Optimal Clustering of Products for Regression-Type and Classification-Type Predictive Modeling for Assortment Planning

Raghav Tamhankar, Sanchit Khattar, Xiangyi Che, Siyu Zhu, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
rtamhank@purdue.edu; skhattar@purdue.edu; che4@purdue.edu; zhu554@purdue.edu;
lanhamm@purdue.edu

ABSTRACT

In collaboration with a national retailer, this study focused on assessing the impact of sales prediction accuracy when clustering sparse demand products in various ways, while trying to identify scenarios when framing the problem as a regression-problem or classification-problem would lead to the best demand decision-support. This problem is motivated by the fact that modeling very sparse demand products is hard. Some retailers frame the prediction problem as a classification problem, where they obtain the propensity that a product will sell or not sell within a specified planning horizon, or they might model it in a regression setting that is plagued by many zeros in the response. In our study, we clustered products using k-means, SOMs, and HDBSCAN algorithms using lifecycles, failure rates, product usability, and market-type features. We found there was a consistent story behind the clusters generated, which was primarily distinguished by particular demand patterns. Next, we aggregated the clustering results into a single input feature, which led to improved prediction accuracy of the predictive models we examined. When forecasting sales, we investigated a variety of different regression- and classification-type models and report a short list of those models that performed the best in each case. Lastly, we identify certain scenarios we observed when modeling the problem as a classification problem versus a regression problem, so that our partner could be more strategic in how they use these forecasts for their assortment decision.

Keywords: product clustering, classification vs regression-type modeling, assortment planning, decision-support

INTRODUCTION

Inventory management is a fundamental part of a firm's operational strategy as it aims to ensure the availability of their product to their customers. This aspect is of paramount importance as it directly impacts the customer experience, which has become one of the key focused area of the firms in our digital world. In addition to this, there are huge costs associated with movement and holding of inventory. Hence, the key is to understand how to operationally optimize distribution of products.

A firm sets its financial objectives such as sales, inventory, margin at various levels which impacts the organization's merchandising strategies and decisions. To meet these financial KPIs, historical data can be paired with effective analytics to identify and anticipate opportunities to improve category-level assortments, and better product placement timing for merchants [1].

Assortment planning is the process of planning, organizing and selecting a group of products which will be designated to particular areas and during specific periods of time [2]. This study in collaboration with a national retailer whom tends to stock large assortments of sparse demand products seeks to answer a few questions. First, what is the best approach to frame modeling such products. For the given industry, retailers will often consider attributes such as product type, store location, geographical demand, product life cycles, and failure rates as often telling predictors or demand. However, modeling demand when demand is a very challenging problem, even more so than traditional intermittent demand. Does the retailer frame the prediction problem as a regression-type problem, which is the most common scenario, or can framing the problem as a classification-problem provide more reliable forecast measures to support the assortment decision?

In our study, we seek to provide some empirical evidence by doing traditional clustering of features commonly used regardless of how the problem is framed, then build predictive models using those clusters as an independent variable to predict the number of units sold. Units sold is often either zero or one in the pre-determined planning horizon, but there are cases where more than one unit is sold.

Thus, our research questions in this study are:
- Can informative clusters be generated using popular unsupervised learning algorithms, and is there a business story about those clusters?
- How does clustering of products improve the predictive performance of models in a regression and classification setting?
- In which scenarios is the regression approach preferred to the classification approach when modeling very sparse demand products?

We structure our paper by first performing a literature review of predictive modeling to support assortment planning decisions, describe the data used in our study, detail our study methodology, discuss the models we investigated to predict demand, summarize our results, and provide answers to our research questions in the conclusions.

LITERATURE REVIEW

We have critically reviewed prior academic literature and articles to provide us guidance in our study. Our research areas cover both methodologies used in assortment planning and statistical learning approaches for clustering and supervised and unsupervised predictive analysis to support this area.

Assortment planning (AP) is defined as the process of optimizing the inventory level of selective types and number of products to be carried at retail stores (Kapoor and Singh, 2014). The objective of AP is to strategically determine the optimal quantity to carry, by reducing the chance of stock-outs (Baker, 2005) and therefore drive bottom line improvements. Traditionally, the assortment decisions were limited to the

consideration of substitution and elasticity demand (Aydın and Hausman, 2009) as well as shelf-space studies (Gajjar and Adil, 2011) as shown in Table 1 below.

