A Predictive Analytics Approach to Understand, Predict and Reduce Distribution Center Stockouts

Ankit Pandey, Mayank Daga, Amjad Hasnain, Nikeeta Brijwasi, Matthew A. Lanham

Purdue University Krannert School of Management

pandey31@purdue.edu; mdaga@purdue.edu; ahasnai@purdue.edu; nbrijwas@purdue.edu; lanhamm@purdue.edu

Abstract

This study provides a solution towards understanding, predicting, and reducing distribution center stockouts for a major CPG conglomerate. The motivation for this study is that distribution center (DC) stockouts have a direct impact on the revenue and profitability of the CPG as well as the retailers they supply. Having the ability to accurately predict and identify stock out root causes at the DC can lead to better business performance measures, and improving customer satisfaction downstream.

Our solution categorizes different products based on ABC classification approach. Then specific predictive models are implemented to predict stockouts for each category. We found that both ARIMA and LSTM performed significantly better than the naïve forecasts on a standalone basis. We saw significant improvements in the forecast accuracy by using a weighted average combination of LSTM and ARIMA methods particularly for "A" and "B" category SKUs

Introduction

Effective Inventory Management is one of the areas that has potential to save huge cost for businesses. Excess inventory results in increase in warehouse management costs and inventory holding costs. Reducing inventories is a great business proposition. At the same time, inventory stockout results in lost sales. Just to avoid excess inventory, a company cannot understock the products. A fine balance between excess inventory and experiencing stockout should be maintained. Many companies are investing heavily in forecasting demand more accurately. Syncing manufacturing and supply chain to demand is the key to have optimum inventory levels.

In collaboration with a national CPG partner, we explored the root-cause of stock outs at the distribution center by identifying whether the stockouts are caused by variability in demand side parameters, such as safety stock calculations and stock allocation rules, or on the supply side, such as production schedules and material procurement policies. For products experiencing high variations in demand we explored different time series forecasting approaches such as naïve forecast, ARIMA methods, and more sophisticated machine learning approaches such as LSTM recurrent neural networks to predict the fluctuations in demand and decide an optimum inventory level.

