**New Product Type Profitability Predictions Report**

In order to assist in determining success of new product in terms of sale volume, analyzing current data can provide valuable insights as to the how do different attributes of product effect sales numbers. This report outlines how specific product types may perform in comparison to each other. In addition, this report investigates how customer service responses affect the sales volumes of the respective product types.

**Methodology**

The data used for this analysis was existing product database which incorporates 80 purchase with the following attributes: Product Number, Product Type, Price, Star Reviews (5,4,3,2,1), Service Review (Positive & Negative), Recommendation, Best Seller Rank, Product Attributes (Depth, Width, Height, Shipping Weight), Profit Margin, and Sales Volume. The second database utilized incorporates As for specific product types, this analysis focused on the following: PC, Laptops, Netbooks, and Smartphones. The new products dataset contained 24 purchases with the same attributes as the existing dataset except that the Sales Volume attribute was left at zero in order predict this attribute.

The process utilized in this multiple regression analysis was to analysis the current Existing Product Data to see which features contributed the most information to predicting new product success. From the existing product data list, specific features were removed to develop a more effective analysis. These features that were removed were Five Star Reviews, Three Star Reviews, One Star Reviews, Negative Service Reviews, Best Seller Rank, and the Product Types not emphasized in this analysis: Product Number, Accessories, Display, Extended Warranty, Printer, Printer Supplies, Software, Tablet, Game Console. Therefore, the following features were included in analyzing new product profitability (see included correlation document for feature selection rationale):

***Product Types***

***Existing Product Attributes***

In order to determine which attributes, contribute the most to understanding the sales volumes based upon product types, a correlation matrix was developed and attributes that exhibited a near perfect correlation with the dependent variable, Sales Volume, were eliminated as well as attributes that exhibited strong positive correlations between each other were analyzed. The following attributes were retained for the analysis:

**Model Selection**

For this analysis, three different models were applied to the Existing Product dataset to develop the most effective training model to more accurately predict the sales volume numbers for the different product types. With each appropriate model, parameters were tuned in order to gain the most accurate model. The three different models applied and their respective model fit statistics are the following:

|  |  |  |
| --- | --- | --- |
| Model | R squared | RMSE |
| Multiple Linear Regression | 0.0189067 | 2900 |
| Support Vector Machine (SVM) | 0.7830149 | 1969 |
| Random Forrest | 0.80993 | 318 |
| Gradient Boost Model | 0.7987539 | 498 |

As illustrated from the chart above, the most accurate models to use for the sales volumes prediction based upon product type would be the Random Forrest and Gradient Boost Model (GBM). Therefore, this model was utilized to predict customer preference for the product types. However, for continuous assurance of model accuracy, all the above model were applied to new products dataset. When the above models were applied to the new products dataset to predict sales volume based upon product type, several models including the Gradient Boost Model, Multiple Liner Regression and Support Vector Machine predicted negative values for the Sales volume attribute. Given that a negative sales volume is not consist with the data, the Random Forrest model with 100 trees illustrated the most accurate prediction for sales volume and no negative values for predicted sales volume and there was utilized for this analysis.

**Predicting Customer Preference**

A key outcome for this analysis was to predict the potential sales volume of the product types (PC, Laptop, Netbook, Smart Phone) based upon existing product sales information. The attached excel sheet (“Product Predication”) provides the corresponding predicted sales volumes for each product type. From this sheet, a couple key insights can be drawn:

* Highest Sales Volume was found with the Tablet category and out category of the Netbook
* Lowest Sales Volume was within the more expensive categories of Laptops and Smartphones
* The price was a significant determiner of sales volume within categories

**Relationship Between Customer/Service Reviews and Sales Volume**

In addition to predicting sales volume based upon respective product type, this report looked at the relationship between customer/service reviews and sales volume. The following chart indicates the various customer reviews based upon product type and their correlation coefficient:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 5 Star Review | 4 Star Review | 3 Star Review | 2 Star Review | 1 Star Review | Positive Customer  Review | Negative Customer Review |
| Sales Volume | 1 | 0.88 | 0.76 | 0.49 | 0.255 | 0.622 | 0.31 |

From the table, above, there was a perfect correlation between 5 star reviews and sales volume while both the 4 star and 3 star as showed significant effect on sales. In the area of customer reviews, a sentiment analysis illustrated that positive service reviews has a much higher significant effect on product sales volumes as compared to negative service reviews. The full correlation coefficient table is attached for further information.