Table 1: Type of Assortment Planning Models and Variables Considered

Author, Year	Model Adopted		Variable Considered
	Space Elastic	Substitution	
(Gajjar and Adil, 2010)	X		Space available, Upper and lower limits on space
(Yücel, et al., 2009)		X	Cost of Substitution, Cost of supplier selection, shelf-space limitations
(Abbott and Palekar, 2008)	X		Direct and cross space elasticity, Inventory, profit, cost
(Shah and Avittathur, 2008)		X	Substitution, Space, Multi-item inventory level, Cannibalization
(Li, 2007a)		X	Substitution, Continuous and discrete in-store traffic, Product profit, Cost
(Kök and Fisher, 2007)		X	Utility, Substitution rate, Cost, Price, facings
(Hariga, et al., 2007)	X		Direct and cross space elasticity, Locational effect, Inventory level
(Gaur and Honhon, 2006)		X	Static and dynamic substitution, Space, Location, inventory, profit
(Rajaram, 2001)		X	Selling and cost price, Set up cost, Budget, Salvage price, Shelf-Space

From our review of the literature, most studies tend to focus on substitution behavior and aspects of computational complexity of the models used to generate assortments under unrealistic scenarios. Thus, in this paper we currently do not provide a clear connection (or extension) from these studies to the problem we investigate with our industry partner. However, we have found close connections to other research streams focused on using machine learning techniques for demand forecasting problems.

Machine learning algorithms have been a hot topic in business and demand forecasting as of late, because of their ability to discover trends and behaviors in the data that traditional approaches are just not able to identify. We posit that these techniques can provide the best guidance to retailers to help improve the AP process. In this study, we combine clustering techniques (i.e. unsupervised learning) with predictive modeling techniques (i.e. supervised learning).

Predictive analytics using machine learning has tremendous advantages over traditional methods because of the often lack of required assumptions needed. Instead of modeling a trend based on historical sales data, predictive analytics using machine learning can incorporate all relevant features such as product attributes,

store capacities, transfer/logistics costs, lead-time variability, assortment rules, geo-demographic diversity, price-elasticity of demand, etc. into the demand model to better understand the product-place-promotion aspects that are fundamental components of purchasing behavior (Cook, 2016). These features could lead to better demand forecasts, which could lead to better assortments that have the right depth and diversity balance, as well as lower the chance of product cannibalization, and frequent order replenishment problems (Krupnik, 2016).

The goal of segmenting the observations into different clusters is to provide users with in-depth understanding of the "internal mechanism" within the data by maximizing in-group similarity and minimizing between-group similarity. Measurement of similarity varies based on specific context, which can generally be categorized as "relationship-based" and "distance-based" clustering (Wu, 2006). Distance-based approaches include K-means and examples of relationship-based methods are linear or logistic regression.

Research has shown that ensemble approaches tend to perform better than single classifiers (Opitz & Macline 1999; Quinlan, 2006), but the ensemble methods discussed there are usually bagging, boosting and stacking. Unlike the traditional studies, Williams' research investigated the ensemble approach of combining clustering and predictive modeling. His research showed that the ensemble approaches increased the accuracy to predict recurrence and non-recurrence for breast cancer patients. (Williams, 2008)

Once the models are deployed, a category manager can very effectively build an assortment plan. While the decisions are made based on the predictive model and constraints of the prescriptive model (e.g. shelf space, budget, etc.), there are other scenarios that should be considered, such as product cannibalization, that also must be considered. The Multinomial Logit (MNL) choice model for example assumes certain properties due to which substitution effects among products (such as cannibalization) cannot be captured (Jena, Dominik, 2017). Thus, while our study does not clearly tie into the academic research, there still necessitates answers to these fundamentally important questions of how to model extremely sparse demand problems to better support the assortment planning decision.

