**Customer Brand Preference Report**

As Blackwell continues to expand its product offerings, understanding customer brand loyalty and selection can be key to determining which products to offer. Therefore, the purpose of this report is to predict potential Blackwell customer brand preference of computer (Acer or Sony) based upon historical survey responses. From using these survey responses, the goal is better understand which brands a customer may purchase and with the opportunities for further data analysis to investigate other potential factors that can assist in determining customer future buying behaviors.

**Methodology**

The data used for this analysis consisted of a dataset of fully completed survey responses and another data set that did not have full response. In the completed dataset, the number of responses was totaled at 10,000 responses with the following categories: salary, age, education level, model of primary car, zip code, credit limit, and brand of computer. While the incomplete survey had only 5,000 responses with the same categories yet the brand of computer attribute did not have any pertinent customer information. Therefore, the process of this analysis consisted of developing and training a model from the completed survey data to predict customer computer preference in the incomplete survey data. The first step in the both data sets was inspecting the data for missing values and transforming specific attributes into data that can be analyzed. This including transitions the attributes type of cars, zip code, and computer brand to factors and changing the attribute education level to ordinal. Then, I proceeded to develop testing and training datasets to determine the best predictive model utilizing the algorithms, K Nearest Neighbor (KNN) and Random Forrest. In building this model, the complete survey data were split into 75% of data (training) and 25$ (testing). Each of the algorithms were applied to the Complete Survey data and the training and testing data sets in order to build an effective model to more accurately predict customer computer brand

**Classifiers used and model selected**

In regards to KNN, the training and testing data were applied to the complete survey data and model fit numbers which included accuracy and kappa to determine which model with the respective number of k values was implemented. Based upon the performance measurements of accuracy and kappa, a KNN model with a 11 values of k led to the highest accuracy (64%) and kappa (.64) measurements. A detailed log of KNN classifier adjustments can be found in Appendix A. The following chart outlines the various parameters of the KNN model that were analyzed to determine performance measurements:

|  |  |  |
| --- | --- | --- |
| KNN Classifiers | | |
| *K Values* | ***Accuracy*** | ***Kappa*** |
| 5 | 0.5685386 | 0.5685386 |
| 7 | 0.5977072 | 0.10290172 |
| 9 | 0.6237562 | 0.15038599 |
| 11 | 0.6380621 | 0.17790761 |
| 13 | 0.6364219 | 0.17196881 |
| 15 | 0.6288896 | 0.15204806 |
| 17 | 0.6175435 | 0.12346740 |
| 19 | 0.6058250 | 0.09176631 |
| 21 | 0.5987324 | 0.06909904 |
| 23 | 0.5950670 | 0.05474685 |
| 25 | 0.5920266 | 0.04192163 |
| 27 | 0.5904675 | 0.03162471 |
| 29 | 0.5908142 | 0.02518717 |
| 31 | 0.5921738 | 0.02086866 |
| 33 | 0.5952005 | 0.02035208 |
| 35 | 0.6001604 | 0.02517731 |
| 37 | 0.6050394 | 0.03028113 |
| 39 | 0.6116250 | 0.04061211 |
| 41 | 0.6156641 | 0.04604348 |
| 43 | 0.6202900 | 0.05330415 |

In addition to applying KNN to the training data sets, the random forest algorithm was applied to the Complete Survey data to determine the most appropriate model to predict customer computer brand preference. Through tuning the model by adjusting the number of trees applied in the algorithm, the best model fit for the data included the using 750 trees which resulted in a () error rate. The following chart logs the different parameters that were tuned within the random forest models and their respective error rates.

|  |  |  |
| --- | --- | --- |
| Random Forrest Performance Measurements | | |
| Number of Trees | ***Error Rate*** |
| 25 | 10.25% |
| 50 | 9.24% |
| 75 | 8.83% |
| 100 | 8.64% |
| 150 | 8.55% |
| 175 | 8.65 |
| 200 | 8.44% |
| 500 | 8.01% |
| 600 | 8.33% |
| 700 | 8.4% |

Based the optimized models for the KNN and Random Forrest, the model selected to predict Blackwell customer computer brand preference was the random forest model with 500 trees with an error rate of 8.01%.

**Brand Predictions**

With the previous optimized model, I applied this model to the incomplete survey data set focusing specifically on the customer computer brand attribute. Based upon the application of the optimized predictive model, more customers could potentially prefer Sony (3147) as compared to Acer (1853). Therefore, it is predicted that more Blackwell customers will likely purchase a Sony computer over an Acer brand. The attached chart provides an overview of the predicted customer computer preferences based upon survey responses. A comprehensive list of the responses is provided within an addition excel spreadsheet titled “Predicted Customer Brands”. Further analysis can be conducted looking at the how other attributes may affect customer buying patterns. This report provides insights into customer survey data that could contribute to the future computer brand purchasing and provide marketing specifics as to tailored marketing approaches based upon such factors as salary, educational level, age, type of care, credit limit, and zip code.

.