Sales Forecasting Using High-Performance Computing

Sowmya Bhatraju, Pavan Ghantasala, Yu Lin Tai, Ashutosh Porwal, Swati Srivastava, Vineet Suhas Soni, Yang Wang
Purdue University, Krannert School of Management, 403 W. State Street, West Lafayette, IN 47907
sbhatraj@purdue.edu; nghantas@purdue.edu; tai21@purdue.edu; porwal@purdue.edu; srivas98 @purdue.edu; vsoni@purdue.edu; yangwang@purdue.edu

Abstract

"Sales Forecasting Using High-Performance Computing" paper is intended to improve inventory management and assortment planning of a national retailer through reliable and efficient sales forecasting. It is focused on building upon the current "bottoms-up approach" model used by the company, by creating a robust regression model with optimal interaction terms, feature engineering, and hyper-parameter tuning using high-performance computing. Exploratory data analysis was performed on three categories of data- batteries, brakes, and filters, which provided information about demographics of customers, the geography of the store regions, gross revenue, and quantity sold for the past two years at a unique SKU-Store combination. Data aggregation and other transformations were done as per business requirements and for ease of modeling. Advanced programming was performed using Bell Cluster and Dask machine learning libraries. Several feature engineering experiments were carried out to create 2nd-degree interaction terms, PCA analysis, etc. The response variable was predicted by building several models such as linear regression, lasso regression, and OLS. The final regression model selected has the highest Adjusted R-square and lease root mean square error, which were the two metrics defined for model selection. This model had the highest interpretability and lowest run time.

Keywords: High-performance computing, sales forecasting, feature engineering, PCA, machine learning

Sales Forecasting Using High-Performance Computing

"The future belongs to those who prepare for it today" (Malcolm X, 1962). This phrase holds for all the companies who aspire to generate promising sales outcomes for their organizations in these uncertain times. Sales forecasting is a vital process to achieve those outcomes that predict an organization's growth and overall health. According to a Forbes article, "It's a tough job, but without a sales forecast, the management team doesn't know whether to conserve cash aggressively or continue current operations. It's not an easy problem to solve, but some insights are possible." (Conerly, 2020, para. 1). An organization's ability to accurately forecast sales is crucial to sustain in the market, competing with its strategic vision, and operating effectively. In addition, it helps to reduce inventory management costs through better sales predictions, better inventory allocation across stores, and proper assortment planning. Accurate sales forecasting is yet a significant challenge for most companies. As per a newspaper published on WSJ, "It's been an adventure over the past year," said Hayley Berg, head of price intelligence at travel-booking app Hopper Inc. "It's really a challenging forecasting problem," she said, referring to the abrupt halt in consumer and business travel last year (Loten, 2021, para. 9). In the absence of a precise sales forecast, organizations can harm their estimation over a long-term duration. At a Gartner press release, "Heads of sales operations are under constant pressure to produce accurate forecasts to help shape decision making," said Craig Riley, senior principal analyst in Gartner's Sales Practice. "Unfortunately, sales forecasting isn't getting easier — as expanding product portfolios and shifting market conditions exacerbate the issue." (Blum, 2020, para. 2). All in all, sales forecasting has the potential to make or break a business. This paper revolves around two primary research questions.

