

# BLACKWELL

NEW PRODUCTS LAUNCHING





# ANALYSING EXISTING PRODUCTS TO PREDICT SUCCESSFUL NEW ONES.

## STATE OF ART

Predicting outcomes using  
Regression Models

Regression analysis entails looking at dependent variables (outcomes) and an independent variable (the action) while also assessing the strength in the association between them. In other words, it looks to understand if there is a relationship between variables and how strong that relationship is.

## DATA QUALITY

Far from perfect, but still useful.

There are some notable issues with the data provided. The Best-Sellers Rank on the existing products csv file, for instance, is lacking some information so we had to delete it in order to have more accurate results. Also, when we performed the Correlation Matrix we noticed a perfect correlation between the 5-Star-Reviews with the product's volume, showing us a potential data error. So, the 5-Star-Review was also not a trustable attribute.

We had to make some adjustments in order to achieve the best model possible, nevertheless an extremely good one would not be possible when reading this data set. But despite all issues we expect to have an acceptable result..



# **METHODS**

## **CORRELATION MATRIX**

It shows us all the linear correlations. If the linear correlation is greater than 0.95 we should exclude this attribute because it is potentially leading us to wrong predictions. Using this tool we decided which attributes certainly needed to be removed.

## **K-NN REGRESSION TOOL**

This algorithm uses ‘feature similarity’ to predict values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set. We trained a model using the existing products and then we applied this model to the new products set, so we could predict its sales volumes.

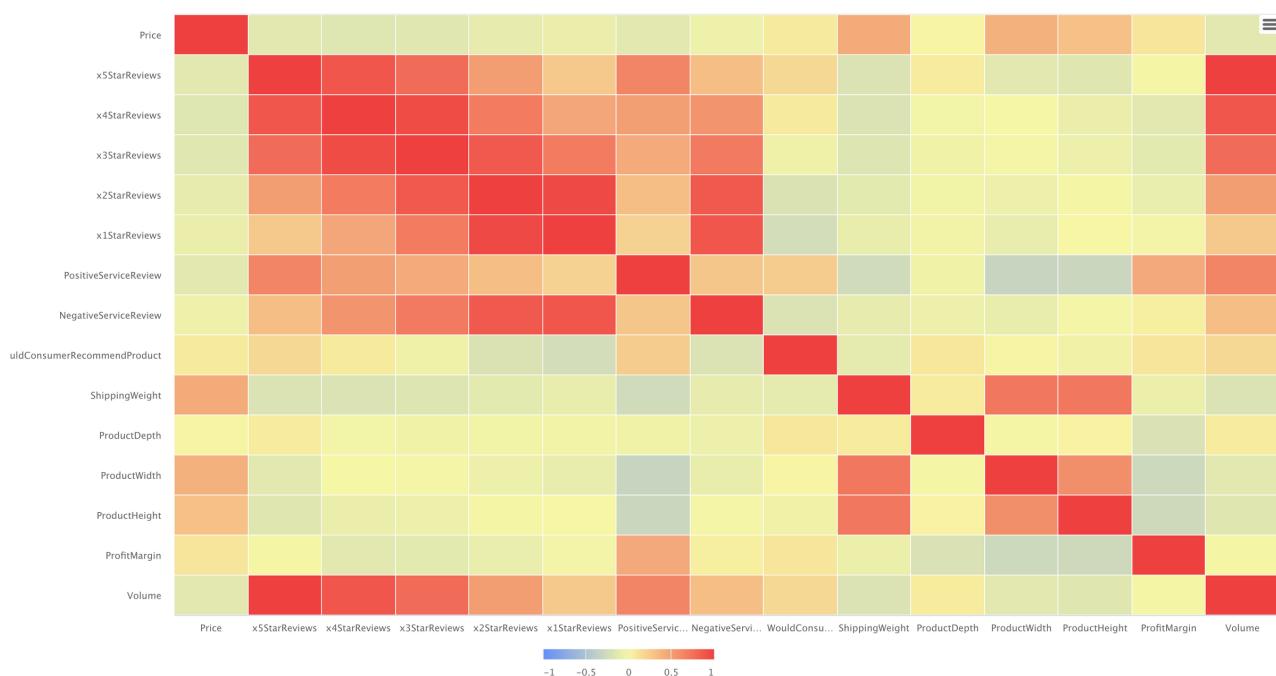
## **EXCLUDING SOME ATTRIBUTES, FOR GOOD.**

In order to improve our model we had to exclude some attributes.

The best possible model we got uses the following attributes: Positive Reviews, Negative Reviews, Volume, Would Customer Recommend, 2-Star-Reviews and 4-Star-Reviews.

# PRE-PROCESSING AND CORRELATION MATRIX

How did we choose which attributes to delete based on the Matrix?



"1" MEANS PERFECT CORRELATION, THAT IS REDUNDANT INFORMATION. THIS IS WHY 5-STAR REVIEWS ARE NOT TO BE TRUSTED IN THIS CASE.

Highly correlated attributes can be easily removed by simply using the Remove Correlated Attributes operator. Correlated attributes are usually removed because they are similar in behavior and will have similar impact in prediction calculations, so keeping attributes with similar impacts is redundant. Removing correlated attributes saves space and time of calculation of complex algorithms.

