**Band Preference Prediction**

Report for Blackwell Electronics

**Introduction**

The purpose of this report is to predict the brand preference from customers for Blackwell Electronic to decide which manufacture should be pursued with a deeper strategic relationship. The prediction is made on two data survey, one is complete dataset, the other one is incomplete, corrupt dataset.

Below are the data features that were collected from customers:

1). 6 Features were used for training / testing models:

| Model Comparison | | | |
| --- | --- | --- | --- |
|  | Random Forest | Stochastic Gradient Boosting | C50 |
| **Accuracy** | 0.9260257 | 0.9153907 | 0.9093633 |
| **Kappa** | 0.8424587 | 0.8212128 | 0.8059751 |

* **Salary** - not including bonus
* **Age**
* **Level of Education** (6 different levels range from Less Than High School Degree to Master’s / Doctoral Degree)
* **Name of Primary Car** (19 different brands given, choice 20 is ‘None of the above’ if customer has a different car from the list)
* **Zip Code** (9 zip codes given from New England to Pacific)
* **Credit**

2). 1 Feature used for prediction

* **Brand** - 0 - Acer, 1 - Sony

**Analysis**

***Model Comparison and Selection for Training / Testing Data & Feature Importance***

3 predictive models were used for training and testing dataset, the testing scores were pretty close between these models. As the above graph shows, Random Forest has the highest overall accuracy and kappa score when applying parameter mtry=4. Gradient Boosting has slight lower score, the tuneLength was set between 1 and 5, tuneLength 5 had higher score. C50 has the lowest score compare to the other two models. Therefore, Random Forest was selected for prediction as it has the highest training / testing score.

All 3 models indicate the most important features from the dataset are Salary and Age, while Random Forest and Gradient Boosting models have Credit as the 3rd import feature, Car and ZipCode have little importance compare to other features, eLevel has 0 importance. The C50 model indicates ZipCode as the next important feature after Salary & Age, the other 3 features Car, eLevel and Credit have 0 importance.

***Brand Preference Prediction***

The above graph shows the number of preferences for the computer brand ( Acer and Sony ) between Complete Survey data, Corrupt Survey data and the entire dataset. More customers preferred Sony brand in each single dataset, over 9000 customers preferred Sony brand, and a bit over 5000 customers chose the Acer brand.

**Conclusion**

The predicted brand preference for customers from Blackwell Electronic is: 62% of customers prefer Sony and 38% of customers prefer Acer brand.

The most important features that would affect the brand preference are (from most to least):

Salary, Age, ZipCode and Credit, the remaining features Car and Elevel are not that relevant to our prediction.