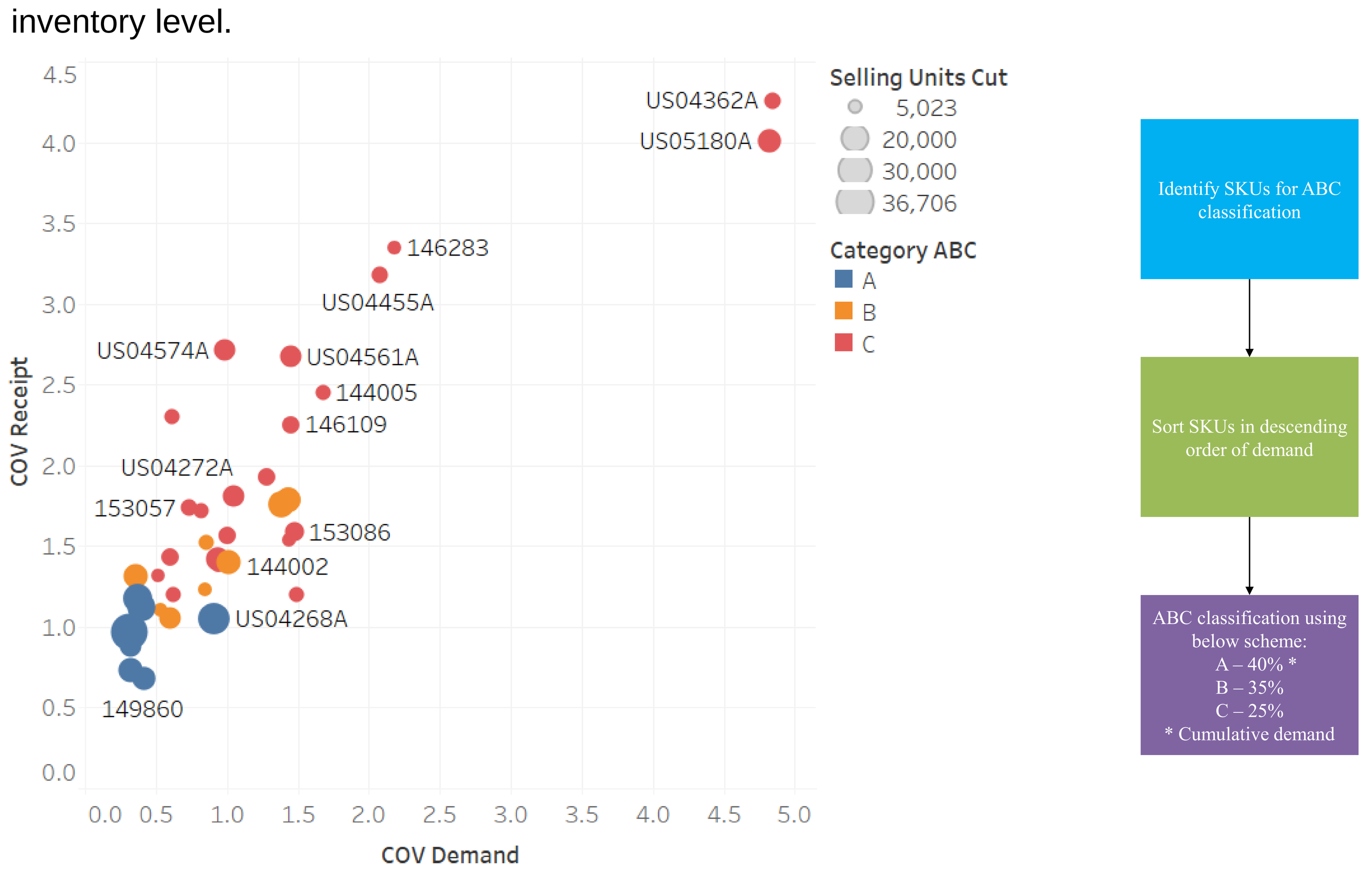


Figure 1. Product categorization by variations in supply and demand; ABC classification

We discuss how business decisions such as production schedules, material procurement policies, shipping and replenishment policies can affect stockouts besides demand patterns, and recommend business actions to mitigate stockouts resulting from such factors. Overall, our cross-validated solution is expected to reduce DC stock-out incidents and provide a better understanding of where improvements can be made based on root-cause identification.

Literature Review

Prior to developing our model, we researched elite journals and studies in time series forecasting to predict the demand in the supply chain domain. Various approaches and methods have been studied to predict future demands.

Author, Year	ARIMA	RF	ANN	SVM	LSTM	RNN	LR	MIP	DNN
WANG Guanghui, 2012				✓					
Ediger and Akar, 2006	✓								
G. Peter Zhang, 2001	✓		✓						
Azzouni and Pujolle, 2017					✓	✓			
Kevin Bonnes, 2015			✓				✓	✓	
Aburto and Weber, 2015		✓			✓				✓

After discussing and evaluating various approaches, we decided to proceed with LSTM and ARIMA models.

Methodology

We followed Box-Jenkins approach to finalize the models and parameters by performing repeated diagnostic checks.

