Comparison and Financial Assessment of Demand Forecasting Methodologies for Seasonal CPGs

By Burak Gundogdu and Jeffrey Maloney
Advisor: Dr. Tugba Efendigil

Topic Areas: Machine Learning, Forecasting, Inventory

Summary: In this research, the performance of the statistical forecast of our sponsor company, King's Hawaiian, is compared to machine learning models. Evaluating multiple machine learning models across different parameters and data, a model that reduced the forecast error was identified. Translating the change in forecast error to reductions in safety stock, the inventory savings were found to be sufficient to offset the incremental costs associated with more advanced analytics. Our results show that machine learning adds value for seasonal CPGs.

Prior to MIT, Burak worked as a management consultant at The Boston Consulting Group. He holds BS degrees in Economics and Mathematics from Middle East Technical University and MBA from the University of Notre Dame.

Prior to MIT, Jeff worked in the supply chain for King's Hawaiian for five years. His experience in supply chain planning and analytics led him to MIT. Jeff holds a BS in Biology from the University of California, Irvine.

KEY INSIGHTS

1. **Machine learning forecasts do not universally outperform traditional statistical methods, requiring a diligent selection and validation process of a model before adoption.**
2. **Including additional features and external data can improve forecast accuracy and allows for additional business insights drawn from feature selection.**
3. **Improvements in forecast error can offset incremental costs associated with increased data and advanced analytics.**

Introduction

For businesses that operate from a demand forecast, achieving a low forecast error can confer a significant financial advantage. High forecast error requires greater investment in safety stock inventory to cover demand variation and can lead to supply chain inefficiencies. It can also lead to missed sales and be detrimental to customer relationships and ultimately damage the brand. Forecast volatility causes a bullwhip effect upstream to manufacturing, which can result in operational inefficiencies.

Improving forecast accuracy is a challenge that many companies face. Accuracy is especially important for consumer packaged goods (CPG) companies that are built-to-stock. As a highly seasonal build-to-stock business, King's Hawaiian must rely on a forecast to build inventory sufficient to cover peak demand. They must balance investments in finished goods to cover forecasted demand while minimizing their financial exposure. Build-to-stock companies that have strong seasonal demand must begin planning and building inventory early, and with less information.

Advancements in forecasting software and analytics have provided modern businesses with many options. The promise of advanced forecasting methods is a more accurate forecast that will yield financial savings to the firm. Advanced forecasting techniques require an investment in not just the software, but the increased amounts of data required to run the model. Companies must also account for the cost of the increasingly sophisticated personnel responsible for maintaining an advanced demand forecasting solution.

Companies like King's Hawaiian need to understand whether the potential savings will offset the incremental costs of a machine learning forecast solution. The development of a framework that can guide the selection process based on potential financial impact can help.

Methodology

Five different machine learning models were evaluated: support vector regression (SVR), artificial neural network (MLP), random forest (RF), gradient boosting (GB), and k-nearest neighbor regression (KNN). The performances were evaluated in a three-dimensional cube search by varying hyper-parameters and feature sets. To augment the machine

learning performance, additional features (consumption, socio-economic, and severe weather data) were incorporated. Shipment and consumption demand were evaluated in both seasonal (S) and de-seasonalized (D) formats.
The accuracies of the models were compared to the current statistical model and the change in projected safety stock was calculated. The resulting inventory savings were then compared against the incremental costs of adopting a machine learning demand forecast.

