Training: | Testing: |
| Accuracy: TP+TN)/(TP+TN+FP+FN) | (431+493)/(431+493+237+199)=67.9% | (199+233)/(199+77+103+233) = 70.59% |
| Specificity: FP/(TN+FP) | 199/(431+199)=31.5% | 77/(199+77)=27.8% |
| Recall: TP/(TP+FN) | 493/(493+237)=67.5% | 233/(233+103) = 69.35% |
| Precision: TP/(TP+FP) | (493/(493+199)=71.2% | 233/(233+77) = 75.16% |

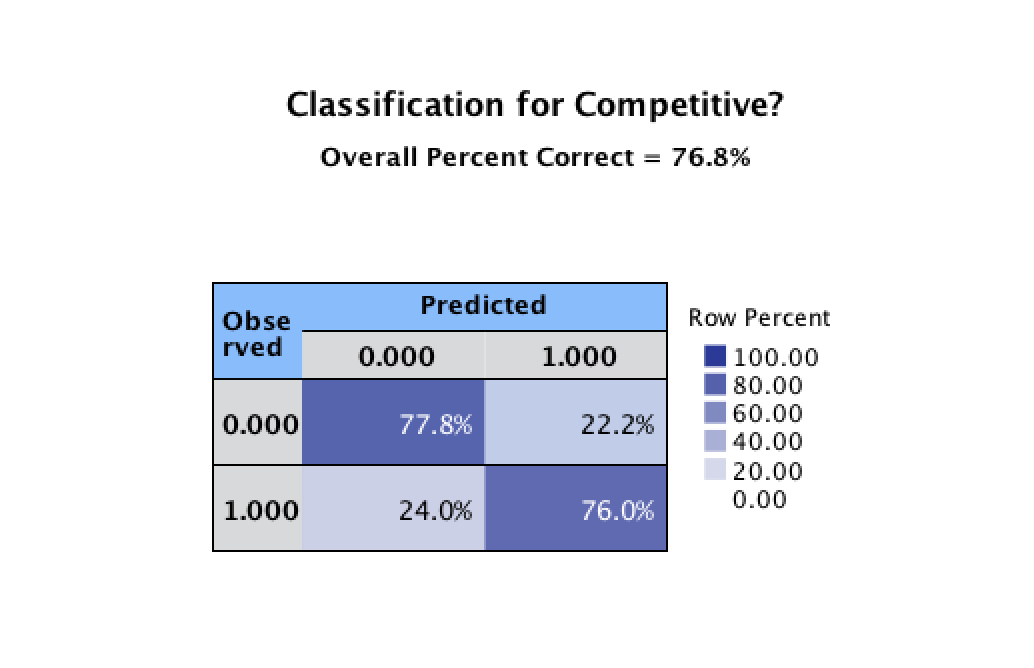
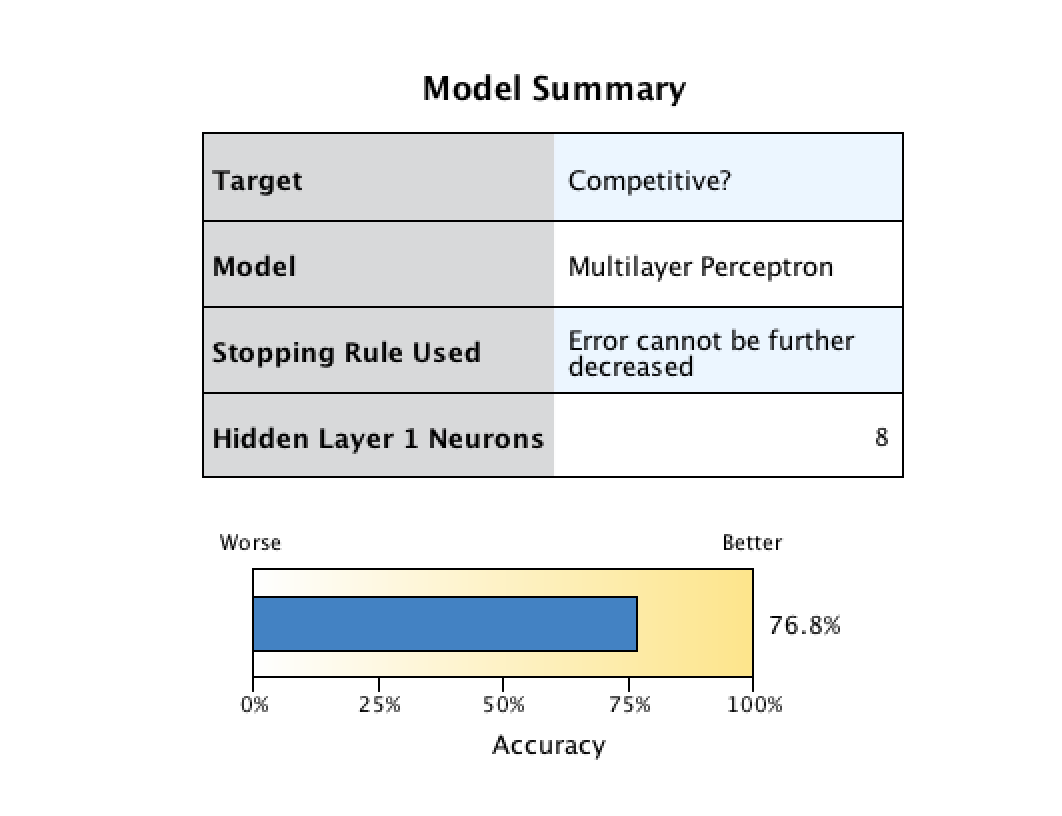
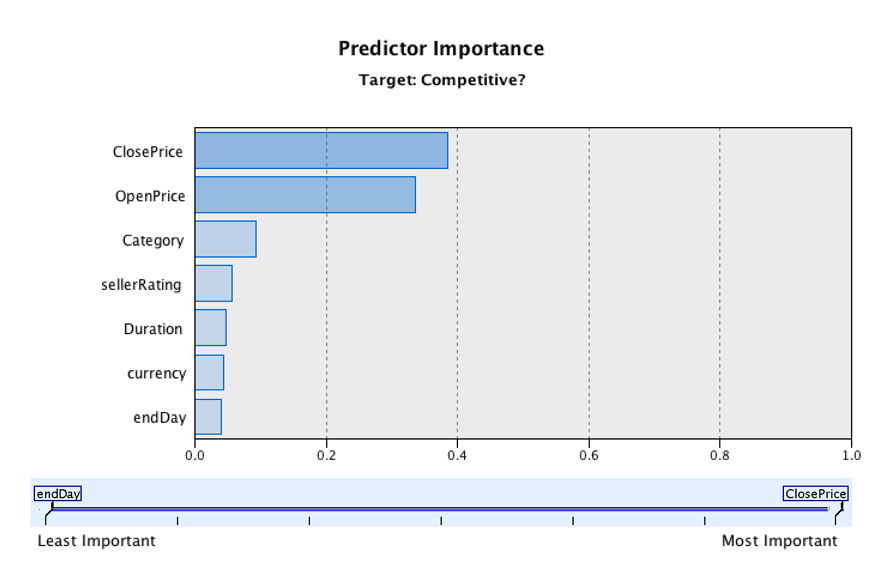
## https://lh4.googleusercontent.com/NcvTUyUap5RPg3ARGYOYHDzuZc9OkFh_zXZnx6lCxNYLFdFp5TErWiREjBsli4L-CyZQX0xHlclxdHigJLOyo08cz4BKfYPEikyzo5D4MkUN43Whl8bu8Ld8yv_WkFinkzebe_Gk

## Appendix E:

Neural Net

|  |  |  |
| --- | --- | --- |
|  | Training: | Testing: |
| Accuracy: TP+TN)/(TP+TN+FP+FN) | (555+490)/(555+490+175+140)=76.8% | (259+218)/(259+218+58+77)=77.9% |
| Specificity: FP/(TN+FP) | 140/(490+140)=22.22% | 58/(218+58)=21% |
| Recall: TP/(TP+FN) | 555/(555+175)=76% | 259/(259+77)=77.1% |
| Precision: TP/(TP+FP) | 555/(555+140)=79.8% | 259/(259+58)=81.7% |





## Appendix F:

Apriori Net

Rule #1 has the most confidence and lift, which make it the best predictor, while rule #2 has lower confidence but more support and a similar lift. Both rules use “endDay= Monday” as a predictor, and the top 3 rules are very similar. The two best predictors are thus “endDay= Monday” and currency=”US”

