**Sales Prediction Report for Blackwell Electronics**

For this project we were asked to analyze historical sales data and then make sales volume predictions for four types of products. This will help the sales team of Blackwell Electronics to better understand how types of products might impact sales across the enterprise. We were asked to predict the sales volume for PC, Laptops, Netbooks and Smartphones.

We were working with the existing product data that had only 80 observations. The data was somewhat clean. There was one attribute with missing values, so we deleted that attribute. Also, there were two attributes, like product number and profit margin, that did not make any impact on the sales volume and therefore we deleted those two attributes as well. We also had to recode some of the attributes to numeric values so we could do our analysis.

We looked at the distribution of the data. The price distribution showed us that most of the products sold were in the price range of up to $500. We looked at the product recommendations that had scale 0 to 1, and most observations fell into .6 to .9 scale of recommendations. We looked at some other data redistribution, but it did not get us a lot of insight on the sales volume of the 4 types of electronics that we were asked to predict.

We also looked at the correlation matrix and it gave us insight on how different variables are correlated to each other. We found that customer reviews based on 1 through 5 stars were highly corelated with each other.

We started to build our predictive models and we found that the linear regression model was not the best choice since customers reviews had a lot of collinearity between all 5 variables. Since we needed to predict the sales volume based on customers reviews and service reviews, we cannot delete those variables and run the model. So, we moved on making sales predictions based on the SVM model and we reached an accuracy of .78. This means that out of 10 sales predictions we can predict 7 out of 10 correctly.

We did not stop there, but rather we used another regression algorithm of Decision Tree. We ran Decision Tree with different running parameters and the best one was with the Tune Length of 1. We reached an accuracy of our model of .81. This accuracy is the highest one, and we chose this model to predict sales volume for the 4 types of electronics requested.

We also ran a Gradient Boosting Model, and we achieved an accuracy of .64.

After we chose the Decision Tree model, we ran a new product file to predict sales for the requested types of electronics and got the following predictions:

|  |  |
| --- | --- |
| **Product Type** | **Sales Volume Prediction** |
| PC | 649 |
| Laptop | 321 |
| Netbook | 1849 |
| Smartphone | 1470 |

We can see that netbook and smartphone has a much higher sales volume prediction, whereas PC and Laptop have much lower sales volumes. We are passing this information on to management, and it will be for them to decide if the sales volume mentioned above will make it profitable for Blackwell Electronics to start selling these products in their stores.