We have referred to several research papers in understanding how clustering has been applied for assortment planning. One of our findings is that in one of the very recently published papers, they propose a dynamic clustering policy as a prescriptive approach for assortment personalization. They introduce a policy that adaptively combines estimation (by estimating customer preferences through dynamic clustering) and optimization (by making dynamic personalized assortment decisions using a bandit policy) in an online setting. Hence, if the customers could be specifically identified, clustering approaches are useful aids in assortment planning from store level to personalization level. (Bernstein, 2017)
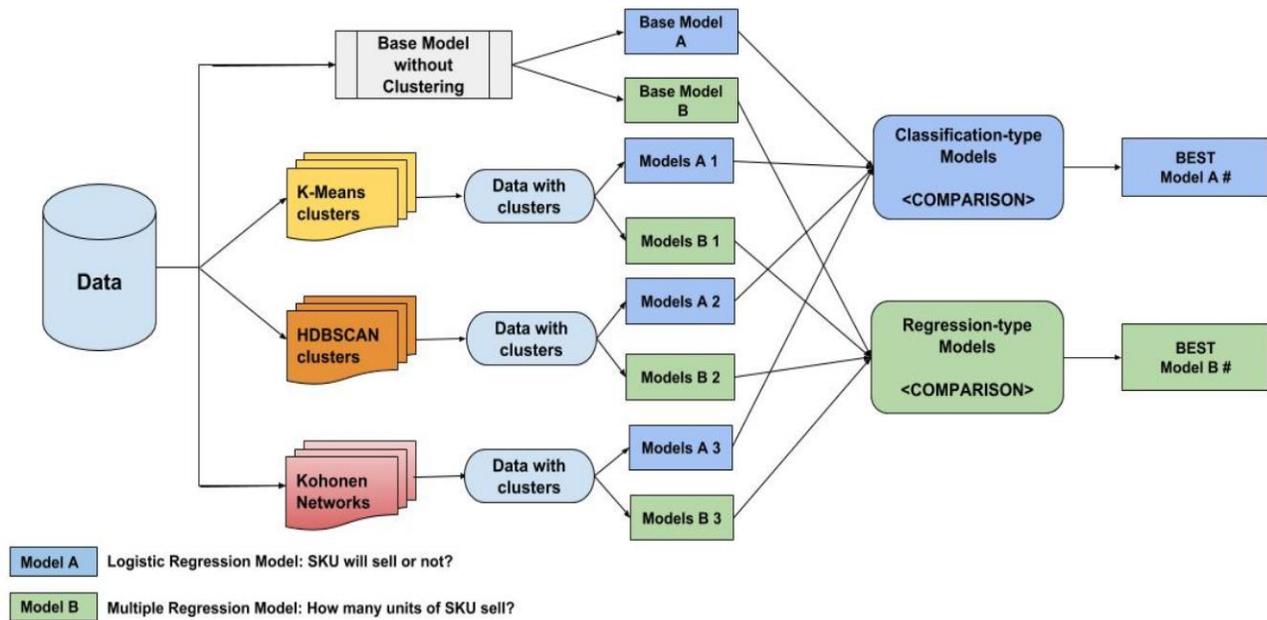
DATA

The dataset was provided by a national retailer and contains information about sales of SKUs coming from a vast number of stores over a two-year time frame. Business parameters such as product lifecycles, failure rates, and number of customers that the SKU supports in a specified area were provided. No additional information can be described about the dataset due to confidentiality concerns.

METHODOLOGY

Figure 1 below shows the overall approach that this study followed. To answer our research questions, we clustered the data three different ways, as well as building a predictive model without clusters. Next, the clusters we fed as inputs along with other features found significant at predicting demand into them for training. Some problems were regression-type models that has the target variable values as 0, 1, 2, etc.,

while the classification-type models had a target response of 0 or 1. One being if the unit had some one or more units during the previous time window and zero if it had not sold any units.

Figure 1: Methodology outline

Before clustering and predictive modes were trained, all irrelevant (insignificant to the objective) features were removed from the dataset. Missing values were imputed using the the mean value. This turned out not to negatively impact the distribution of the variable as is often the case. Numerical features were normalized to maintain scale-uniformity, which should allow the models to train without being biased by the range of the features. Data-partitioning was performed using an 80-20% train-test set all the predictive models.

CLUSTERING MODELS

Model 1- Base Model without Clusters
The base model was developed to provide a reference point in evaluating the performance of models with different clusters. It was directly used in the regression models to forecast sales, and in the classification-case to predict purchase propensities.

Model 2: k-Means Clustering Model
K-means clustering was performed on failure rates and percentage lifecycle remaining variables, which are common features to use in this application area. K-means is a popluar unsupervised learning algorithm, which is used for unlabeled data (i.e., data without defined categories or groups). The goal of this algorithm is to find groups in the data. The algorithm works iteratively to assign each data point to one of K groups based on the features that are provided. Data points are clustered based on feature similarity which is generally calculated using Euclidean distance. The appropriate number of clusters was obtained by analyzing an elbow plot after implementing the algorithm for different values of 'k'. This plot provides a heuristic to gauge the average cohesion within clusters. Once the optimum value was identified, the cluster number was used as a categorical variable for classification and regression models along with other predictor variables. The variables used to cluster were omitted from the forecasting techniques to prevent bias towards those specific variables.

Model 3: HDBSCAN Clustering Model

The HDBSCAN clustering technique was implemented on the same set of variables to obtain clusters based on the density of data points. It extends the popular DBSCAN (Density-Based Spatial Clustering of Applications with Noise) clustering algorithm. It is a density-based clustering algorithm that provided a set of points in some space will group together points that are closely packed together (points with many nearby neighbors), marking as outliers, points that lie alone in low-density regions (whose nearest neighbors are too far away).

Model 4: Kohonen Network Clustering Model

Kohonen Network is a type of unsupervised learning Artificial Neural Network to produce a low-dimensional representation of the input space. Kohonen Networks also known as Self-Organizing Maps (SOMs) help to understand high dimensional data by reducing the dimensions of data to a map. The final map reveals which observations are grouped together. The algorithm works by having several units compete for the current object. Once the data have been entered into the system, the network of artificial neurons are trained by providing information about inputs. The weight vector of the unit is closest to the current object becomes the winning or active unit. During the training stage, the values for the input variables are gradually adjusted in an attempt to preserve neighborhood relationships that exist within the input data set. As it gets closer to the input object, the weights of the winning unit are adjusted as well as its neighbors.

