Prediction of Customers' Preference of Brand

Prepared for: Danielle Sherman

Executive Summary

Two sets of data were provided by Blackwell Electronics. Both sets of data included 7 attributes about each consumer: salary, age, level of education, make of primary car, region, amount of credit available, and consumer's preferred brand. The first 6 attributes were used as the independent variable to predict the 7th (consumers' preferred brand). The first set of data was used to train models. The first model trained was a Random Forest model. Default number of trees for this model was 500. The error rate is plotted in Figure 1. From this data the number of trees was reduced to 300. Next "mtry" values were plotted against error rate (Figure 2), and the model "mtry" value was tuned to 3. The final results from of this model are shown in Figure 3. The next model trained was C5.0. This model was tuned, and the model parameters are shown in Figure 4. Finally this trained C5.0 model was used on the second dataset to make predictions about consumer brand preference. This is shown in Figure 5.

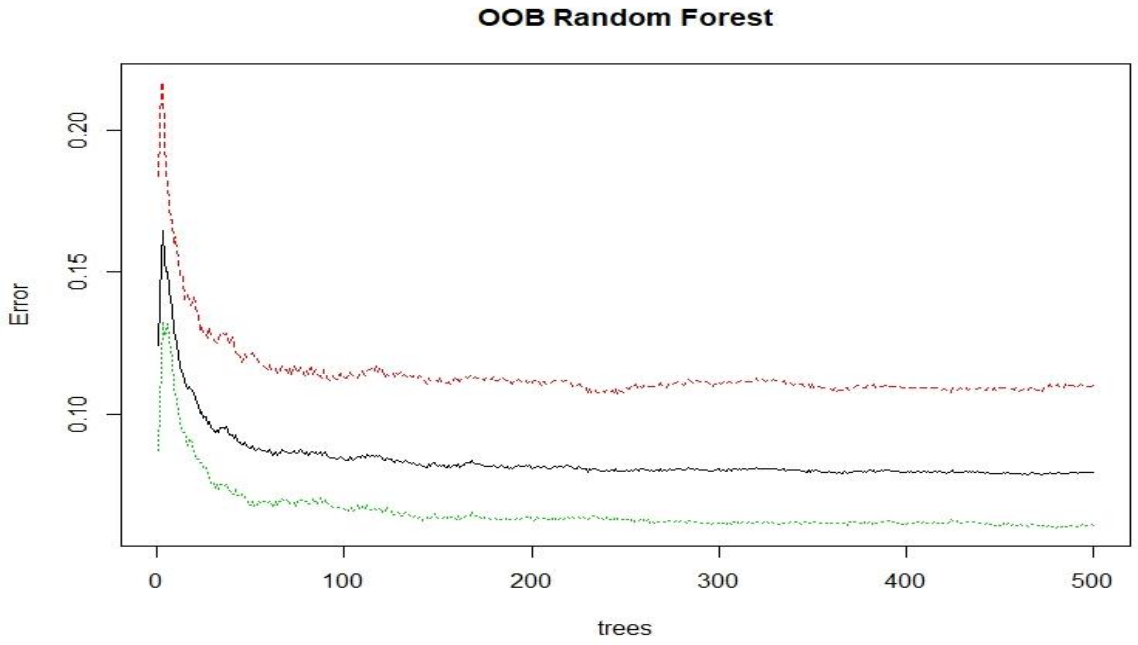

Figure 1

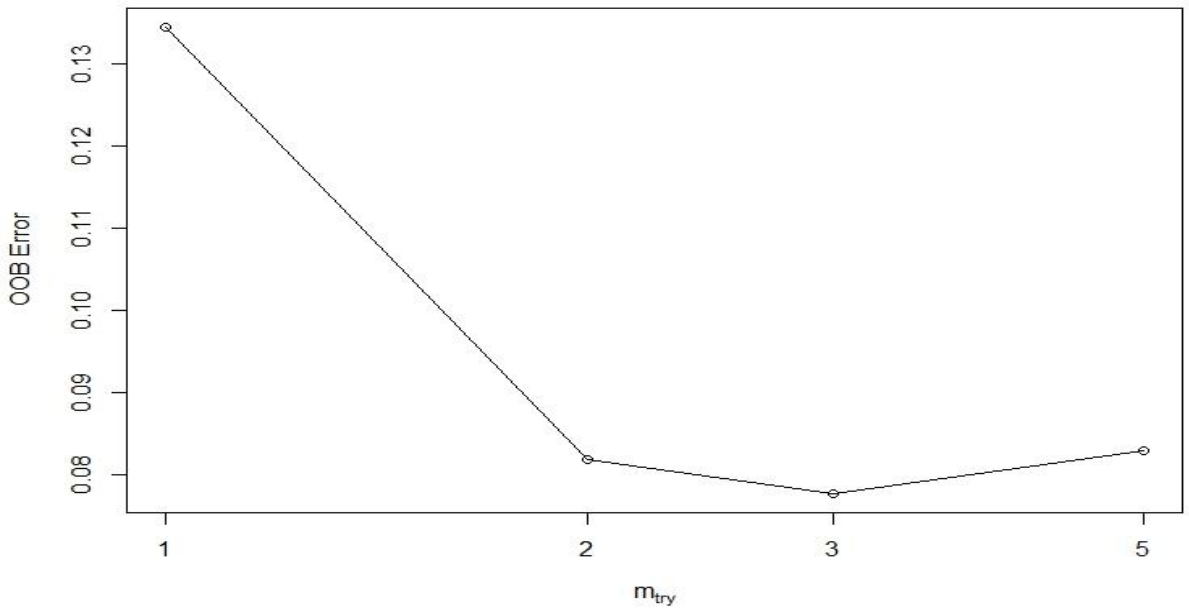

Figure 2

```
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1843    4
         1 3094   59

