Blackwell Electronics

|  |  |
| --- | --- |
| To: | Danielle Sherman, CTO, Blackwell Electronics |
| Cc: | E. Logan Crawford |
| From: | Kate Koebbe |
| Date: | 18 January 2022 |
| Re: | Sales Prediction Report |

Our team recently performed analyses on the sales performance of specific product types in one our company stores in an effort to reassess sales predictions. Utilizing existing and new product attribute data, we selected three different algorithms to determine the best performing model to make our predictions. Below are our findings.

Our analysis is based upon data from 104 sales records, 80 of which are existing products and 24 that are new products. Prior to running our models, we performed some basic exploratory data analysis and utilized correlation analysis to get a better understanding of how certain attributes may or may not impact our sales volumes. Through this process, we created new datasets with some highly correlated features removed to tests against our full (“out-of-the-box”) dataset to which may perform better.

We utilized Random Forest Regression, Support Vector Machines Linear, and Extreme Gradient Boosting models in our analyses of the two different datasets. To determine the best model to use in our predictions, our team compared the R-squared values and root mean squared error (RMSE) values of each model. R-squared values show the proportion of variance in the dependent variable that can be explained in the independent variables and the RMSE measures the difference between predicted values and actual values. For the out-of-the-box dataset, our Random Forest Regression model yielded an average R-squared value of 83.44% (RMSE 912.29). The Support Vector Machines model resulted in an R-squared value of 92.79% (RMSE 170.99). Finally, the Extreme Gradient Boosting model returned an R-squared value of 93.93% (RMSE 655.92). Models performed on the limited attribute datasets (i.e. correlation datasets) yielded R-squared values significantly below those of the out-of-the-box sets and were, therefore, excluded from selection. Our team ultimately decided to use the Support Vector Machines model to make predictions due to the relatively high R-squared value and the significantly lower RMSE compared to the other models.

The Support Vector Machines model determined the following predictions for the volume of sales for the products in question: PCs (1078), Laptops (548), Netbooks (1661), and Smartphones (2380). We utilized the same model to validate our results, which returned an R-squared value score 83.83%, which is largely consistent with the metrics in our training set.

One of the most significant findings of our Support Vector Machine model was the impact of service reviews on product sales volume. The chart below illustrates the relative importance of the top 20 attributes in our model: of those 20, reviews are the most significant factor in determining the volume of sales. Our sales department can benefit from this information by actively seeking reviews from our customers after purchase.

Please do not hesitate to contact me or my team should you have any questions.

Chart

Description automatically generated

Respectfully,

Kate Koebbe

Data Analyst

Blackwell Electronics