****

**Report**

**Predicting Profitability**

**January 2020**

**Version 1.0.0 | By Murilo Miranda Nüsslein**

Content

[Context 3](#_Toc30589004)

[Objective 3](#_Toc30589005)

[Methods 3](#_Toc30589006)

[Results 4](#_Toc30589007)

[Algorithm performance 4](#_Toc30589008)

[Profitability of the new products 5](#_Toc30589009)

[Recommendations 5](#_Toc30589010)

# Context

Our sales team is developing a new business strategy, so they want to add some new products to the catalog. However, they do not know which products should be selected. So far, they have a list of 17 new products, but they only want five new ones. In this analysis, we assume that certain attributes are associated with highly successful (current) products and, therefore, any potential new products that also have these attributes will be similarly successful.

# Objective

Our goal was to create a predictive model to select five new products from a list of 17 products, with greater profitability. Therefore, we need to forecast the sales volume for each of the potential new products.

# Methods

We received two datasets, one with information about products in the catalog and the other with information about new products. Both have variables such as **customer preference** (5-1 Star Reviews, Positive Service Review, Negative Service Review, Would recommend and Best Sellers Rank), **product characteristics** (Category, Number, Weight, Depth, Height and Width) and **economic** (Shipment, Price, Profit margin, Volume sales). In the pre-processing phase, we removed some attributes (Product Type, Product Number and Best Sellers Rank) because they have no relevant information for the models or have missing values. Then, we checked the matrix correlation and decided to remove the 5, 3 and 1 Star Reviews, because the 5 Stars Reviews has a perfect correlation with sales volume and 3 stars and 1 star have high collinearity (correlation > 0.90) with other variables. As the remained variables have different units, they were Z-transformed, except sales volume was not transformed.

Subsequently, we trained our models using three algorithms: (1) k-Nearest Neighbor (kNN), (2) Gradient Boosted Trees (GBT) and (3) Support Vector Machine (SVM). Before using the models to predict sales volume, we had to find the best parameters for each algorithm via an operator which seeks the best combination of parameters. The parameters that varied were: number of folds in the cross-validation (between 2 and 20), C value in the SVM (between 0 and 100), number of trees (between 1 and 100) and learning rate (between 0 and 1) in the GBT, and number of k in the kNN (between 1 and 20). Then, we compared the models using root mean squared error (RMSE) and squared correlation (R-Squared). In the end, we decided to use the best model for each algorithm. The best models were used to make predictions about sales volume for new products. After predicting the sales volume for each potential new product, we calculated the profits by multiplying the predicted sales volume by the price of the product and its profit margin. Finally, we ranked all products in order from highest to lowest profit.

# Results

## Algorithm performance

We preferred to use the models with the highest squared correlation values. Their values are listed in the table below.

Table 1. Performance of the different algorithms. Models in bold values were chosen to predict the volume sales of the new products. Root mean squared error (RMSE) and squared correlation (R-Squared).

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Support Vector Machine** | | | | |
| Number of folds | Kernel type | C | RMSE | R-Squared |
| 13 | Dot | 48.48 | 558.92 | 0.742 |
| 12 | Radial | 96.85 | 868.05 | 0.451 |
| 12 | Polynomial | 6.88 | 1038.76 | 0.505 |
| **13** | **Dot** | **6.88** | **794.26** | **0.827** |
| 12 | Radial | 93.66 | 993.66 | 0.471 |
| 12 | Polynomial | 6.88 | 1253.52 | 0.769 |
| **Gradient Boosted Trees** | | | | |
| Number of folds | Number of trees | Learning rate | RMSE | R-Squared |
| 19 | 70 | 0.03576 | 634.67 | 0.864 |
| **19** | **70** | **0.03216** | **650.79** | **0.912** |
| **k-Nearest Neighbor** | | | | |
| Number of folds | Weighted vote | k | RMSE | R-Squared |
| 12 | True | 8 | 673.55 | 0.689 |
| 12 | False | 8 | 679.39 | 0.690 |
| 12 | True | 8 | 789.79 | 0.794 |
| **12** | **False** | **8** | **754.59** | **0.803** |

## Profitability of the new products

On average profit between the models, the highest profitability product is **Dell PC Number 172** ($238,188.73) in the new product list, following by **Motorola Smartphone 196** ($93,918.24), **Dell PC 171** ($87,138.57), **Asus Netbook 181** ($58,732.73) and **Razer Laptop** ($58,288.33).

Table 2. The top 5 products with the highest average profitability. Profit (SVM) has values predicted by Support Vector Machine algorithm, Profit (GBT) predicted values by Gradient Boosted Trees and Profit (kNN) predicted values by k-Nearest Neighbor. Average profit has the average predicted values using three algorithms.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Product | Profit (SVM) | Profit (GBT) | Profit (kNN) | Average profit |
| Dell PC 172 | $125,876.06 | $446,790.12 | $141,900.00 | $238,188.73 |
| Motorola Smartphone 196 | $73,407.32 | $99,727.90 | $108,619.50 | $93,918.24 |
| Dell PC 171 | $30,414.36 | $163,809.98 | $67,191.38 | $87,138.57 |
| Asus Netbook 181 | $25,546.03 | $114,338.09 | $36,314.08 | $58,732.73 |
| Razer Laptop | $57,224.20 | $50,054.61 | $67,586.19 | $58,288.33 |

# Recommendations

We should think about diversifying our product catalog. For example, we could include tablets in the list of new products, even if the models predicted that they will not be highly profitable (between $19,293.77 and $26,005.99). However, looking at the rank of the best sales, tablets are in the best positions (rank 1st Amazon tablet and 34th Apple tablet). So, we suggest exchanging Motorola Smartphone 196 for one of the tablets because it has a bad ranking (rank 44,465th) in the list of the best sales.

We should also switch from Razor Laptop to Apple Laptop. On average, they had similar profitability, however Razor Laptop has a worse position on the bestseller list (rank 2820th) compared to the Apple Laptop's position (rank 111th).

To put it briefly, the five new products should be: Dell PC 172, Dell PC 171, Asus Netbook 181, Apple Laptop and Apple/Amazon Tablet.