Annals of Operations Research (2021) 303:159–174
https://doi.org/10.1007/s10479-020-03666-w

S.I. : DATA MINING AND DECISION ANALYTICS

A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches

Majd Kharfan[1] · **Vicky Wing Kei Chan**[2] · **Tugba Firdolas Efendigil**[3]

Published online: 23 June 2020

Abstract

Companies in the fashion industry are struggling with forecasting demand due to the short-selling season, long lead times between the operations, huge product variety and ambiguity of demand information. The forecasting process is becoming more complicated by virtue of evolving retail technology trends. Demand volatility and speed are highly affected by e-commerce strategies as well as social media usage regards to varying customer preferences, short product lifecycles, obsolescence of the retail calendar, and lack of information for newly launched seasonal items. Consumers have become more demanding and less predictable in their purchasing behavior that expects high quality, guaranteed availability and fast delivery. Meeting high expectations of customers' initiates with proper demand management. This study focuses on demand prediction with a data-driven perspective by both leveraging machine learning techniques and identifying significant predictor variables to help fashion retailers achieve better forecast accuracy. Prediction results obtained were compared to present the benefits of machine learning approaches. The proposed approach was applied by a leading fashion retail company to forecast the demand of newly launched seasonal products without historical data.

Keywords Forecasting · Machine learning · Fashion retailer · Newly launched products

T. Firdolas Efendigil: This project was advised and performed at MIT Center for Transportation and Logistics.

✉ Tugba Firdolas Efendigil
tugbae@mit.edu

Majd Kharfan
majd.kharfan@converse.com

Vicky Wing Kei Chan
chanwkvicky@hotmail.com

1 Converse Inc., 1 Lovejoy Wharf, Boston, MA 02114, USA

2 Google Hong Kong Ltd, Suites 2501, Tower 2, Times Square, 1 Matheson St, Causeway Bay, Hong Kong

3 Massachusetts Institute of Technology (MIT), Center for Transportation and Logistics, 77 Massachusetts Avenue, E40, Cambridge, MA 02139, USA

1 Introduction

The fashion industry has immensely evolved especially after the introduction of e-commerce. The industry is influenced by internal and external factors including personalization, omnichannel competition, social media influencers, economic fluctuations, political movements and others. These shifts have led to shorter product lifecycles and the obsolescence of the retail calendar as well as globalization of sourcing and manufacturing, and consequently an increase in demand variability. Consumers have become more demanding and less predictable in their purchasing behavior that expects high quality, guaranteed availability and fast delivery. Besides, the imported items from far countries are needed to be predicted up to almost 1 year before the raw materials are ordered (Thomassey 2010; Mostard et al. 2011; Hui and Choi 2016).

As a result of these changes, fashion retailers need to build an agile supply chain that adapts quickly to the unsteady consumer demand. Running an agile supply chain crucially requires a well-established demand management. It serves as a major input for planning across different supply chain and business functions, including raw materials planning, supply planning, inventory management, sales and merchandising. Poor forecasting results can lead to a significant inventory shortages including stock outs and loss in revenues and market share to competitors, or to excessive inventory, i.e., frozen capital and high obsolescence. Therefore, having significant demand forecasting capability is essential to support supply chain functions of the companies. In addition to the difficulties in forecasting, fashion industry has a complex structure regarding the product selling approach which is based on a change in each season. Since, limitations of in-season replenishment are observed due to the global sourcing and manufacturing, an accurate demand forecast is required before the beginning of the season when historical data is not available.

This study presents a data-driven methodology to optimize demand forecasting process for fashion retailers by leveraging machine learning (ML) techniques in order to have an improved forecast accuracy for newly launched products without historical data in a changing market environment. Moreover, our aim is to present how to take advantage of using ML tools for a real industrial forecasting problem. The rest of the paper is organized as follows; Sect. 2 reviews the contributions in literature. Section 3 includes the data-driven methodology and introduces the company problem, while Sect. 4 gives the comparative results of the methodology. The conclusions, the limitations and the further opportunities are covered in Sect. 5.

