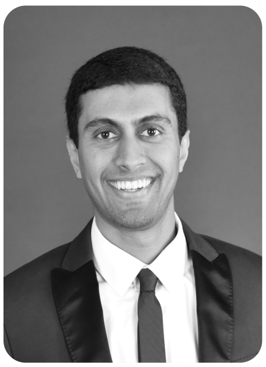
Enhancing S&OP Performance with Analytics

By: Minhaaj Khan and Srideepti Kidambi
Advisor: Dr. Tugba Efendigil

Topic Areas: Risk Management, Forecasting, Predictive Analytics

Summary: This study focused on predicting risks in the Sales and Operations Planning (S&OP) of a Consumer Packaged Goods (CPG) company. Using S&OP data for a specific brand, the objective was to build models that could accurately predict and mitigate risks in the plan, allowing the business to improve S&OP forecast accuracy. One model proved viable and helped improve profit margin for the studied brand by 3.8% over a three-month period. Extrapolating this methodology to all brands within the company can increase gross profits by $17MM annually.



Minhaaj Khan has over 10 years of consulting and project experience in the realm of Supply Chain Management. He has a Bachelor's in Operations Research & Management Sciences from UC Berkeley. He will join a startup in the automotive industry as the VP of Strategy & Operations upon graduation.

Deepti Kidambi has over 10 years of experience in Supply Chain Transformation & Operational Expansion with Schlumberger, Accenture Strategy & Rent the Runway. Deepti has a Bachelor's in Mechanical Engineering from BITS, Pilani, India. She will join Google as a Technical Program Manager upon graduation.

KEY INSIGHTS

1. With appropriate business data, supervised classification models effectively predict risks in the S&OP plan.
2. Supervised classification models applied towards S&OP risk mitigation have the potential to deliver substantial improvement in forecast accuracy and gross profit.
3. By capturing planning data and acquiring the knowledge necessary to leverage predictive analytics, companies can drive a large increase in profit and gain significant competitive advantage.

Introduction

The S&OP process delivers a consensus forecast that helps businesses better align different functions of their organization towards a common goal. Despite the importance of this process, businesses lack the ability to assess risks in their S&OP plan. This inability leads to suboptimal planning, lower forecast accuracy and reduced profits.

In this project, predictive analytics techniques were applied to 1) identify patterns of risks in the S&OP process and 2) mitigate those risks to improve consensus forecast accuracy for a CPG company that sells health and nutrition products. This project focused on a protein bar brand that dealt with $7MM in obsolescence due to missed risks in its S&OP plan. Research on the application of predictive analytics methodologies for risk assessment in S&OP plans is scarce. The focus of this study was to answer the following three questions:

1. Can predictive analytics models effectively predict high-probability risk patterns in the S&OP plan?
2. How much can these models improve consensus forecast accuracy and what is the financial impact of this improvement?
3. What factors are important to the success of other CPG companies that want to pursue a similar risk assessment methodology in their S&OP plan?

Operational Context

Consumer packaged goods (CPG) are products that sell quickly and at relatively low cost. CPG products have a short shelf-life and obsolescence is a common risk when supply isn't correctly matched with demand. This misalignment of supply and demand is more likely in the case of the sponsoring company, as they must make supply decisions three months in advance of demand for their contract manufacturers to deliver products on time. Most companies attempt to mitigate this risk by carrying surplus inventory, but when dealing with short shelf-life products, unrealized demand leads to high levels of obsolescence and other supply chain issues.

Data and Methods

To effectively predict risks in the S&OP plan, the methodology laid out in Figure 1 was followed. The project began with understanding the business context and identifying data requirements; relevant data was then extracted and scrubbed; four different risk outcomes were chosen for analysis; five different data mining classification techniques were applied; and finally, viable models were chosen for each risk outcome and evaluated for accuracy improvement and financial benefit.

