Final: Predicting Supply Chain Product Backorders

by

Heather DeSimone

Fundamentals of Machine Learning MIS-64060-001

5/8/2022

Problem Overview

We often forget the importance of our supply chain in business. We concentrate on maximizing profits through surface level approaches such as marketing, product/service sold, sales, etc. But the fundamental stonework for any business is the supply chain. It is how we manufacture our product and get it to the consumer. Without it, we have nothing to sell.

Currently, we are struggling within supply chain on almost every level. Raw materials are in higher demand than supply, we are facing labor shortages in our factories as well as our logistics, and ports are clogged with containers waiting months to be unloaded.

Backorders are creating major issues throughout the market. Without inventory, consumers are unable to purchase product, and without consumers purchasing the product, revenue is in decline. Furthermore, backorders cause a firm to eat the costs of any unsold goods because the shipment arrived too late. For instance, each holiday season we see sales skyrocket as consumers purchase their holiday gifts. If a product were purchased and placed on backorder then it would not be available for the holiday rush, but would still arrive in the warehouse to sit and wait to be sold. Most likely, the product would need to be placed on sale in order to move the supply. This is a loss of revenue not only because the product needed to be placed on sale, but also because the firm lost the potential sales if it had been available to the consumer during the peak buying season.

If we were able to predict backorders ahead of time, we could mitigate some of these losses. Through machine learning, we can build a predictive model that will use historical data from other products that went on backorder, to predict if a future product will do the same.

Data

The data set used to build the predictive model was a real-world data set pulled from Kaggle (https://www.kaggle.com/datasets/chandanareddy12/back-order-prediction). The variables of the data set were sku code, current inventory level of component, transit time, quantity in transit, forecast sales for the net 3, 6, 9 months, sales quantity for the prior 1, 3, 6, 9 months, minimum recommended amount in stock, indicator that the item may have a potential issue, parts overdue from source, source performance in the last 6 and 12 months, amount of stock orders overdue, general risk flags, and if the product went on backorder. The last variable – if the product went on backorder – is the target variable. I used all variables in building the model except for sku code.

Methodology

I created a k-NN classification model to predict if a particular product would go on backorder, based on the variables used in training the model. I used the k-NN algorithm for this specific problem because of its level of simplicity and accuracy when making a specific prediction based on the same variables for similar items. Furthermore, I wanted to make my prediction of the 3 highest sold skus, rather than a large number of items.

Since the k-NN classification model is a lazy algorithm (memorizes data) it can easily overfit to the training set. For this reason, I split my data into a training set, validation set, and a test set to avoid overfitting when making my final predictions.

Analysis

I chose the 3 products with the highest forecasted sales over the next 9 months, to predict if these products would or would not go on backorder. The 3 skus were sku#3449673 (9 month forecasted sales of $3,162,260), sku#3301406 (9 month forecasted sales of $3,124,704), and sku#3337435 (9 month forecasted sales of $1,858,864). Per the model, none of these 3 items were predicted to go on backorder.

Conclusion

Since none of the 3 products are predicted to go on backorder, the firm should not overbuy product or prematurely buy product in preparation. Normal buying practices should be conducted for these 3 skus based on current trends and forecasting.