2 State-of-art

In demand forecasting, the fashion industry is extremely challenging with the volatile consumer demand, the strong seasonality of sales, the wide number of items with short lifecycles and the lack of historical data (Thomassey 2010). Consumer demand is the result of the interplay among a number of factors that ideally serve as predictor variables in generating demand forecasts. In practice, sometimes the effect of these factors can be difficult to decouple, e.g., price and seasonality are interdependent on each other, since the same sales price may lead to different demand values at different weeks of the year (Kaya et al. 2014). Besides these complicated factors, some varying fashion trends like design and style which shall be up to date and not be renewed for the further collections, are other reasons not to have a historical sales data for an accurate forecasting process (Thomassey 2010).

For the past few decades, traditional forecasting methods, including time series (exploratory) and regression (explanatory) techniques, have been widely used in demand forecasting. Naíve, moving average, trend, multiple linear regression, Holt–Winters, exponential smoothing and ARIMA are among these traditional techniques. In recent years, their performance has been benchmarked against advanced ML techniques, which have gained attention and popularity due to the advancement in technology. The emergence of big data, cloud computing and improved computing storage and processing capabilities has led to increased availability and accessibility to large volumes of data, making ML techniques a viable option for demand forecasting in the industry.

Literature shows that ML and predictive analytics provide an advantage over traditional forecasting methods that use only limited demand factors to create more accurate demand forecasts. Traditional forecasting methods usually only take into account a single factor or at most a few factors such as trend, seasonality and cycle, so part of the variation remains unexplained in the forecasting model when in fact there may be patterns undiscovered. ML based forecasting combines learning algorithms to identify underlying demand drivers and uncover insights by processing an excessive number of predictor variables, and determining the ones that are significant. The data source for traditional demand forecasting is restricted to only the demand history, while ML based techniques can make use of various data sources and take advantage of limitless data determining what's significant, then prioritizing available consumer insights (demand sensing) to influence future demand using "what if" analysis (demand shaping). Moreover, considering different type of data such as numeric/categorical/nominal data, the need of converting qualitative data into quantitative, aggregating of data are not possible for time series data. While modeling the non-linear relationships between variables, ML outperforms with the help of technological developments. In traditional approaches, multiple single-dimension algorithms are used separately for different product styles or categories based on different data constraints. Thus, more manual data manipulation and cleansing work is required and the algorithms are less generalizable. In ML, an array of general algorithms is used to fit demand patterns across the entire product portfolio, creating a synchronized and integrated forecast (Chase 2016).

Different techniques have been applied to demand forecasting in literature for the fashion industry, including neural networks, support vector machine (SVM), fuzzy inference system, extreme learning machine, extended extreme learning machine, deep neural networks, harmony search algorithm and grey method (Thomassey et al. 2005; Das and Chaudhury 2007; Sun et al. 2008; Au et al. 2008; Carbonneau et al. 2008; Thomassey 2010; Wong and Guo 2010; Choi et al. 2011; Xia et al. 2012; Choi et al. 2014; Lu 2014; Kaya et al. 2014; Brahmadeep and Thomassey 2016; Pillo et al. 2016; Hui and Choi 2016; Loureiro et al. 2018). In addition, a hybrid combining different techniques tend to perform better than a single method (Wong and Guo 2010; Choi et al. 2014). Some papers investigated this problem in perspective of inventory management (Kogan and Herbon 2008; Mostard et al. 2011). Studies focusing on the fashion industry are limited due to the lack of large historical data set, and the findings are sometimes contradictory. Many authors considered long manufacturing and shipping lead times, seasonal effects, sales promotions, the purchasing power of customers that make the forecasting process extremely complicated in the fashion retail industry. The problem is being more complicated when it comes to adding a new request related to fashion trends. Throughout the fashion supply chains including many entities such as raw materials suppliers, manufacturers, distributors and retailers, the orders being placed should be well organized depending on demand level of the items. If a company wants to launch a new seasonal product with a different style as a certainly new item, both predicting

