1/25/2022

Danielle sherman

CTO, Blackwell Electronics

Overview:

As per your request, I analyzed the provided historical sales reports and created predictive models to forecast the following product types: PC, Laptops, Netbooks, and Smartphones. Below you will find an overview of the algorithms tested and the sales predictions.

**Random Forest Regression**:

I built 10 models using various random forest methods. I determined that Quantile Random Forest performed the best with the least amount of overfitting, and also had the highest Rsquared result. No negative values and thus this algorithm was chosen for the final predictions. Therefore, Random Forest resulted in the highest level of confidence.

**Support Vector Machine Regression**:

I built 9 SVM models using those available in caret. L2 Regularized SVM (dual) with Linear Kernel returned a perfect score of 1.0. This was a high indication of overfitting and this model was rejected. Two other SVM models were rejected as they did not fit the dataset and errored immediately, from the lack of “x” not being a character matrix with a single column for string kernel methods. The other models all produced negative values, which is a indicator of a faulty prediction.

**Gradient Boosting (Regression)**:

I built 5 Gradient Boosting models for regression. The following 3 models all returned values higher than 88%, xgbDART, xgbLinear, xgbTree, but contained negative values in the models. Again, this is an indicator of a poor prediction and were omitted. The final 2 gbm models were built using Gradient Boosting Machines and Stochastic Gradient Boosting, and they yielded results under 50%, which is too low to accurately predict.

**Sales Predictions**:

Making sales predictions with such a small dataset is not usually advised. This is due to a high occurrence of overfitting and a large amount of warnings indicating the dataset is small and may not yield optimal results. That being said, building of the models was adjusted to account for the small dataset, i.e. 3 crossfolds instead of 10.

The predictions are located in the file: C3.T3output.rf.csv.

The forecasted sales for the for the product types are as follows:

PC – 636

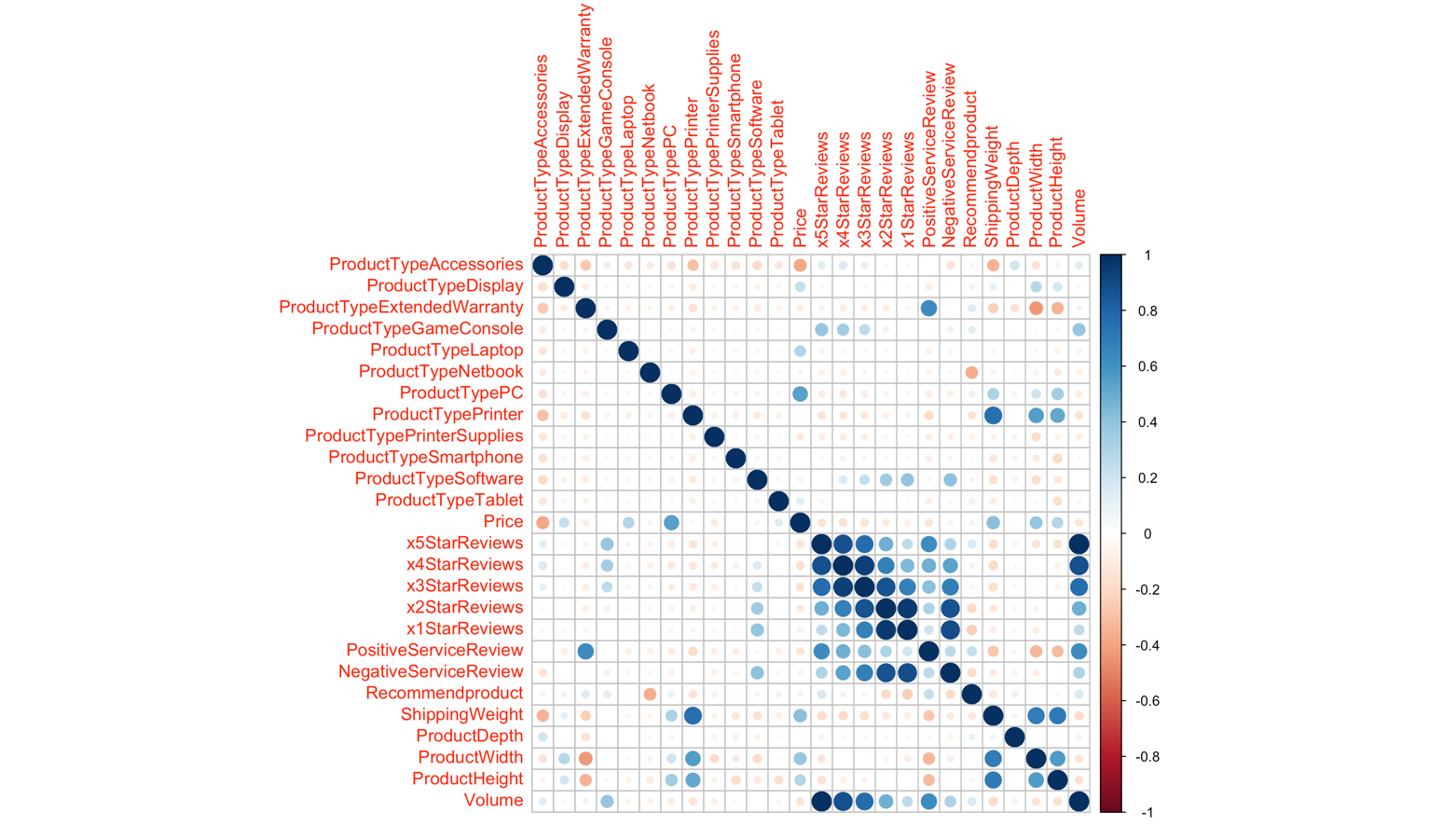
Laptop – 336

Netbook – 1404

Smartphone – 1292

**Impact of Customer Reviews**:

Below is a heatmap chart of all the features in the data set and how strong or weak of a connection each feature has to another. Looking at the Reviews on the left side of the chart and following to the intersect point with Volume (last column on the right), one can see how important positive reviews are for sales volumes. The higher the rating, the darker the blue dot, meaning the more sales volume. It is extremely important to follow-up with customers after a sale is complete in order to get a product rating/review.



GRANT JOHNSON