Monika Spreitzer

C3T3 Report

1. The algorithms you tried.

Rf  
svmLinear  
svmLinear2  
xgbLinear

1. The algorithm you selected to make the predictions, including a rationale for selecting the method you did and the level of confidence in the predictions.

Below are my results of the 8 tests I performed.

A screenshot of a cell phone

Description automatically generated

\*\*Group one had the following features removed – ProductNum, BestSellerRank, ProfitMargin

\*\*Group two had the following features removed – x5StarReviews, ShippingWeight, ProductDepth, ProductWidth, ProductHeight, ProductNum, BestSellerRank, ProfitMargin

I used process of elimination to decide which algorithm to use. Here is my logic:

* **UID 8:** (Gradient Boost) The entire dataset was sampled to train. When using test data we achieved100% accuracy -- I deemed this to be overfit.
* **UID 5:** (Random Forest) **I consider this my best fit model.** It had fairly good Train R2 accuracy and even better Test R2 accuracy.
* **UID 3:** (SVM2) The entire dataset was sampled to train and predicted the test data with 98% accuracy. I deemed this to be overfit. This was also part of group one which included x5StarReviews.
* **UID 1:** (Random Forest) This one performed well; however, it was part of group one (which included x5StarReviews) so I believe it’s overfit.
* **UID 4:** (Gradient Boost) The entire dataset was sampled to train and predicted the test data with 98% accuracy. I deemed this to be overfit. This was also part of group one which included x5StarReviews.
* **UID 2:** (SVM1) This was part of group one (which included x5StarReviews) so I believe it’s overfit.
* **UID 6:** (SVM1)The train accuracy was significantly higher than the train accuracy which is likely overfit.
* **UID 7:** (SVM2) The train accuracy was significantly higher than the train accuracy which is likely overfit.

1. Your sales predictions for four target product types found in the new product attributes data set
2. A chart that displays the impact of customer and service reviews have on sales volume.

A screenshot of a social media post

Description automatically generated

Star reviews and positive service reviews play a big role in predicting sales volume.

1. Was it straightforward to rerun your projections of sales volume using both models?
   1. Yes. Creating a pipeline and modifying a few things here-and-there made it easy.
2. What are the main lessons you've learned from this experience?

The biggest lesson learned here was ALWAYS run your predictions on your train and test sets and compare!

1. What recommendations would you give to the sales department regarding your findings relating to the different types of reviews?

I would recommend focusing sales on netbooks and smartphones. Also, encouraging buyers to write reviews will help to create more robust data.