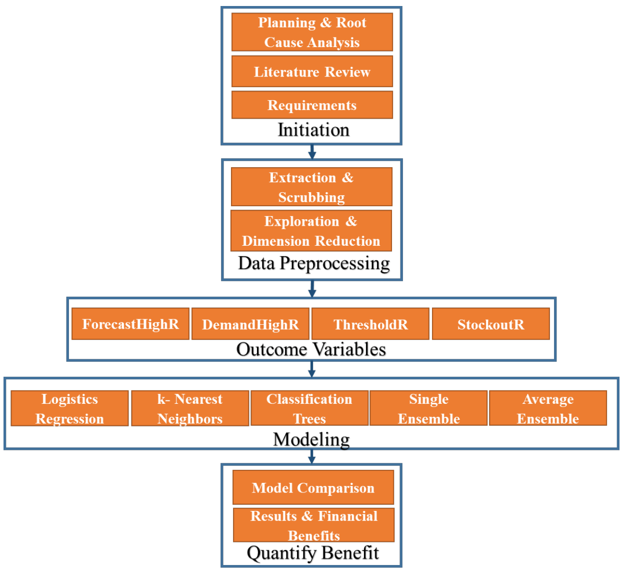


Figure 1: Methodology of Study

<u>*Initiation:*</u> Root cause analysis utilizing a fish-bone diagram identified the S&OP process as an area of target for risk assessment using predictive analytics. As product brand categories shared the same supply chain, a single brand was chosen for this study.

<u>*Data Processing & Variable Selection:*</u> All data attributes were manually extracted from S&OP Excel files provided by the sponsoring company. This data existed in weekly buckets and was scrubbed of missing data, leaving 2,477 records for analysis. As S&OP risks are categorical in nature, continuous variables provided by the business were used to generate four categorical outcome variables that classified risks in the models. The four categorical risk outcome variables used are summarized below:

1. ForecastHighR – 50% over forecast risk, forecast greater than 100 cases.
2. DemandHighR – 50% under forecast risk, forecast greater than 100 cases.
3. ThresholdR – inventory weeks of supply less than 4 weeks.
4. StockoutR – 1 if total demand in a week greater than total supply across the entire supply network.

Data was portioned into a 60:40 split for training and validation. Dimensionality reduction was done with a predictor correlation heat map. Further reduction of predictors was achieved for each risk outcome utilizing variable importance plots and by selecting coefficients with low p-values in logistics regression models. As an example, for predicting ForecastHighR, 14 predictors were reduced to four – (1) coefficient of variation last 8 weeks, (2) mean absolute deviation last 8 weeks, (3) minimum total quantity ordered last 8 weeks and (4) the consensus forecast.

<u>*Model Selection and Build:*</u> Shmueli et al. (2018) describe supervised learning as, "the process of providing an algorithm (regression tree, etc.) with records in which an output variable of interest is known and the algorithm "learns" how to predict this value with new records where the output is unknown." This fits in exactly with the objective of this study, where S&OP history with known output variables was used to predict risks in future plans where the risk is unknown and must be determined.

This study applied k-Nearest Neighbors, Classification Trees, Logistics Regression and Ensembles. Application of data mining algorithms was an iterative process, attempting multiple variants, and often choosing different variables or settings within the algorithm.

In selecting the best models for risk classification, confusion matrices, lift charts, and decile-wise lift charts were used to interpret the results. The most important aspect of this study was to evaluate the ability of each model to identify high probability risks correctly (i.e., Sensitivity). Therefore, only models with higher sensitivities were considered. P-value and accuracy were also used to determine the potential viability of models.

Results and Limitations

Supervised classification algorithms used to assess S&OP risks achieved mixed levels of success. One of the key limitations for this study was the unavailability of promotion data in a usable format. Promotions drive demand fluctuations in the CPG industry and DemandHighR, ThresholdR, and StockoutR risks weren't accurately predicted due to the unavailability of promotional data. ForecastHighR was the only risk outcome where a model was determined to be viable (Figure 2).

The average ensemble methodology, using logistics regression, k-NN and classification trees, proved most effective in identifying ForecastHighR risks in the S&OP plan. In fact, when applied against a separate subset of data outside the training and validation set (i.e., 12-week test set February-April 2018), model performance showed better accuracy than the validation set. This indicates that the Average Ensemble model was not over fit.