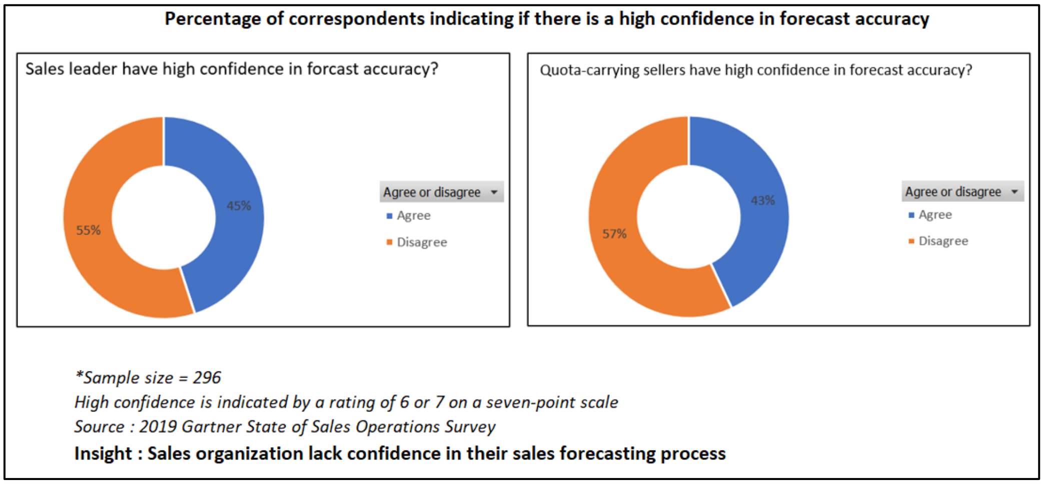
How to forecast sales using different linear regression techniques, including interaction terms and feature engineering?

Sales can be forecasted by identifying predictor input variables for the machine learning model, then manipulating and transforming input variables into features that can boost the models' prediction accuracy, performance, and efficiency. There is an interaction among the predictor variables at times, indicating a relation between input variables that impact the response. Therefore, poor feature engineering or the absence of acknowledgment of interaction among variables in the sales forecasting model can harm the model.

How would using HPC capabilities eventually allow us to find an optimal robust linear model that performs much better than trying a generic backward selection type approach?

High-performance computing (HPC) systems can solve complex computational problems swiftly through enhanced processing capability and storage capacity. It can help create robust models taking advantage of vast volumes of data with higher computing and performance specifications, which can't be done through mere feature selection. It utilizes dense computer clusters that sync with one another, allowing a model to run on optimized hardware with advanced processing capabilities. With such implications, a model with optimal interaction terms, feature engineering, and hyper-parameter tuning using High-Performance Computing can perform sales forecasting with higher interpretability and lower run time.

The remainder of this paper is organized as follows: The related work elaborates on the literature on various methods used to date. We delve deeper into understanding the current methodologies and understand how our analysis caters to the gap in the existing literature. Data analysis entails the exploratory and descriptive analysis performed on the store level SKUs in the retailer store. The proposed methodology is presented in further sections, and the criteria formulation is discussed along with its performance against various models. Finally, section 6 concludes the paper with a discussion of the implications of this study, future research directions, and concluding remarks.



**Sample size = 296*
High confidence is indicated by a rating of 6 or 7 on a seven-point scale
Source : 2019 Gartner State of Sales Operations Survey
Insight : Sales organization lack confidence in their sales forecasting process

Fig 1. %age of correspondents indicating if there is a high confidence in forecast accuracy

Dealing with high dimensionality and computation issues

Sales forecasting is the process of predicting the number of sales a firm will generate within a specific timeframe such as a week, month, quarter, or year. Creating a predictive model to forecast sales with high dimensionality issues requires careful consideration so that model computation time is within an acceptable time frame. Sales forecast modeling with high dimensionality concerns in the retail domain can be approached differently. Multiple variable selection methods in cases of many explanatory variables can be combined, and the union of selected variables from those methods can be taken to reduce the risk of eliminating variables having a significant impact on the response variable (Stock and Watson, 2004). PCA can be performed to deal with high dimensionality, but it runs the risk of eliminating some features that might be strong predictors of the dependent variable. A way around this is to categorize variables in broad groups according to their common attributes and run PCA on each group separately to reduce the likelihood of missing out on important predictor variables (Ma et al., 2015). An evolutionary algorithm is a less-known concept of feature selection in high-dimensional data. The principles of human evolution inspire evolutionary algorithms in that new candidate solutions are created by transforming past best-performing candidates. Evolutionary algorithms, although practical, in the case of high dimensional data demand high computing power. Therefore, high-performance computing is employed to leverage evolutionary algorithms for feature selection in high-dimensional datasets (Alwadei, Dahab, and Kamel, 2017).

