****

**Report**

**Predicting Profitability**

**January 2020**

**Version 1.0.0 | By Murilo Miranda Nüsslein**

Sumário

[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. Up to now, they have a list with 17 new products, but they only want five new products. 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 on a list of 17 products, which have the highest profitability. Therefore, we need to forecast the sales volume of each of the potential new products.

# Methods

We received two datasets, one has information about existing products in the catalog and another has information about the new products. Both has variables like **customers’ preference** (5-1 Star Reviews, Positive Service Review, Negative Service Review, Would recommend and Best Sellers Rank), **product’s features** (Category, Number, Weight, Depth, Height and Width) and **economic** (Shipment, Price, Profit margin, Volume sales). In the preprocessing phase, we removed some attributes (Product Type, Product Number and Best Sellers Rank) because they either do not have relevant information for the models or they have missing values. After, we checked the matrix correlation and decided to remove 5, 3 and 1 Star Reviews because the 5 Star Reviews has a perfect correlation with the volume sales, and 3 star and 1 star have high collinearity (correlation > 0.90) with other variables. As the remained variables have different units, they were Z-transformed, except volume sales was not transformed.

Later, we trained the 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 the volume sales, we had to find the best parameters of each algorithm via an evolutionary optimize parameters. The parameters that could vary 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). Next, we compared the models by root mean squared error (RMSE) and squared correlation (R-Squared). At the end, we decided to use the best model of each algorithm. The best models were used to make predictions about volume sales of the new products. After predicting each potential new product’s sales volume, we calculated the profits by multiplying the predicted sales volume by the product’s price and its profit margin. Finally, we ranked all products in order of 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 also think about to diversify our catalog. For instance, we could also include tablets. Even though they have low profitability (between $19,293.77 and $26,005.99), they are in great position in the best sellers rank. Amazon tablet and Apple tablet are in the rank 1 and 34, respectively.

Motorola Smartphone 196 is the second best in the profitability rank, but in the best sellers’ rank is on 44465 position.

We could also remove Razer Laptop from this list and add Apple Laptop, on average they have similar profitability, but Apple Laptop is the best sellers (rank 111) than Razer Laptop (rank 2820).

In conclusion, the final five products should be: Dell PC 172, Dell PC 171, Asus Netbook 181, Apple Laptop and Apple/Amazon Tablet.