PREDICTIVE MODELS

In our study, we combined supervised and unsupervised learning models and evaluated the best models that predicting sales units and propensity to sale. We researched and investigated both machine learning clustering techniques and business segmentation approaches to apply to our specific business problem and data set.

Classification-Type Modeling:

The goal of the classification type modeling is to predict the probability of a SKU being sold at a given store and eventually classify it as a seller or non-seller based on a specific probability cutoff threshold value. Several classic machine learning classification algorithms like Logistic Regression, Classification Tree (CART), and Linear Discriminant Analysis (LDA) models were developed. More sophisticated techniques such as Bagged Classification Trees, Boosted Logistic Regression, a multilayer perceptron feedforward artificial neural network (MLP) were also explored to evaluate these values. Bagged (or boosted aggregated) trees tend to have lower variance than individual trees. Boosted algorithms tend to shrink both bias and variance as records that have been misclassified are provided greater weights and essentially probabilities are generated that will tend to fall more closely to their respected class (e.g. 0 or 1) than not boosting. Neural networks provide the ability to fit complex relationships which often exist in sparse demand patterns.

Logistic Regression:

Logistic regression is the simplest and most commonly used classification approach in order to predict the binary outcome (independent variable). Similar to the linear regression model, the purpose of logistic regression is to describe the relationship between independent variables and a number of independent variables. Unlike the linear regression, the maximum likelihood estimation is used to estimate the parameter coefficients where the response variable (0 or 1) has been transformed into log-odds via link function known as the logit.

CART

The Classification and Regression Trees (CART) algorithm, introduced by Leo Breiman, refers to the commonly known decision tree model used for classification or regression type of problems. The advantage of a decision tree model is easy to interpret. A tree can be "learned" by splitting the total dataset into subsets

based on the attribute value test, and this process is done in a recursive manner. The number of splits determines the tree complexity. A popular approach to obtaining a tree that generalizes well on a test set is to grow a large tree that overfits to the training data, then prune it back until a consistent test error is obtained. This is how we developed our tree model in this study.

Bagged Classification Tree
A bagged decision tree was investigated because there has been strong evidence that bagging weak tree classifiers can lead to reduction in overall error, by reducing the variance component in the error. The variance in the bias-variance tradeoff is the error that is seen by applying the algorithm on different data sets. An analyst would hope to achieve a robust model, meaning that it will generalize well from dataset to dataset, meaning the error obtained from each dataset is similar. When multiple decision trees results are aggregated (often by averaging), the resulting model results tend to reduce overfitting and lead to a more robust model.

LDA
Linear discriminant analysis (LDA) is a generalized Fisher's linear discriminant method, used to find a linear combination of features as linear classifier to separate two or more classes of objectives. It is also viewed as an approach for dimensionality reduction before classification. LDA is also closely related to principal components analysis (PCA) as they both look for the linear combinations of variables to further explain the data. However, LDA tries to differentiate the class of data, while PCA looks for similarity to build feature combination.

MLP
A multilayer perceptron feedforward artificial neural network (MLP) consists of at least three layers of nodes. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. MLP utilizes a supervised learning technique called backpropagation for training. Its multiple layers and non-linear activation distinguish MLP from a linear perceptron. It can distinguish data that is not linearly separable and seems appropriate for the sparse demand nature of this study.

Regression-Type Modeling:
Regression-type models were explored to predict the exact quantity of products that will be sold at a store in an upcoming planning horizon. The models were built using only significant variables which helped in forecasting the sales. The clusters obtained from different algorithms were also included in the model as categorical type variables. Multiple Linear Regression and several machine learning algorithms such as Neural Networks and Decision Trees were explored to predict sales. Additionally, a zero-inflated Poisson Regression model was developed for cases having a large number of '0's in the response variable.

Multiple Linear Regression
By fitting a linear equation to the observed model, multiple linear regression tries to model the relationship between the continuous predicted variable and two or more predictor variables. Both forward and backward selection approaches were used to identify a consistent set of drivers that explain sales.

Neural Networks
Artificial Neural Networks (ANN) were originally a biologically-inspired programming paradigm that enabled scientists to learn from brain activity. It has the ability of deriving meaning from complicated data and can be used to extract patterns and trends that are identifiable by normal algorithms. ANN is based on a collection of connected nodes referred to as artificial neurons. Data serves as signal inputs and the output of each neuron is calculated by nonlinear function that aggregate all inputs.