There is substantial literature in causal forecasting assuming a functional form for the effect of different explanatory variables on sales. Fader and Hardie (1996) argued that consumer choices and demand for SKUs are not formed at an individual level. Instead, a set of discrete attributes like package, size, and type can be used to characterize their future demand. Marketing Literature has augmented the sales prediction models, including promotional activities as predictors, e.g., Discounts in (Cooper et al., 1999), frequency of promotions for similar products (Christen et al., 1997), category or product group in (Narasimhan, Neslin & Sen, 1996). Trapero, Kourentzes, and Fildes (2014) have proposed novel promotional modeling combining PCA and Regression analysis at the SKU level. Retail studies for demand estimation took a different direction historically dominated by consumer choice models with a focus on likelihood estimation and expectation maximization algorithms using Multinomial Logit Models and Utility functions (Vulcano, Ryzin, and Ratliff, 2011; Kok and Fischer, 2007; Anipudi, Dada, and Gupta, 1998). Most of these models are based on modeling stock-out event data (Flides, Ma, and Kolassa, 2019). However, there is evidence that demand depends on the inventory level for at least some products, with higher inventory levels leading to higher sales called the "billboard effect" (Koschat, 2008; Ton & Raman, 2010; Flides, Ma, and Kolassa, 2019).

Machine Learning Models such as Decision Trees, SVM or Neural Networks do not assume such direct functional form relationships. Their model complexity also means computational expense in a Big Data Environment. Linear Regression has been the optimal choice for sales prediction for practitioners finding the balance between interpretability, time complexity, and accuracy. Linear Regression is easy, simple, and feasible to fit at SKU level high dimensional space. Several approaches to capture non-linearities like multiplicative interactions log and exponential transformations can be applied to variable space (Flides, Ma, and Kolassa, 2019). Divakar et al. (2005) designed a dynamic regression model to capture the effect of past sales, trends, and regional factors.

Data Description

Data used in the project has been received by a national retailer and is primarily divided into three categories- filters, brakes and batteries. Fields are consistent across the three categories with over 80 predictors. Each record refers to a unique SKU-Store combination. More than 15 features describing the demographics and geography (population density, road quality index, avg. age, percentage of blue collar, median household income, vehicles in operation in the region etc.) of the store regions have been included. Features pertaining to sales (gross revenue, net revenue, quantity sold etc.) made in the past two years have also been considered. Stores have been clustered into various groups based on sales performance and location. Average performance (avg. cluster unit sales, avg. cluster lost sales etc.) for clusters have also been used as input. Other features like missed sales, lost quantity of an item, no. of times an SKU is looked up in the catalogue and life cycle of the SKU have also been used as predictors.
Separate tables for inventory, sales, SKUs and stores have also been received. The data dictionary for these tables goes as under:

Table 3
Inventory

Variable	Type	Description
storenum	Numeric	Store number
skunum	Numeric	SKU number
inv_year	Numeric	Year of inventory
inv_period	Numeric	Period of inventory
wk_of_period	Numeric	Week of period
wk_of_year	Numeric	Week of year

smaxi	Numeric	maximum if smaxi >0 then sku is stocked at the store
scoh	Numeric	quantity of units on hand in store
wk_ending_dt	Date	Week ending date
merchandise_group_desc	Text	Merchandise group decsription(batteries etc.)