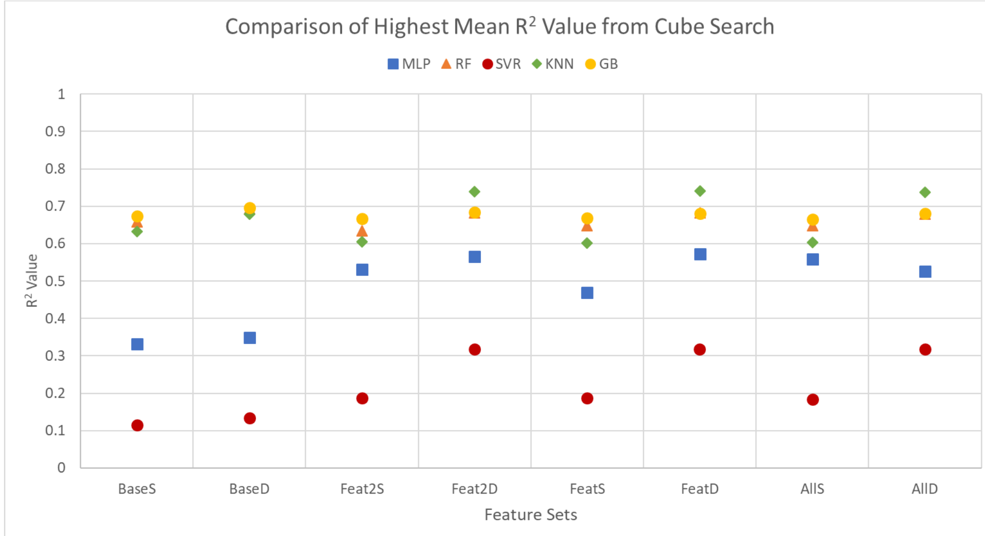

Figure 1: Comparison of Error from Cube Search

Results

A comparison of the performance of all models run through the cube search process indicated that a KNN model was the best fit for predicting shipments when running on a machine learning generated de-seasonalized feature set. The selected KNN model scored the highest mean R^2 value on cross-validation during the cube search, while maintaining a low Train-Test variance. The selected model was tested on the unseen year 2018 data set. The machine learning model achieved a lower total forecast error (WAPE) than the current King's Hawaiian statistical forecast model, which translates to financial savings due to lower safety stock inventory levels required to cover demand variance.

Model, Features, and Hyperparameters Analysis

Machine learning models were first compared using the hyper-parameters that yielded the highest mean R^2 for every feature set evaluated. Figure 1 compares, for each feature set, the highest mean R^2 achieved during cube search of each model from the hyper-parameters evaluated.

Based on the comparison, KNN consistently presented the highest R^2 value when additional features were added above baseline and the demand was de-seasonalized. While KNN scored the highest mean R^2 value for de-seasonalized feature sets, it was outperformed by GB for seasonal feature sets. In general, the highest mean R^2 values came from models run on de-seasonalized feature sets. The addition of features tended to improve mean R^2 for most models above the Baseline feature set. SVR was the worst performing model, while also requiring the longest run time.
The Train-Test variance (the absolute difference between mean R^2 scores on test and train data sets during cross-validation) was also evaluated for all models with the highest mean R^2 values. Models with lower deltas between test and train scores are considered less likely to be overfit. As shown in Figure 2, a comparison of the mean R^2 values and the Train-Test variance helped resolve the differences between models with similar mean R^2 values and aided in final model selection.

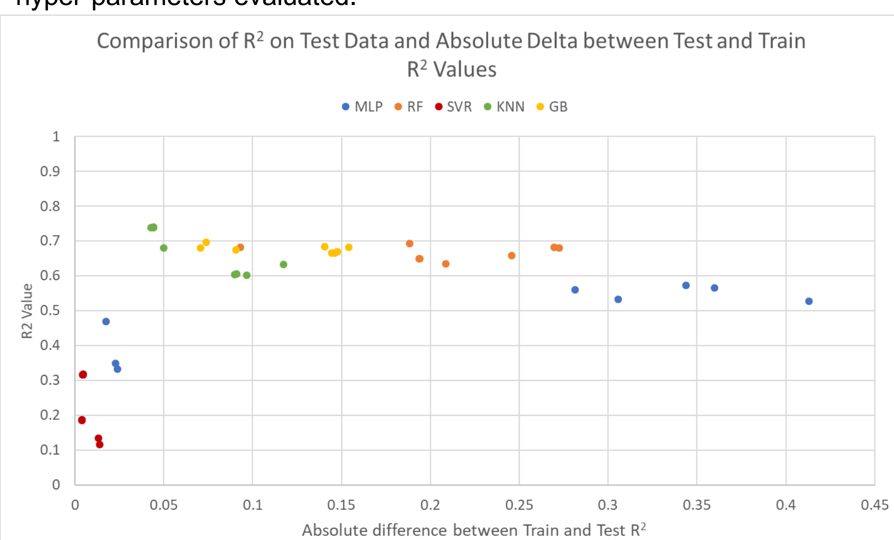

Figure 2: Comparison of Train-Test Variance to mean R2 values.

The selected KNN model achieved the best results when running on the de-seasonalized feature set with the hyper-parameters set, as determined from the search space, to 15 neighbors and the Manhattan distance calculation. The selected KNN model achieved a mean R^2 of 0.74 on cross-validation during the cube search. The train time ranged from 0.79 seconds to 4.5 seconds, while the score time range from 47.9 seconds to 153.2 seconds. The KNN achieved an R^2 of 0.67 on the out of sample 2018 data.