Zero-Inflated Poisson Regression
Zero-inflated Poisson (ZIP) Regression is a specialized multiple linear regression which particular fits well for data having skewed distribution of 0s in the dependent variable. ZIP models assume that some zeros

occur via a Poisson process, but others were not even eligible to have the event occur. Thus, there are two processes at work—one that determines if the individual is even eligible for a non-zero response, and the other that determines the count of that response for eligible individuals.

RESULTS

When developing the clusters, we decided to use 6 clusters based on the elbow plot shown in Figure 2. In this plot we were looking for an "elbow" or kink that would suggest past that number of k clusters there are dimensioning returns the plot itself shows the mean squared error (MSE) for different number of 'k' clusters.

Figure 2: k-Means elbow plot
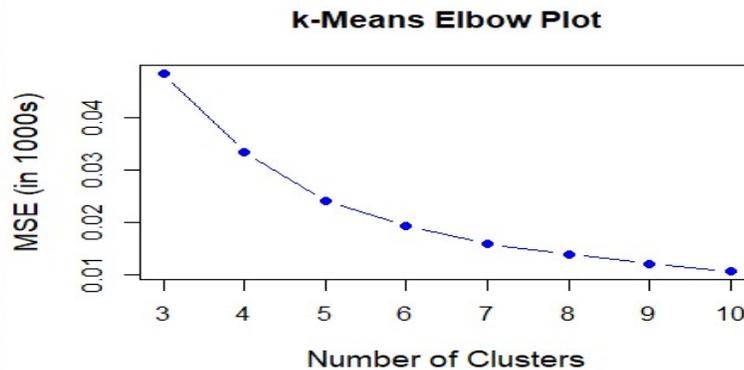
K-Means clusters are built based on vicinity of data points compared to a centroid. In Figure 3 below, we found that the green colored clusters have imminent demand and should be expected in stores frequently.

Figure 3: Graph showing the optimal number of cluster for K-Means
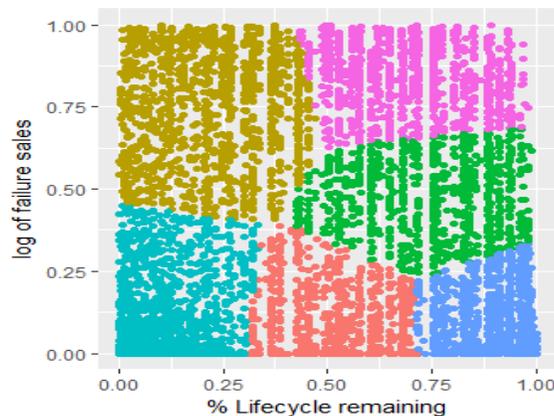
Kohonen clusters as shown in Figure 4, were built based on the magnitude of failure sales and percentage lifecycle remaining values. The clusters with no green or white color cross section represents products with low value of both the variables

Figure 4: Vicinity based clustering based on neural networks (Kohonen Networks)

% Lifecycle Remaining   □ Log of Failure Sales
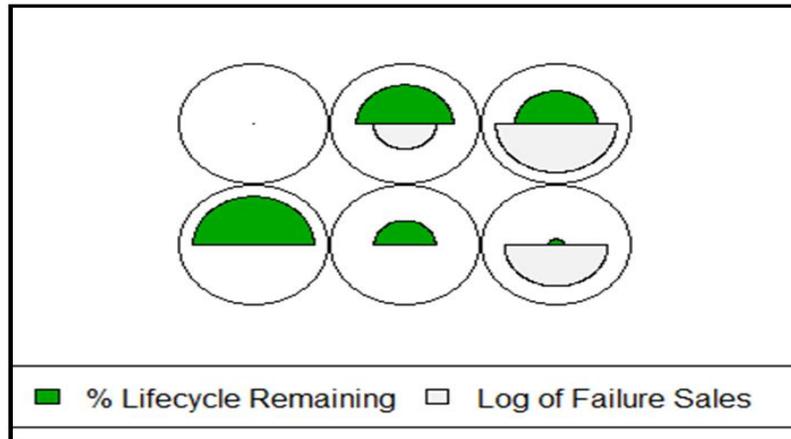
Lastly in Figure 5, cluster 3 represents products which have intermediate value of lifecycle remaining and relatively high failure sales. HDBSCAN, on the hand classified a large chunk of data points as noise (black-colored dots) making difficult to infer any business segmentation of the clusters.