Table 4
Sales

Variable	Type	Description
store_number	Numeric	Store number
sku_number	Numeric	SKU number
customer_type	Text	Type of customer (DIFM)
qty_sold	Numeric	Quantity of SKUs sold
gross_sales	Numeric	Gross sales for the fiscal year
sales_cost	Numeric	Cost to company for the fiscal year
fiscal_year	Date	Fiscal year
fiscal_period	Numeric	Fiscal date

Table 5
SKU

Variable	Type	Description
sku_number	Numeric	Store number
mpog_description	Text	Master Plan of Gram description
sku_description	Text	Description of SKU
length	Numeric	Length of SKU
width	Numeric	Width of SKU
height	Numeric	Height of SKU
cubic_inches	Numeric	Volume of SKU
retail_price	Numeric	Retail price of SKU
unit_cost	Numeric	Cost per unit
min_app	Numeric	No. of applications needed to complete the project
discontinued_flg	Categorical	SKU discontinued or not
stocking_location	Text	Location of the stock
epc_sku_type_description	Text	Type of SKU
store_acquisition_cost	Numeric	Store acquisition cost

Table 6
Store

Variable	Type	Description
store_number	Numeric	Store number

dma_id	Numeric	Designated Market Area ID
sector_id	Numeric	Sector ID
market_class	Categorical	Market class
latitude	Numeric	Latitude of store
longitude	Numeric	Longitude of store
platform_cluster_name	Categorical	Platform cluster name
platform_cluster_name_py	Categorical	Platform cluster name previous year
open_date	Date	Store opening date
market_share	Numeric	Market share
capture_rate	Numeric	Ratio of sales in per store, sector and sales per Store
rolling_capture_rate	Numeric	Rolling capture rate

Methodology

The sales forecasting approach is formulated as a regression problem with the optimal prediction built on the performance benchmarks of RMSE. Since such an optimal solution is highly dependent on computational intricacies, the linear regression model's potential is maximized by using parallel computing and the high-performance abilities of enhanced storage and performance(increasing nodes). The overall methodology for regression is divided into four parts: exploratory data analysis, data pre-processing, feature extraction and feature engineering, model building, and performance benchmarking.

The sales data enlists SKU and store-level information documented as monthly aggregates and yearly trends. The SKUs under consideration belong to three categories - brakes, filters, and batteries, which constitute the bulk of the sales for the retailer. An exploratory data analysis was aimed at summarizing the totality of data. The purpose was to extract maximum insights from the data to discover patterns and anomalies. Moreover, it was also a step for investigating potential interactions and their significance in the regression problem.

Out of the 12121 unique SKUs identified, 40% belonged to the heavy-duty application-specific. Regarding profitability, the brake pads brought in maximum revenue, whereas break shoes incurred fewer profit margins to the firm. There were 1207 SKUS discontinued across stores, and negative sales implied imputation requirements. Therefore, this step set the foundation for specific pre-processing, such as scaling, imputations, and outlier handling to improve prediction capability.

The data pre-processing stage dealt with the following to decrease unwanted information in the data - data reduction, transformation, and determination of potential interactions between variables. The data reduction was performed to eliminate low business impact variables, null and missing occurrences in the dataset: all current year sales-related data, part types, cluster, adjusted vehicles in operation, and failure-specific information were removed. Since the actual sales from the current year bring in the factor of bias, removal of these columns assisted in ensuring greater accuracy in prediction and reduce non-predictors in the model. Moreover, all SKUs with zero sales were considered insignificant for an accurate prediction and eliminated from the dataset. Multiple variables such as actual vehicles in operation, establishments, and road quality indices were imputed with central aggregates and replaced negative sales with zero to avoid skew.

Further transformations included the creation of dummies for the categorical variables aiding the use of multiple groups in regression, scaling and standardization of the data, boosting the need for variable interactions via correlation check in the correlation matrix.

To increase the linear regression's efficacy and include the insights from the correlation map, we created select interactions. These interactions were second-order terms, and inclusion of these brought in the parameters of first and second powers(equation 2). The scope of this process was to analyse variables that yielded a positive business impact on the overall score and MSE value of the regression model, increasing the prediction accuracy. The equations (1) and (2) addresses the difference in models with interaction terms. These interaction terms brings in the capability of inclusion of synergies and combined impact of sales based on both the variables. Consideration on the interaction terms was on the basis of the statistical significance observed. Moreover, it was important to analyse terms which had insignificance on its own but raised the statistical significance to the overall predictions when they were a part of an interaction. This brings in the feasibility of identification of such variables with intrinsic impact on sales. An aim of proving the business impact using variables can also be achieved with interaction variables. A history of SKUs with good sales indirectly implies the availability of that particular item in the store. Such variable combinations and their significance are explored using the interaction terms as below:

$$Sales_{regression\ with\ no\ interactions} = \beta_1 x_1 + \beta_2 x_2 \ (1)$$

$$Sales_{regression\ with\ interactions} = \beta_1 x_1 + \beta_2 x_2 + \beta_{11} x_1^1 + \beta_{22} x_2^2 + \beta_{12} x_1 x_2 \ (2)$$

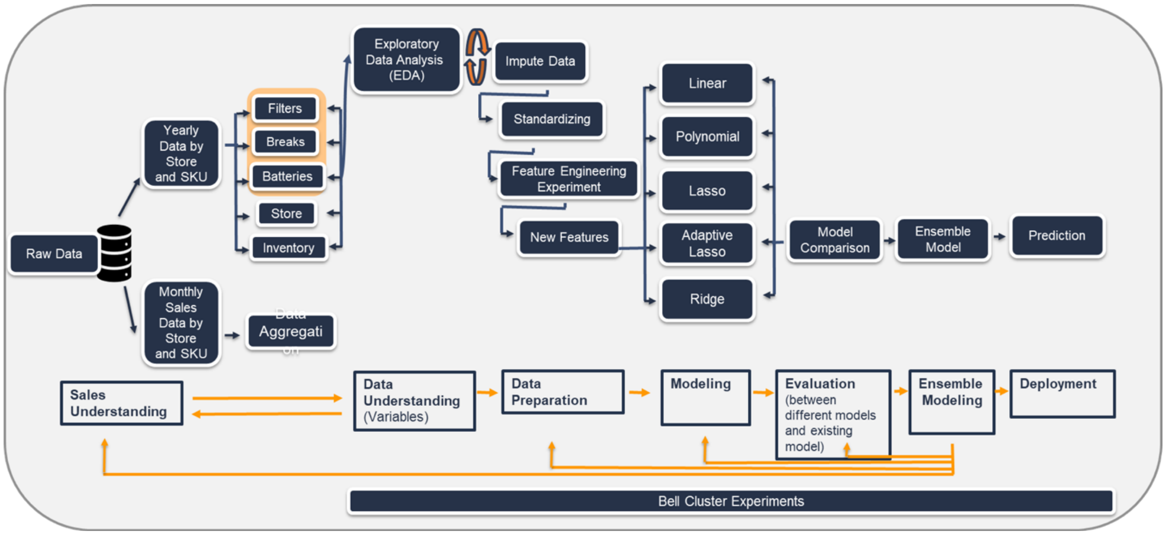


Fig 2. Study Design

Model

Libraries and functions

Dask_ML library was used for modelling provides scalable and distributed machine learning to cater to the challenges of large data size, high dimensionality, and computation issues. Dask helps in breaking down the dataset into small chunks which helps in parallelizing tasks and reducing computational load. Every function run on dask dataframes are stored in a workflow and not actually computed unless explicitly called. This ‘lazy’ computation greatly reduces computational requirement for executing operations on large datasets. The library comes with its own set of functions such as make_regression, StandardScaler, PolynomialFeatures, PCA. Dask dataframes were used for data pre-processing and the processed dask dataframe was converted back into pandas dataframe to run the model.

Data transformation

Several transformations were carried out on the data as per business requirements and for ease of modelling. Stores with outlier sales were dropped. To perform feature engineering using interaction term, variables with null values were imputed with business sense and some null values were dropped. The current year variables were not used in the model as they were not predictors. The null values for Part type were replaced with “Miscellaneous”. SKUs with 0 values for previous year gross sales were dropped, since gross sales is a significant predictor. The data was normalized using standard scaling method.

