**Course Three Task3 Report**

**by**

**Shifeng (Steve) Li**

**03/04/2022**

In this project, the Blackwell Electronics historical sales data and new sales data sets are analyzed using R caret data analysis techniques to predict sales of difference product types, and access the impact the reviews have on sales.

1. **How the service review and customer reviews have the impact to the sales volume?**

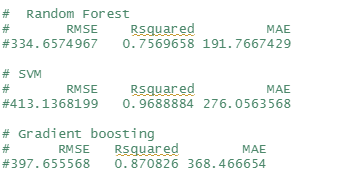
From the quantified correlation analysis, found that the correlation value between 5-star customer review and sales volume is **1**, the correlation value between 4-star customer review and sales volume is **0.87**, the correlation value between 3-star customer review and sales volume is **0.76**, and the correlation value between PositiveService review and sales volume is **0.62**.

Therefore, the service review and customer reviews do have the impact to the sales volume. And the 5-star customer review has the highest impact.

1. **Can the new sales data set predict the sales of the PC, Laptop, Netbook, and Smartphone very well?**

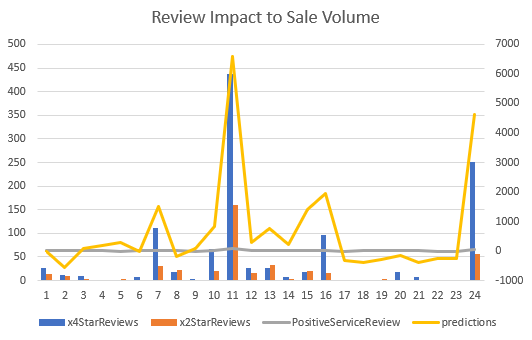
To find out the answer, applied three different regression algorithms to do the machine learning process on the historical data set. Build the models with the Support Vector Machines algorithm, Gradient Boosting algorithm, and Random Forest algorithm. The conducted the machine learning on the training data that are focused on the four product types, then tested against the test data set to collect the quantified prediction accuracy level for forecasting the four product types sales volume.

The following listing is the top level summary of the prediction accuracy levels from the three models.



Based on the comparisons of the three models, the Support Vector Machines algorithm based model is chosen due to its highest R-squared value 0.968884 and its RMSE value 413 which is not significantly higher than other two models.

As the additional illustration for the review related features’ impact to the sales volume, the following chart shows the how the several review attributes in the 24 data records of the new sales data set impact the sales volume. Note the previously mentioned identified 5-star review and 3-star review are not in the chart as the data set has been modified by removing a couple of fields in order to fit the model training data set layout.

****

**Uploaded to: https://github.com/UTOct21DaPtSteve/Steve\_Course3**