Forecast Accuracy

Models	ForecastHighR	DemandHighR	ThresholdR	StockoutR
(1)Logistic Regression	79.92%	82.53%	88.71%	72.51%
(2)k-nearest neighbors	66.20%	85.25%	89.37%	76.82%
(3)Classification Tree	74.28%	81.41%	87.93%	74.66%
(4)Single Ensemble	78.48%			
(5)Average Ensemble (Models 1, 2 & 3)	**78.87%**			

p-Value

Models	ForecastHighR	DemandHighR	ThresholdR	StockoutR
(1)Logistic Regression	<2e-16	0.017	0.574	0.584
(2)k-nearest neighbors	<2e-16	5.70E-06	0.348	0.007
(3)Classification Tree	1.819E-12	0.109	0.807	0.132
(4)Single Ensemble	<2e-16			
(5)Average Ensemble (Models 1, 2 & 3)	**<2e-16**			

Sensitivity

Models	ForecastHighR	DemandHighR	ThresholdR	StockoutR
(1)Logistic Regression	75.30%	28.50%	8.20%	6.90%
(2)k-nearest neighbors	66.20%	63.00%	15.30%	34.70%
(3)Classification Tree	74.90%	35.00%	10.60%	29.70%
(4)Single Ensemble	75.60%			
(5)Average Ensemble (Models 1, 2 & 3)	**79.40%**			

Note: **Bold** designates viable model values

Figure 2: Model Performance Metrics on Validation Data

Average Ensemble model performance also showed promise in instances where the model misclassified ForecastHighR=1 when the forecast happened to be less than 50% over forecast (i.e., ForecasHighR=0). 71% of misclassifications still had forecast exceed demand by a large margin (near, but below 50%), allowing the model to capture improvement in accuracy and reduction in bias through risk mitigation.

Accuracy over the 12-week period from February-April 2018 improved by 5.7% as shown in Figure 3. Bias in that same period went from 3.7% over forecast, down to -0.3%. This was a significant improvement in three-month lag SKU/Week accuracy.

To quantify financial benefit, we used the findings from an AMR Research study conducted in 2008. This leading research firm in global supply chain best practices stated that a 3% increase in forecast accuracy increases profit margin by 2%. Given the 5.7% improvement in accuracy over the 12-week period, the finance department for the sponsoring company extrapolated a $1.8MM annual increase in gross profit through the application of the ForecastHighR average ensemble model. Under the assumption that the remaining brands in the business (93% of the revenue) would show a more modest 3.5% improvement in S&OP forecast accuracy, an additional $15MM in annual gross profit could be captured by the business. This is under the assumption that the business manages to apply key missing predictors (i.e., promotions) so that all risk models, not just one, are viable for business application.

Improvement in Model Accuracy & Bias

Accuracy	Feb	Mar	Apr	Total
Baseline	50.4%	55.3%	44.8%	50.0%
Predictive Model	54.1%	57.2%	55.8%	55.7%
Improvement	3.6%	1.9%	10.9%	**5.7%**

Bias	Feb	Mar	Apr	Total
Baseline	-1.0%	2.4%	8.5%	3.7%
Predictive Model	-3.9%	-1.0%	3.4%	**-0.3%**

Figure 3: ForecastHighR Average Ensemble Model Benefits

Conclusion

This study showcases the potential of predictive analytics to successfully capture business risks in the S&OP plan despite missing key predictors and without big data. This is especially true if business levers that drive these risks are available as predictors in the model, as was the case with the 50% over forecast risk. This finding should encourage companies hesitant on pursuing predictive analytics to search for value in any company data that stretches far enough back to capture risk patterns but doesn't fit the big data mold.

However, where outcomes are significantly impacted by predictor data that is unavailable or where recent demand patterns are insufficient in predicting risks that are dependent on demand three months into the future, supervised classification models show poor accuracy. This necessitates further effort from the sponsoring company to capture key planning inputs in a format suitable for data mining applications.

The key to successful implementation of the methodology described in this study begins with the buy-in from top-level management. This is important, as the S&OP forecast operates as a negotiation across different functions of the business to match supply and demand; making risk-based adjustments to an agreed upon plan will require all functions of the business to trust and align with this process. Companies that improve the quality of their data and acquire the knowledge necessary to leverage predictive analytics, not only for S&OP risk assessment but also for wider supply chain decision-making, can better plan for what to make, when, and for whom, to drive large increases in profit and gain significant competitive advantage.