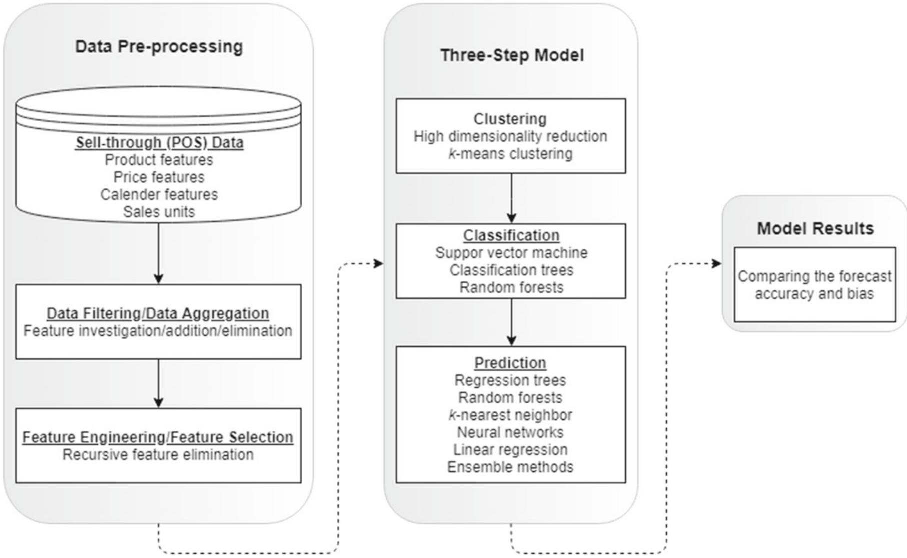


Fig. 1 The proposed methodology

the sales and ordering the raw materials to their suppliers are arising as complex issues to solve.

This study intends to contribute to the academia and industry by proposing a demand forecasting approach as an agile supply chain strategy that meets today's challenges. The distinguishing part of the study is recommending a prediction system for newly launched products, which have no time series data, based on the similarities considering various features of previous collections. Our purposes are, first, to propose a demand forecasting process for newly launched products having no historical data, second, to leverage ML techniques to recommend solutions for an improved prediction accuracy, and third, to maximize the utilization of the point-of-sale (POS) data to unveil new insights that contribute the company's strategic objectives. This study explicitly expresses the features as predictor variables, which are used for fashion retailers and presents the features' transformation to give a perspective to academics and industry practitioners.

3 The proposed methodology

This section demonstrates the steps of the proposed methodology. It presents both collecting data to identify significant features and building a model to forecast the demand for newly designed and launched seasonal products, which have no sales history. Due to the lack of sales information of newly designed products, this methodology provides a visibility for hidden features to discover look-alike products and enrich the forecasting process as much as possible. Figure 1 shows the methodology defining the scope and granularity of the data involved, describing the process of feature engineering and selection, building a three-step forecasting model which uses various techniques for different purposes, and finally comparing the results of the techniques. Each stage is explained in the following parts.

3.1 Data pre-processing

The proposed methodology was applied to a data set of a leader company of the apparel and footwear industry in USA. Two types of data were collected from the company as sell-in (shipment) and sell-through (POS) data. The POS data collected was at daily-style-location level from 115 retail outlet stores. However, since style-level was considered, not all stores sold the same style, so that some styles were sold in one store only while others were sold in many stores. The original data set included features in terms of *product*, *calendar*, *store*, *price and promotion*, and *sales units*. The total number of records in the sell-through data was 13,295,485, spanning a total of nine and a half seasons from July 2013 to March 2018. The Spring/Summer season consists of January to June while the Fall/Holiday season consists of July to December. Since this study focused on determining of the number of each style to order from the manufacturer for the whole season, the data were aggregated to the level at which this decision is made; i.e., across all stores at the monthly level. The list of all features is shown in Table 1.

The products sold at outlet stores may either be discounted products from regular inline stores or products made exclusively for launching at the outlet stores. In the context of demand forecasting, we were only interested in the latter category. Products with excess inventory after the intended product lifecycle are discounted and this distorts the demand. Meanwhile, the data showed that seasonal products live in the market for 4 months on average. Therefore, only records for outlet-exclusive products with full-price status and a product lifecycle of 1–4 months were included in our analysis.