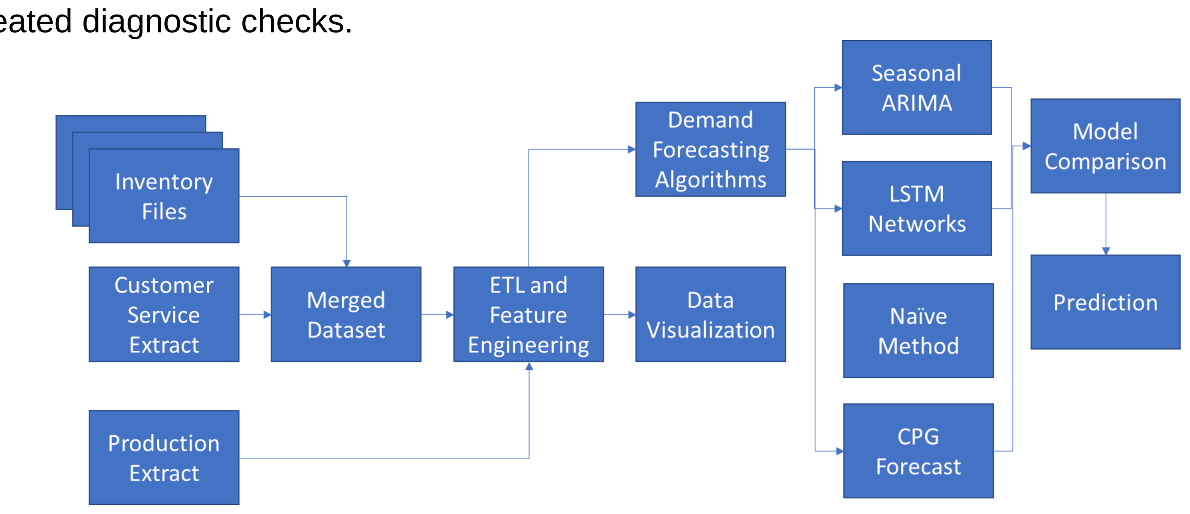


Figure 2.a Study Design

Data : We received daily inventory and supply, demand, and production data for each product for past three years in three different files

Data Cleaning & Pre-Processing : We merged data from all the three sources. Some of the products were obsolete (no production) or had little or no demand. Since data was not continuous for these SKUs, we filtered such SKUs for further analysis. Post processing, we had about 225 SKUs which were further analyzed for stock cuts and demand and supply variations

ABC classification: We further categorized SKUs into "A", "B", and "C" categories based on their demand. Category "A" contains high demand SKUs responsible for 40% of demand, "B" contains moderate demand SKUs responsible for next 35% of demand and category "C" contains low demand SKUs responsible for remaining 25% of the demand.

Patterns and Trends: To identify trends, seasonality, and the stationarity in the demand we plotted time series plot of for different products along with rolling mean and rolling standard deviations of demand and also used ACF plots to identify periods which were correlated to current demand. We tested various combinations of model parameters

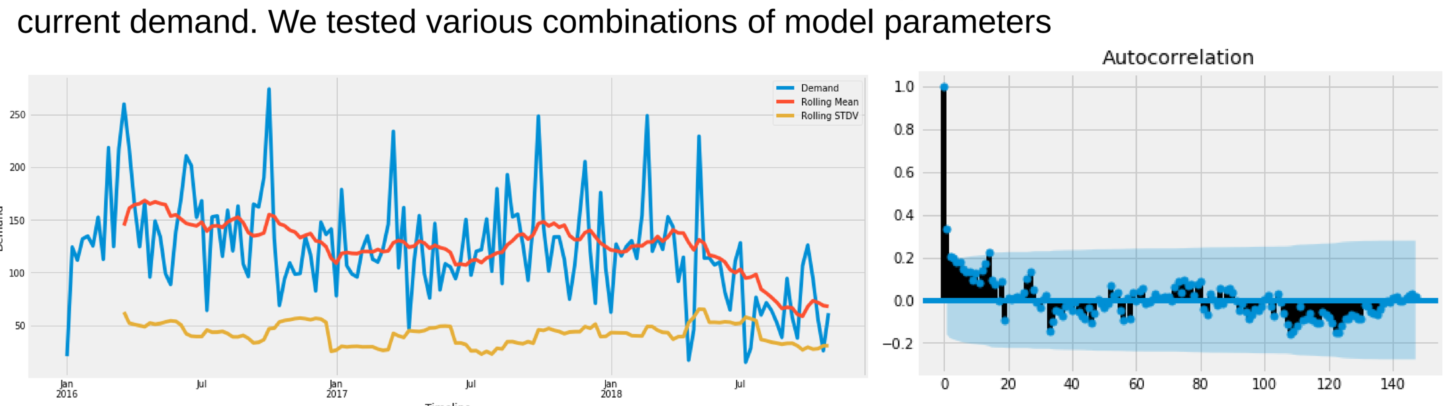


Figure 2.b Time series plot of demand, ACF plot

Methodology (Approach) Selection : We noticed that the demand follows a seasonal pattern for most of the products. We used two different forecasting approaches for predicting the demand