               Accuracy : 0.3804
                 95% CI : (0.3669, 0.394)
    No Information Rate : 0.9874
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0123

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.37330
            Specificity : 0.93651
         Pos Pred Value : 0.99783
         Neg Pred Value : 0.01871
             Prevalence : 0.98740
         Detection Rate : 0.36860
   Detection Prevalence : 0.36940
      Balanced Accuracy : 0.65491

       'Positive' Class : 0
```

Figure 3

```
> confusionMatrix(p1, incomplete$brand)
Confusion Matrix and Statistics

          Reference
Prediction    0    1
         0 1884    5
         1 3053   58

               Accuracy : 0.3884
                 95% CI : (0.3749, 0.4021)
    No Information Rate : 0.9874
    P-Value [Acc > NIR] : 1

                  Kappa : 0.0121

 Mcnemar's Test P-Value : <2e-16

            Sensitivity : 0.38161
            Specificity : 0.92063
         Pos Pred Value : 0.99735
         Neg Pred Value : 0.01864
             Prevalence : 0.98740
         Detection Rate : 0.37680
   Detection Prevalence : 0.37780
      Balanced Accuracy : 0.65112

       'Positive' Class : 0
```

Figure 4

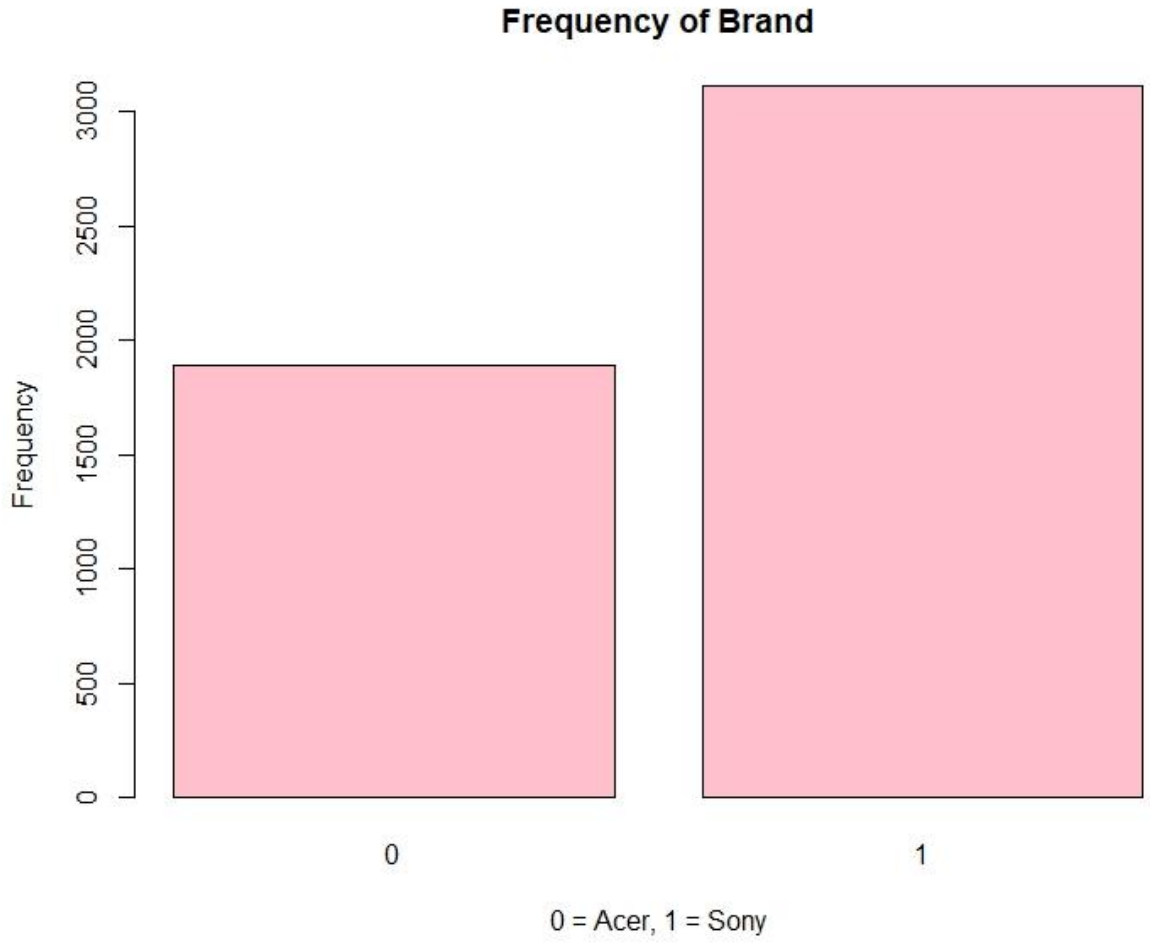
Frequency of Brand
Frequency
0
500
1000
1500
2000
2500
3000
0
1
0 = Acer, 1 = Sony

Figure 5