**Normalising attributes:** We made a range transformation (from 0 to 1) in order to have an accurate result, once the measurements in the data are very distinct (there are weight, price, and other different values that would cause trouble on results).

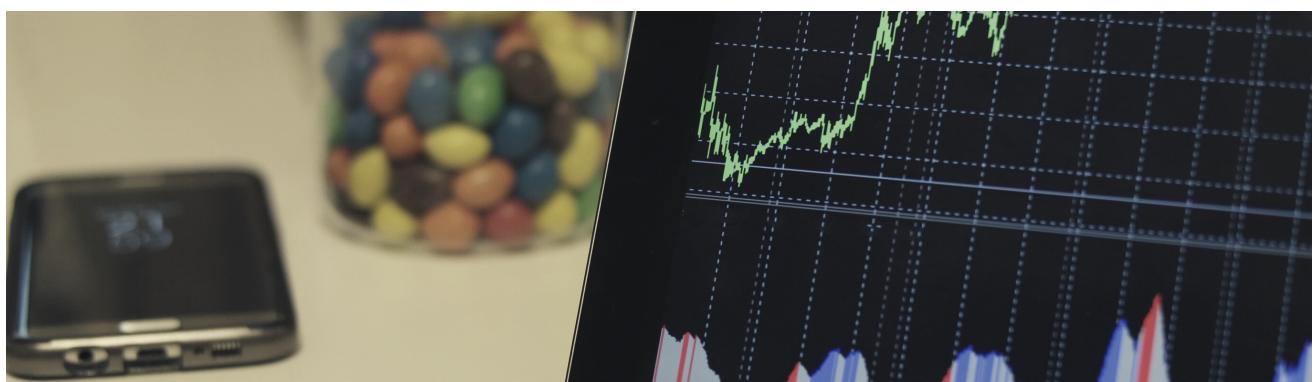
Product Type was removed because it cannot be measured (no numerical value). Product Number is an integer and was removed to avoid "noise" on data (labeled as "ID" so it could still be presented on results).

# K AND PERFORMANCE METRICS

## Indicators of Good Performance

**K Number:** It is the number of neighbours which are going to be used to calculate the result. In KNN, finding the value of k is not easy. A small K value suggests that noise is probably highly influencing the result and a large value makes it computationally challenging. Data scientists will usually choose as an odd number if the number of classes is 2. One other simple approach would be to select  $k=\sqrt{n}$ .

**Absolute Error:** It is the amount of error in your measurements. It is the difference between the measured value and “true” value. Our model presented an absolute error of 338.565 +/- 366.018 (micro average: 338.565 +/- 1195.851).



**AFTER TESTING A MYRIAD OF NUMBERS,  
WE USED K = 9 FOR OUR BEST MODEL**

**Squared Correlation:** R-Squared is a statistical measure of fit that indicates how much variation of a dependent variable is explained by the independent variable(s) in a regression model. Our model's squared correlation is 0.854 +/- 0.129 (micro average: 0.412).

**Root Mean Squared Error:** It is a standard way to measure the error of a model in predicting quantitative data. RMSE can be thought of as some kind of (normalized) distance between the vector of predicted values and the vector of observed values. Our model presented a root mean squared error of 755.297 +/- 1040.413 (micro average: 1242.854 +/- 0.000).

# WHAT BLACKWELL SHOULD SELL:

Product Type	Brand	Price	Profit Margin	Sales Volume	Monthly Profit
<b>TOP 5:</b>					
PC	Dell	\$699	0.25	904	\$157933
Tablet	Amazon	\$199	0.2	3.154	\$125518
Laptop	Apple	\$1199	0.1	1.021	\$122367
Tablet	Apple	\$629	0.1	1.275	\$80225
Laptop	Razer	\$1999	0.23	103	\$47455
<b>OTHER OPTIONS TO CONSIDER:</b>					
Netbook	Acer	\$329	0.09	1136	\$33642
Game Console	Sony	\$249.99	0.09	1275	\$28685
PC	Dell	\$860	0.2	149	\$25574

Applying our trained model we could get the predicted sales volume for each product and therefore calculate the monthly profit. The products are ranked according to the monthly profit (descending)



## FINAL CONSIDERATIONS

### SELLING LESS, EARNING MORE.

We can easily realize that by selling 103 Razor's laptops per month Blackwell's profit would be greater than selling 1136 Acer's. So Blackwell should definitely choose Razer, right? **Maybe not.** Sometimes having more transactions is more desirable because it means more customers in the store (or even online) so this is great opportunity to try to sell them more items. More people entering your store should be taken into considerations.

### FEED THE GAMERS.

Gaming has not been Blackwell's first focus, but maybe you should give it some highlight. We are talking about one of the most ascending markets. It would attract both children and older people. Consider including Sony's Playstation to your store, as it would gather more attention for the gamers, who are likely buying more items than just the console. Razor's top laptop is also a great option: and a plus - it offers incredible profit margin.

### YES, APPLE...

Selling Apple goods is certainly a must. It attracts people to the store. Everyone loves testing their products and eventually buying it. Although it doesn't offer a high profit margin it is totally worth it because of its sales volume.

### THE KINDLE TREND.

Adding the top Kindle to your store is perfect. Your mid-aged consumers will love it, and they are more likely to make higher purchases, including a reading tablet upgrade.