Customer Brand Preferences Report
Blackwell Electronics

By Samantha Goodman

Background

For this task I was assigned the job of predicting brand preferences given a survey dataset and an incomplete dataset. The customer survey contained demographic data (age, zipcode, salary, education) as well as credit available and car make. Customers then indicated which brand of computer they used, Acer or Sony. The second dataset had the same demographic and background information, but the brand preference question was incompletely captured and thus unusable. I used machine learning to predict the final question of brand preference for the second incomplete dataset and thus can present a complete picture of which brand the customers prefer. Through this process I also uncovered which feature (age, salary, etc) has the most influence on brand preference, which could be valuable to the marketing and sales team.

Classifier Test Results

I tested three different classifiers, C50, Stochastic Gradient Boosting, and Random Forest. The C5.0 classifier was found to have an Accuracy of 0.92 with a Kappa of 0.84

model	winnow	trials	Accuracy	Kappa
rules	FALSE	1	0.8291921	0.6591105
rules	FALSE	10	0.9199940	0.8291753
rules	FALSE	20	0.9179717	0.8254053
rules	TRUE	1	0.8271706	0.6550959
rules	TRUE	10	0.9237643	0.8373373
rules	TRUE	20	0.9248417	0.8400902
tree	FALSE	1	0.8290575	0.6588619
tree	FALSE	10	0.9218762	0.8336765
tree	FALSE	20	0.9220117	0.8345244
tree	TRUE	1	0.8267668	0.6541188
tree	TRUE	10	0.9213388	0.8324620
tree	TRUE	20	0.9226858	0.8357900

The Stochiastic Gradient Booting was very close, with an Accuracy of 0.92 and a Kappa of 0.83

Resampling results across tuning parameters:

interaction.depth	n.trees	Accuracy	Kappa
1	50	0.7315500	0.4358430
1	100	0.7296625	0.4304315
1	150	0.7299344	0.4308008
2	50	0.8092559	0.6028575
2	100	0.8856438	0.7617999
2	150	0.9127121	0.8161628
3	50	0.8768823	0.7462695
3	100	0.8988307	0.7889326
3	150	0.9218691	0.8353337

Random Forest classifier had an Accuracy of 0.89 with a Kappa of 0.76.

```
Resampling results across tuning parameters:

  mtry  Accuracy   Kappa
  1     0.6217673  0.000000000
  2     0.6220366  0.000884002
  3     0.7401585  0.381831172
  4     0.8511556  0.675831552
  5     0.8867191  0.758707641
```

Therefore, the best classifier to continue with is C5.0.

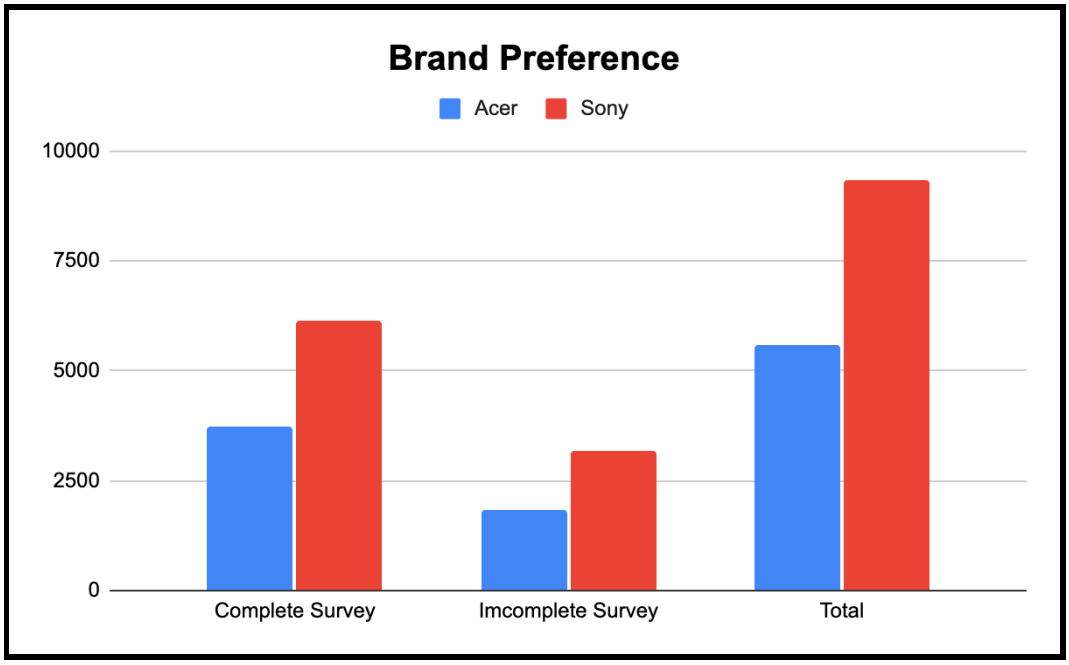
Important Variables

Using the variable importance function, I was able to determine that the most influential factor (given the limited data that we have) was salary. The second most influential factor on a customer's preference was age. The top 3 influential variables are Salary>Age>Available Credit.

Brand Preference

Using the uncorrupted data as a guide, I found the brand preferred by 63% of customers was Sony. The overall predicted brand preference by the customers is shown in the table and charts below.

Customer Preference		
	Acer	**Sony**
Complete Survey	3744	6154
Incomplete Survey *(predicted)*	1828	3172
Total Customers	5693	9205
Percentage of Customers	37%	63%


Brand Preference
Acer
Sony
10000
7500
5000
2500
0
Complete Survey
Imcomplete Survey
Total