ARIMA Modelling:

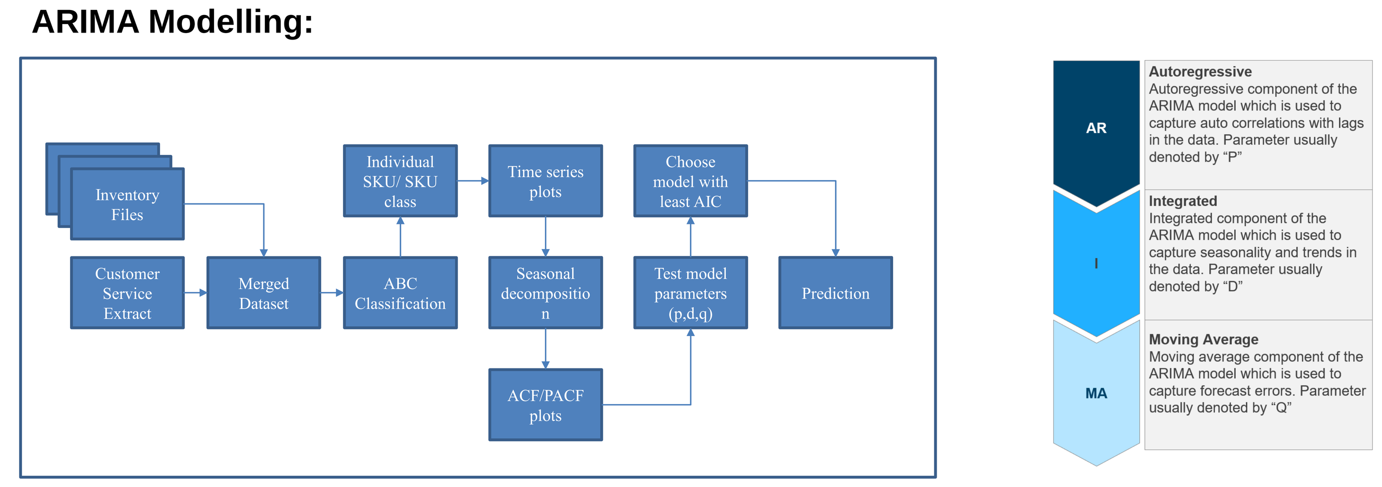
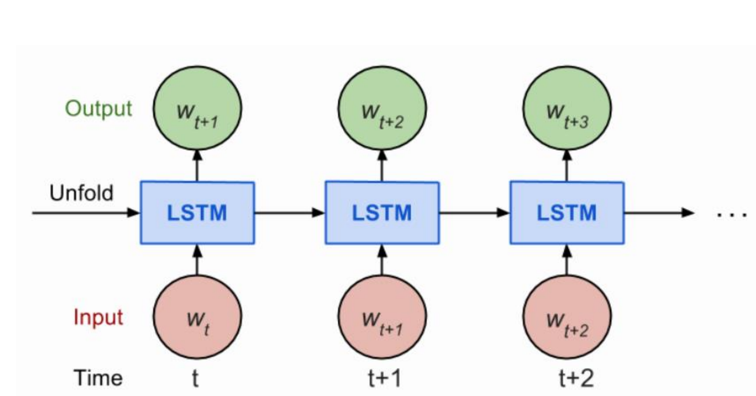


Figure 3 Time-series analysis and ARIMA modelling

ARIMA forecasting involves estimating the model parameters (p,d,q) for each product using the minimum AIC approach. Additional model terms (P,D,Q) are added for products displaying seasonality in the demand. The forecasts are then compared with actual demand and the model performance is evaluated using the RMSE (root mean squared error)

LSTM Modelling:



Long short-term memory (LSTM) units are units of a recurrent neural network (RNN). An RNN composed of LSTM units is often called an LSTM network (or just LSTM). A common LSTM unit is composed of a cell, an input gate, an output gate, and a forget gate. The cell remembers values over arbitrary time intervals and the three gates regulate the flow of information into and out of the cell.