Figure 5: Density based clustering algorithm with 4 unique clusters (HDBSCAN)
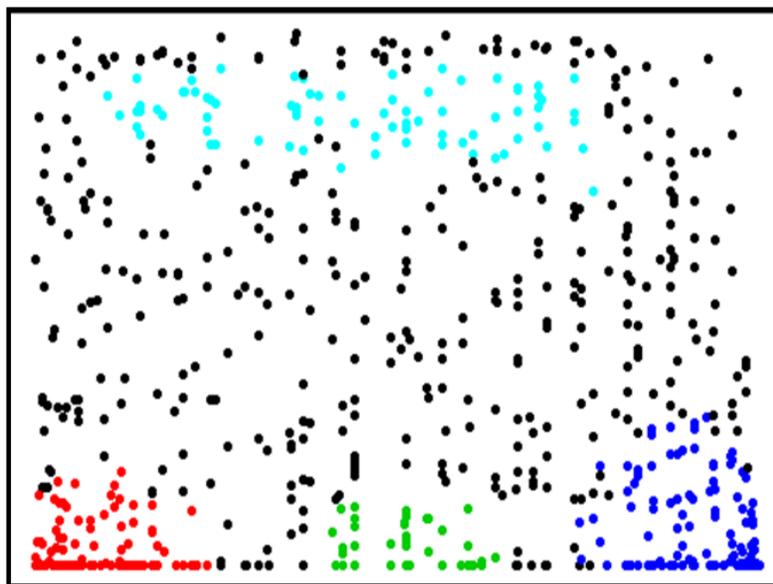
We chose k-means as the final cluster to publish the results as it made business segmentation of SKUs more interpretable. The Kohonen model predicted almost the same clusters as k-means and hence was not substantially differentiable. The HDBSCAN model classified the major chunk of the data points as noise leaving a small subset with clusters. Thus, due to the shortcomings of Kohonen and HDBSCAN models, we proceeded our analysis with k-means clusters.

Classification Model Results
Table 2 summarizes the statistical performance for each classification model we investigated. The Bagged Logistic Regression without clustering had the greatest area-under-the-curve (AUC) with 0.9841. LDA performed the best (AUC = 0.9825) among the predictive models that used the k-means clusters.

Table 2: Baseline model vs Models with clusters comparison

Model	Baseline Model				With Clusters					
	Accuracy	Sensitivity	Specificity	AUC	Accuracy	Sensitivity	Specificity	AUC		
Logistic Regression	0.95241	0.98397	0.93889	0.95843	0.88535	0.95422	0.85584	0.89847	▼	-6.3%
Logistic Boosted	0.92294	0.88446	0.93943	0.91561	0.92552	0.92462	0.92590	0.92535	▲	1.1%
Logistic Bagged	0.98517	0.97958	0.98756	0.98410	0.96308	0.97917	0.95618	0.96615	▬	-1.8%
CART	0.93942	0.95344	0.93342	0.94209	0.95826	0.96357	0.95599	0.95927	▲	1.8%
LDA	0.93885	0.96511	0.92759	0.94385	0.98660	0.96511	0.99580	0.98250	▲	4.1%
Neural Network	0.95868	0.93217	0.97004	0.95363	0.97756	0.93503	0.99578	0.96946	▲	1.7%

Since most of the models performed similar, we used a probability calibration plot as another means to discriminate among competing models. Since the sparse demand assortments using propensity scores are essentially ranks of which SKUs will be in the assortment or not, choosing a model that is best calibrated would be idea. Figure 6, shows the probability calibration plots for the models we investigated. On the x-axis are binned predictive probabilities, and on the y-axis reveals the percentage of actual sellers for each bucket. In theory, lines in this plot that are closer to a 45-degree line are better calibrated than those that are not. Thus, such a model would provide better ranked propensities for the assortment planning decision. We selected the Bagged Logistic model as our optimal classification model.

Figure 6: Calibration Plot for all logistic regression models
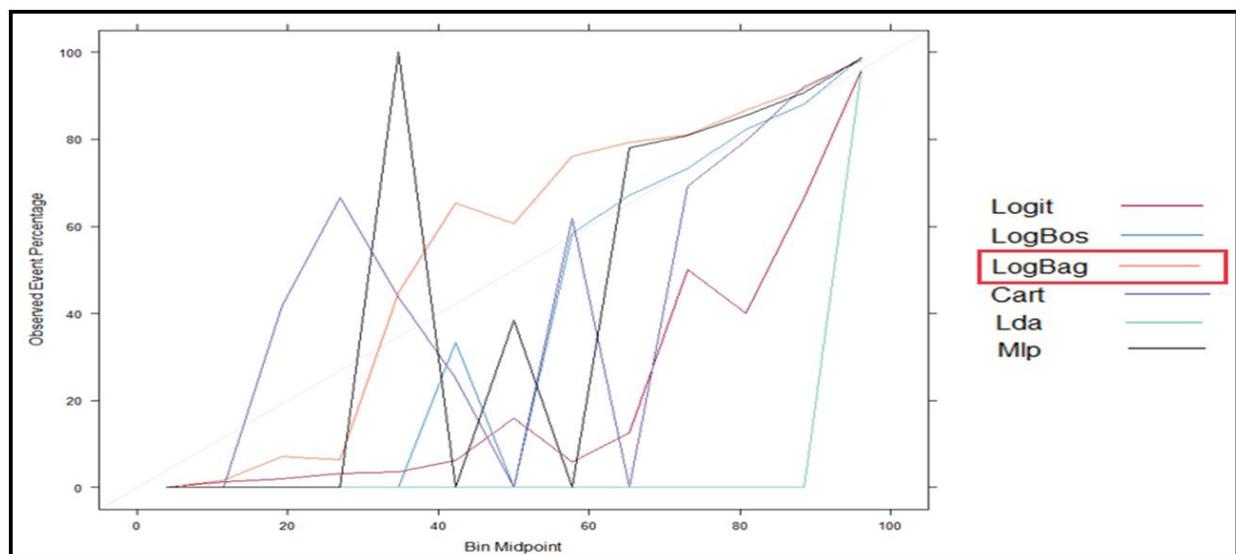
To evaluate the regression models, root mean-squared error (RMSE) and adjusted R-squared were evaluated on the training and test datasets. Both MLR backward and forward selection produced exactly the same set of feature combinations and therefore produced the same output. Based on the coefficients for the corresponding variables, we observed that the quantity sold in previous years had great impact on this year's sales number. We observed certain features had interesting business implications but could not disclose those in our paper.