Feature engineering

2nd degree interaction terms were created. Principal Component Analysis (PCA) was carried out. To project the data to a lower dimensional space, PCA was used for reduction of linear dimensionality using Singular Value Decomposition. The number of components were varied depending upon variance. Quantity of SKUs sold in current year was treated as the response variable.

Data modelling

1. Linear regression: Linear regression is simple to implement, and the output coefficients are easier to interpret. It is less complex as compared to other algorithms, and hence is easier to use. There is a possibility of over-fitting, but it can be avoided through dimension reduction, regularization, and cross-validation.
2. Lasso regression: Least Absolute Shrinkage and Selection Operator or LASSO is a type of linear regression that uses shrinkage in the loss function and automates feature selection in the model. It could be useful in a situation where there is a mixture of large coefficients and small coefficients for the predictor variables.
3. OLS: The OLS estimator is reliable when the Gauss-Markov assumptions, also called OLS assumptions are met. If the data fulfil all the required assumptions, it is considered the best linear unbiased estimator available, i.e., BLUE.

Results

The graph illustrates the best few experimental models with high R-square and low mean squared error. The R-square ranges from 64.6% to 82% and MSE was between 0.465 and 4.722. Models such as Linear regression (1,2,12), Lasso (4) and OLS (3,13,5-11) were explored. Interaction terms were created by combining demographic variables, lost sales and quantities in previous year, quantity sold in previous and previous to previous year. PCA analysis was also performed. Model 1(Linear regression model) had the minimum MSE of 0.71 but the predictors used in the model were not significant which led to lesser R-square value. A few OLS models

(6,8) had decent R-square of ~75% but had high MSE of 4.72 indicating high biased or high variance estimate. The best model obtained was a Linear regression model (12) which used the interaction between quantity sold in previous and previous to previous year, with R-square of 81% and MSE of 0.71. The Mean Absolute Error averaged to ~0.186.
Similar results were observed consideration was on the brakes data instead of the filters using linear regression, lasso regression and OLS were performed on the brakes dataset. The ranges of R-square observed across these models are 0.81 to 0.835. The absolutes of the mean squared errors were 0.62-0.707. In terms of the batteries data under discretion, r squared of 0.911 was observed with the interaction terms implying a better fit on the model on the data.

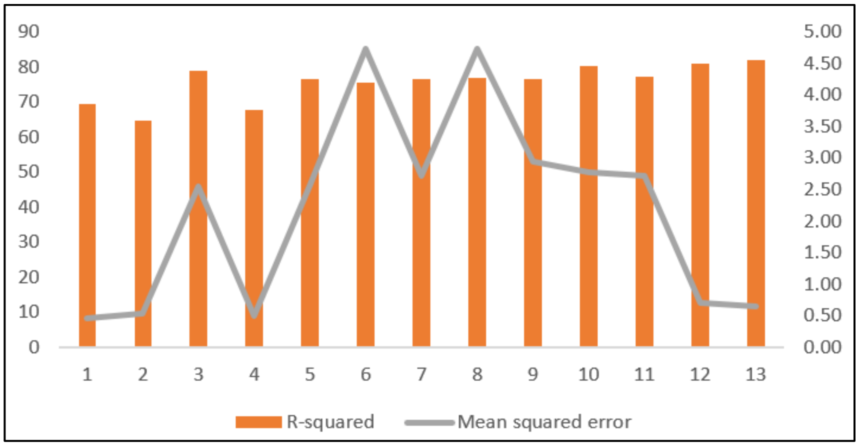


Fig 3. Model results

Conclusion

Through our use of high-performance computing on bell cluster, we were able to perform hundreds of experiments for different models and reached an MSE of <1 with a fairly good R-square value of >90%. Our model could play a crucial role on creating a strong business impact for the company. Below are the monetary benefits that could be availed through our model:

- Through percentage decrease of mean absolute error by ~90%, sales could be more accurately predicted. This would further reduce the inventory management costs for the company.
- With model-run time being as low as 2 min, it can provide another cost-saving opportunity

Future Scope

Since parallel computing of this scale demands high performance with computationally taxing requirements, inclusion of interactions can be incorporated via parallel computing to enhance the predictive power of the regression model.