Results

We compared the forecasts produced by LSTM, Naïve, and ARIMA over a six month period from January to June. Visually, ARIMA and Naïve forecasts appeared to closely follow the demand fluctuations whereas LSTM provided a relatively stable forecast

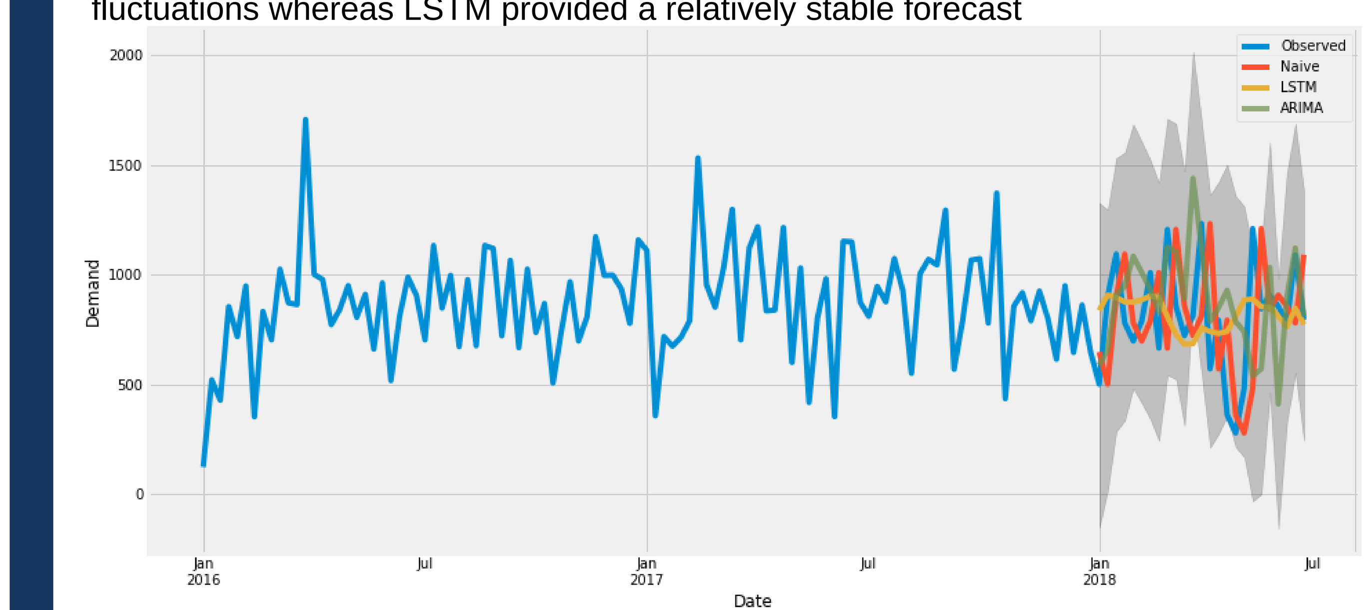


Figure 4. Model Comparison

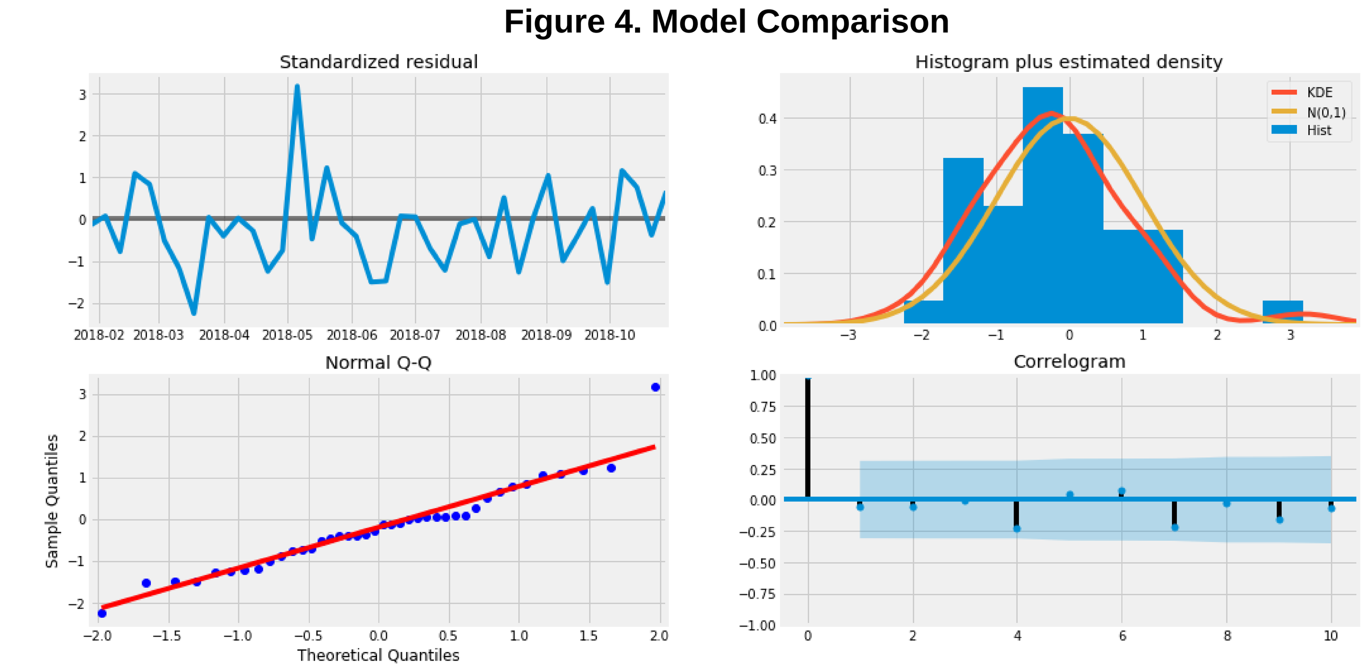


Figure 5. Residual plots for ARIMA models

The residual plots for the ARIMA model indicated that the residuals are normally distributed and the correlogram indicates that there is no correlation left in the residuals that can be extracted. This indicated that the chosen ARIMA parameters performed reasonably well

Category	CPG	ARIMA	LSTM	LSTM+ARIMA
A	84%	-14%	-19%	-29%
B	48%	-2%	-14%	-25%
C	-10%	-17%	-22%	-30%

Error relative to Naïve forecast: Low / Moderate / High

Figure 6. Model comparison with respect to naïve forecasts

We compared the performances of forecasts produced by CPG firm, ARIMA, LSTM, and LSTM+ARIMA (weighted average of the two models) to find out how well these models performed relative to the naïve forecasting methods. We found that ARIMA+LSTM model performed significantly better than the standalone ARIMA, LSTM, or CPG forecast models. Therefore, it is recommended that the CPG firm review their forecasting methodology particularly for the "A" and "B" category SKUs. We believe that while standalone forecasting approaches may be efficient for these categories, considerably better forecasts can be achieved using a combination of different methods such as LSTM and ARIMA

Conclusions

Distribution Center stockouts can be a nightmare for firms particularly for businesses that have large number of SKUs. At times, it may not be computationally feasible for firms to identify model parameters for each and every SKU. In this research, we noted the importance of product classification using demand and supply patterns for identifying parameters for forecasting models. We also found that combining LSTM and ARIMA model forecasts can lead to substantially more accurate predictions than standalone models.

For future work, we can investigate combinations of different models such as Holt's Winter, exponential smoothing, LSTM, ARIMA/SARIMA using varying weights and assess the accuracy of prediction from such combined models

Acknowledgements

We thank our industry partner and professor Matthew Lanham for constant guidance on this project.