Table 3: MLR Forward and Backward results for train and test for base model and mode with clusters

Regression		Training		Testing	
Model	RMSE	R-Squared	RMSE	R-Squared
MLR_Forward	0.4278469	0.3864196	0.5913151	0.4898889
MLR_Backward	0.4278469	0.3864196	0.5913151	0.4898889

	With Clustering				
Regression	**Model**	RMSE	R-Squared	RMSE	R-Squared
	MLR_Forward	0.4292657	0.3738107	0.6002026	0.4934494
	MLR_Backward	0.4292297	0.3738949	0.6001385	0.4936345

Table 4 shows the results for the regression-type models with and without clusters we obtained. We observed that the zero-inflated Poisson regression outperforms all other models with an adjusted R2 of 0.905. Additionally, its RMSE error is was the lowest among all models.

Table 4: Table comparing after clustering the statistical performance of baseline regression models

	Baseline Model		**With Clusters**		
Model	RMSE	Adjusted R-Squared	RMSE	Adjusted R-Squared	
MLR_Forward	0.56785	0.58642	0.52345	0.64960	▲ 10.8%
Poisson Regression	0.46397	0.86643	0.43014	0.90592	▭ 4.6%
Neural Network	0.67834	0.56827	0.72970	0.55075	▼ -3.1%
Decision Tree	0.72543	0.47462	0.73080	0.46838	▼ -1.3%
Decision Tree Bagged	0.65633	0.50946	0.61970	0.51980	▭ 2.0%

CONCLUSIONS

Assortment planning is one of the most important operational aspects that all retailers face. A key component to the assortment planning decision is having the ability to estimate demand. Intermittent demand is always a challenge, but extremely sparse demand is the most challenging demand forecast to generate, only behind predicting demand for new products. In our study we investigated the impact that clustering has on improving predictions for sparse demand models, but also investigated how one might frame such a scenario given that the target variable might have an abundance of zeros. This study is the groundwork to help our business partner to recommend whether to build classification-type or regression-type predictions for its assortment planning decision-support.

Thus, can informative clusters be generated using popular unsupervised learning algorithms, and is there a business story about those clusters? We employed k-Means, Self Organizing maps (Kohonen networks), and HDBSCAN to build clusters. We found that k-Means performed the best among the three. We also found that there was indeed a business story in these clusters.

Secondly, how does clustering of products improve the predictive performance of models in a regression and classification setting? We found that four of the six classification models we tried had improved performance with clusters than without. However, the best model Bagged Logistic Regression which performed the best among all competing models was achieved without clustering. In the regression setting, only two of the five models had a higher adjusted R-square using clustering than not clustering. However, our best model for this set of SKUs had an adjusted R2 improvement of 4.6%. These results do not allow us to have a definitive claim that clustering or not clustering is the best way to go in all situations. The evidence we provide suggests that most likely clustering should be used on a case by case basis where it performs better.

Lastly, in which scenarios is the regression approach preferred to the classification approach when modeling very sparse demand products? We obtained a highly accurate classification model with an accuracy of 98.5% with an AUC of 98.4% via the Bagged Logistic Regression. Also, the probabilities seemed decently calibrated and thus could be used for ranking purposes. In the regression setting we obtained an adjusted R-square using Poisson Regression that is also considered very high (90.5%).

Since we were obtained models that performed very well in both cases, we believe this study should be extended by developing assortments using both models via our partners assortment recommendation engine. We believe only then can we tie in the statistical performance measures to the business performance measures (e.g. sales, lost sales, inventory costs, lost opportunity costs, etc.)