References

Alwadei, Sahar and Dahab, Mohamed and Kamel, Mahmod (2017). A Feature Selection Model based on High-Performance Computing (HPC) Techniques. *International Journal of Computer Applications*. Vol. 180, p.11-16. https://doi.org/10.5120/ijca2017916054

Anupindi, R., Dada, M., & Gupta, S. (1998). Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Science*, *17*(4), 406-423. https://doi.org/10.1287/mksc.17.4.406

Blum, Kelly. (2020). Gartner Says Less Than 50% of Sales Leaders and Sellers Have High Confidence in Forecasting Accuracy. *Gartner* https://www.gartner.com/en/newsroom/press-releases/2020-02-12-gartner-says-less-than-50--of-sales-leaders-and-selle

Christen, M., Gupta, S., Porter, J. C., Staelin, R., & Wittink, D. R. (1997). Using market-level data to understand promotion effects in a nonlinear model. Journal of Marketing Research, *34*(3), 322-334. https://doi.org/10.1177/002224379703400302

Conerly, Bill. (2020). Forecasting Sales In These Uncertain Times. *Forbes.* https://www.forbes.com/sites/billconerly/2020/07/02/forecasting-sales-in-these-uncertain-times/?sh=25c256d35070

Cooper, L. G., Baron, P., Levy, W., Swisher, M., & Gogos, P. (1999). PromoCast™: A new forecasting method for promotion planning. *Marketing Science*, *18*(3), 301-316. https://doi.org/10.1287/mksc.18.3.301

Divakar, S., Ratchford, B. T., & Shankar, V. (2005). Practice prize article-chan4cast: A multichannel, multiregion sales forecasting model and decision support system for consumer packaged goods. Marketing Science, *24*(3), 334-350. https://doi.org/10.1287/mksc.1050.0135

Fader, P. S., & Hardie, B. G. (1996). Modeling consumer choice among SKUs. *Journal of marketing Research*, *33*(4), 442-452. https://doi.org/10.1177/002224379603300406

Kök, A. G., & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. *Operations Research*, *55*(6), 1001-1021. https://doi.org/10.1287/opre.1070.0409

Koschat, M. A. (2008). Store inventory can affect demand: Empirical evidence from magazine retailing. *Journal of Retailing*, *84*(2), 165-179. https://doi.org/10.1016/j.jretai.2008.04.003

Loten, Angus. (2021). Companies Adjust Predictive Models in Wake of Covid. *The Wall Street Journal.* https://www.wsj.com/articles/companies-adjust-predictive-models-in-wake-of-covid-11625160587

Ma, S., Fildes, R., & Huang, T. (2016). Demand forecasting with high dimensional data: The case of SKU retail sales forecasting with intra-and inter-category promotional information. *European Journal of Operational Research*, *249*(1), 245-257. https://doi.org/10.1016/j.ejor.2015.08.029

Malcolm X. (1962). Malcolm X's Fiery Speech Addressing Police Brutality, https://www.youtube.com/watch?v=6_uYWDyYNUg

Narasimhan, C., Neslin, S. A., & Sen, S. K. (1996). Promotional elasticities and category characteristics. *Journal of marketing*, *60*(2), 17-30. https://doi.org/10.1177/002224299606000202

Ton, Z., & Raman, A. (2010). The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production and Operations Management*, *19*(5), 546-560. https://doi.org/10.1111/j.1937-5956.2010.01120.x

Trapero, J. R., Kourentzes, N., & Fildes, R. (2015). On the identification of sales forecasting models in the presence of promotions. *Journal of the operational Research Society*, *66*(2), 299-307. https://doi.org/10.1057/jors.2013.174

Vulcano, G., Van Ryzin, G., & Ratliff, R. (2012). Estimating primary demand for substitutable products from sales transaction data. *Operations Research*, *60*(2), 313-334. https://doi.org/10.1287/opre.1110.1012