Forecast Error Analysis

The current King's Hawaiian statistical model utilizes a Holt-Winters process that is trained on data from 2012. The model uses the same aggregate product categories for limited time offer products and product refreshes as used in this analysis. King's Hawaiian updates its model monthly as part of its S&OP process. The statistical forecast is run at the national level and disaggregated to the regional 3PL network level. For an equivalent comparison of regional forecast error between the models, actual shipments, King's Hawaiian regional forecast, and the KNN model forecast were all aggregated to five primary distribution regions that align to the principle 3PL service regions in the King's Hawaiian network.

The performance of the 2018 statistical model evaluated on the same product mix, geographic resolution, and timeframe, had 34.4% WAPE annually. The selected KNN model running on the de-seasonalized feature set achieved a WAPE of 30.5%, 3.9% lower than the statistical model.

Inventory and Financial Analysis

King's Hawaiian's regional forecast is at the weekly resolution; however, for calculating statistical safety stock they use the root of the mean of the weekly forecast error squared (RMSE) for each region for each month. To evaluate the impact on safety stock, the RMSE of the KNN model was calculated at a monthly resolution to match the current safety stock methodology. The RMSE is then scaled by the corresponding standard deviation value associated with King's Hawaiian's 99.5% customer service level. Using the forecast accuracy of King's Hawaiian current statistical methodology, the average periodic pounds of the safety stock for the sum of all regional geographies and products selected equates to 6.5% of the total annual demand for 2018. The same value using the KNN model's forecast error is 5.8% (Figure 3).

The machine learning forecast error represents a decrease in the total level of safety stock of 10% as a direct result from the increase in forecast accuracy of 3.9%. Other associated positive financial impacts from decreases in the forecast error can be expected as well.

Increased costs are incurred by running a machine learning model over a statistical model. Approximately 30-40% of the incremental cost is for the collection of data, and 10-20% for the incremental compensation for employees with higher technical specialization to initialize and maintain the models. The remainder of the cost comes from the hosting and licensing fees associated with commercially available demand forecasting systems.

The KNN model's lower forecast error results in a projected reduction in the value of safety stock carried of >$900k. The savings exceed the annual incremental costs of ~$500k in data, software and personnel support required to maintain the KNN model, justifying the pursuit of the more advanced forecasting methodology.

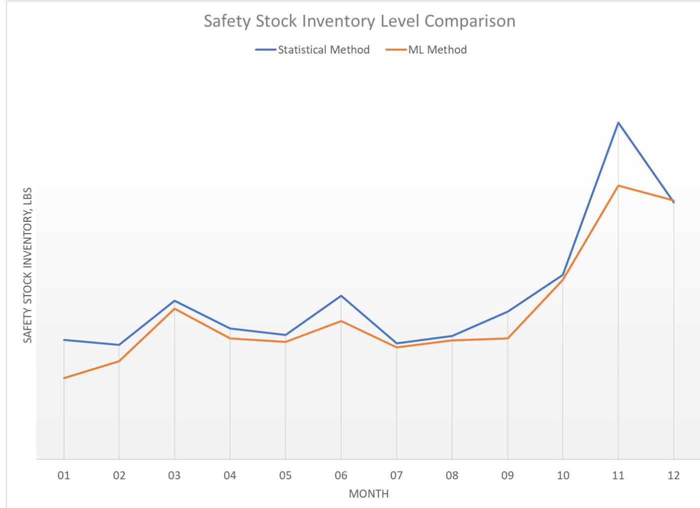

Figure 3: Comparison of Safety Stock Levels for Statistical vs Machine Learning.

Conclusions

We believe the results of this research help frame the impact that more advanced analytical techniques can have at King's Hawaiian or similarly seasonal CPG businesses. The model selection process identified models that were good predictors of the demand and those that were not, demonstrating that machine learning models are not universally better than traditional statistical methods. The variable feature sets showed that forecast accuracy tended to improve with the addition of relevant attributes from external sources. The improved demand forecast accuracy, achieved with the KNN model, had a beneficial impact on inventory, resulting in decreased safety stock. The inventory benefits demonstrate that the incremental costs associated with a more complex analytical technique, and the data required to run them, can be overcome by savings in other areas of the supply chain.