# C3T3 – Report

Objectives

Accurately predict the sale volumes of specific product types at the Blackwell Electronics stores. Using historical sales data, I must build 3 models to compare and predict sale volumes using R and RStudio. The 3 models used in this analysis were Gradient Boosting (GBM), Random Forest (RF), and Supported-Vector Machines (SVM).

Exploratory Data Analysis (EDA)

There are several features which are useless when making the predictive models such as Product Number or Product dimensions. Therefore, these were taken out of the modeling data and the rest was replaced by ‘dummy’ variables to allow the modellign algorithms to easily run. Furthermore, this would make it easier to see correlations between product types and other features using a confusion matrix (see below).

Chart

Description automatically generated

Additional information can also be gathered using the explore library on r, the results of which can be found in GitHub.

Models

The predicted feature used was the Volume variable and the three models where trained, tested and compared using the root-mean squared error values (RMSE). Generally, the lower the RMSE, the better the model. Here are the results:

|  |  |  |
| --- | --- | --- |
| **Model** | **RMSE** | **Rsquared** |
| RF | 587 | 0.939 |
| GBM | 982 | 0.813 |
| SVM | 182 | 0.970 |

At first glance it looks like the best model is the SVM model as it has the lowest RMSE, however, when looking at the actual model results and predictions, both GBM and SVM models give negative values for predictions. Since this is volume data (quantity of products), negative values are meaningless. Therefore, the best model that fits is RF.

 Predictions

Predictions were created using the Random Forest model and then written in a new .csv file. Predictions for not only the required product types but also all the product types given were created and can now be used to make better informed business decisions such as marketing and advertisement adjustments, inventory volumes for each category, and even expected profit margins.