REFERENCES

All-in-one Category Management and Planogram Software
https://www.dotactiv.com/assortment-planning

Getting More from Retail Inventory Management, Jan 2013
http://deloitte.wsj.com/cio/2013/01/22/getting-more-from-retail-inventory-management/

Abbott, H. & Palekar, U. S. (2008). Retail replenishment models with display-space elastic demand. European Journal of Operational Research, 186(2), 586-607.

Baker, Stacy. "Assortment planning for apparel retailers-Benefits of assortment planning." ABI/Inform Collection, June 2005.

Berkhin, Pavel. Survey of Clustering Data Mining Techniques. 2002, www.cc.gatech.edu/~isbell/reading/papers/berkhin02survey.pdf.

Bernstein, Fernando, et al. "A Dynamic Clustering Approach to Data-Driven Assortment Personalization." By Fernando Bernstein, Sajad Modaresi, Denis Saure :: SSRN, 9 June 2017, papers.ssrn.com/sol3/papers.cfm?abstract_id=2983207.

Cook, Chris. "Building the Optimal Retail Assortment Plan with Predictive Analytics.", 19 Feb. 2016. www.linkedin.com/Pulse/Building-Optimal-Retail-Assortment-Plan-Predictive-Analytics-Cook

Gajjar, H. K. & Adil, G. K. (2010). A piecewise linearization for retail shelf space allocation problem and a local search heuristic. [Article]. Annals of Operations Research, 179(1), 149-167.

Hasmukh K. Gajjar; Gajendra K. Adil. "Heuristics for retail shelf space allocation problem with linear profit function, International Journal of Retail & Distribution Management." DeepDyve, Emerald Group Publishing Limited, 15 Feb. 2011, www.deepdyve.com/lp/emerald-publishing/heuristics-for-retail-shelf-space-allocation-problem-with-linear-oGV9fT1lk3.

Hariga, M. A., Al-Ahmari, A. & Mohamed, A. R. A. (2007). A joint optimisation model for inventory replenishment, product assortment, shelf space and display area allocation decisions. European Journal of Operational Research, 181(1), 239-251.

Kapoor, Rohit. and Singh, Alok "A Literature Review on Demand Models in Retail Assortment Planning." Academia.edu, 2014,

www.academia.edu/5179110/A_Literature_Review_on_Demand_Models_in_Retail_Assortment_Planning.

Kok, A. G. & Fisher, M. L. (2007). Demand estimation and assortment optimization under substitution: Methodology and application. Operations Research-Baltimore, 55(6), 1001-1021.

Krupnik, Yan. "Say Goodbye to Inventory Challenges With Predictive Analytics." News | Apparel Magazine(AM), 4 Jan. 2016, www.apparelmag.com/say-goodbye-inventory-challenges-predictive-analytics.

Li, Z. (2007a). A single period assortment optimization model. Production and Operations Management, 16(3), 369-380.

Jena, Sanjay Dominik, et al. "PARTIALLY-RANKED CHOICE MODELS FOR DATA-DRIVEN ASSORTMENT OPTIMIZATION." Data Science for Real Time Decision Making, Sept. 2017. http://cerc-datascience.polymtl.ca/wp-content/uploads/2017/09/Technical-Report_DS4DM-2017-011.pdf

Opitz, D., & Macline, R. (1999). Popular ensemble methods: An empirical study. Artificial Intelligence Research, 11, 169 – 198.

Quinlan, J.R. (2006). Bagging, Boosting, and C4.5. Retrieved January 29, 2008, from http://www.rulequest.com/Personal/q.aaai96.ps

Shah, J. & Avittathur, B. (2007). The retailer multiitem inventory problem with demand cannibalization and substitution. International Journal of Production Economics, 106(1), 104-114.

Williams, Philicity K. CLUSTERING AND PREDICTIVE MODELING: AN ENSEMBLE APPROACH ... 9 Aug. 2008, www.bing.com/cr?IG=9E4CBDB21E994715B83382B14FD36D28&CID=107E8AD488686F8115A9818289C76E3B&rd=1&h=cjGUbFZYHKNHQ4lTsXtd0YNVvSZyNRsYhTXRRQz9lB8&v=1&r=https%3a%2f%2fetd.auburn.edu%2fbitstream%2fhandle%2f10415%2f1546%2fWilliams_Philicity_49.pdf%3bsequence%3d1&p=DevEx,5064.1.

Wu, Senlin. "DATA MINING USING THRESHOLD-VARIABLE AIDED RELATIONSHIP-BASED CLUSTERING." Doctor of Philosophy in Management Information Systems, University of Chicago, 2006.

Yücel, E., Karaesmen, F., Salman, F. S. & Türkay, M. (2009). Optimizing product assortment under customer-driven demand substitution. European Journal of Operational Research, 199(3), 759-768.