Some features were modified, dropped or added by the processes of both filtering and aggregating in data pre-processing step. For example, in order to make the *colour group* feature more meaningful, colours were aggregated into groups based on similarities. Apart from this, the list was investigated and the features which had same data under different features were dropped off the list not to lead a confusion. Due to the correlation between *pillar* and *category* features, *pillar* was dropped as *category* already captured the same information. Besides *retail outlet sub-department* feature was excluded regarding that only seasonal styles would be evaluated. Another feature *store count* was added to refer to the number of stores at which a style was sold. As seasonal styles were launched at different times of the year with short lifecycles, the sales were dependent on the lifecycle features in addition to the calendar features; i.e., sales were not only related to which calendar month the sale occurs in, but also to which month the product was launched. Three new features were introduced for the data analysis as *lifecycle length* which refers to the length of the lifecycle of a product style in number of months, *lifecycle month* which refers to the number of months since product launch, and *lifecycle start month* which refers to the month when the product style is launched.

After the process of filtering, aggregating and adding new features, the data was partitioned into three sets as train, validation, and test by considering the overlaps caused by the company's fiscal year which was between June–May. In this respect, each season is divided into quarterly periods such as (i) January–March, Spring (ii) April–June, Summer (iii) July–September, Fall and, (iv) October–December, Holiday. The train set included all the sales records occurred before fiscal year 2017, except for products with sales overlap in both fiscal years 2016 and 2017, which were allocated to the validation set. The validation set covered the sales records in fiscal year 2017 and the overlap from 2016 plus 7 months of records from fiscal year 2018. The test set included three months of records from fiscal year 2018. Table 2 gives an overview of the data separation was complied with the quarterly split for the three-step model.

Table 1 List of features

Feature category	Feature	Description	Notes	Actions on feature
Meta data	Style	Unique style number of each product	740 (# of styles)	1327 (Original # of styles before data pre-processing)
Meta data	Style description	Description of the style	740 (# of styles)	
Calendar	Year	Fiscal year	2013–2018	
Calendar	Month	Fiscal month	12 months	
Product	Color group	Color code	16 unique values (black, blue, yellow, etc.)	
Product	Basic material	Type of material	12 unique values (canvas, cotton, leather, etc.)	
Product	Gender	Gender or age group description	7 unique values (men, women, unisex, infant, etc.)	
Product	Category	Product family	5 unique values	
Product	Sub-category	Classic versus modern	2 unique values	
Product	Cut	Ankle height	5 unique values	
Product	Product class	Product main feature	7 unique values	
Product	Pillar	Product sub-brand basic versus seasonal	3 unique values	Dropped in data pre-processing as it correlates with "Category"

Table 1 continued

Feature category	Feature	Description	Notes	Actions on feature
Product	Retail outlet sub-departments		3 unique values (core, semi-core, rebuy)	Dropped in data pre-processing as only Rebuy (fashion products) were considered
Price and promotion	Price status	Full price versus markdown	2 unique values (FP, MKD)	Dropped in data pre-processing as only FP (full price products) were considered
Price and promotion	Manufacturer's suggested retail price (MSRP)	Ticket price	Range \$25–\$110 [\$53.94 ± \$14.66]	
Price and promotion	Average unit retail (AUR)	Actual selling price	Range \$14.29–\$88.49 [\$36.36 ± \$11.16]	
Lifecycle	Lifecycle length	The total number of months in the lifecycle of a style	4 unique values (1–4)	Added in data pre-processing step
Lifecycle	Lifecycle month	The number of months since product launch	12 unique values (1–12)	Added in data pre-processing step
Lifecycle	Lifecycle start month	The month at which the lifecycle started	12 unique values (1–12)	Added in data pre-processing step
Store	Store count	Number of stores selling a style	Range 1–115 [63.73 ± 19.61]	Added m data pre-processing step
Sales units	Retail sales units (target variable)	Retail sales units	Range 1–7167 [806.17 ± 760.41]	

Table 2 Overview of data set partitioning

Data set	Number of months for sales	Number of styles	Percentage (%)
Train	35	539	67
Validation	19	201	25
Test	3	58	8

Before implementing ML models, feature engineering and feature selection processes were applied to decide on valuable features. Feature engineering covers the transformation of raw data into suitable features to model, whereas feature selection involves removing unnecessary features. We used recursive feature elimination (Guyon and Elisseeff 2003), a backward feature selection method, to eliminate features based on their contribution to improving forecast accuracy. A random forests algorithm, which is frequently used with backwards selection, was used on each iteration to evaluate the model with different subsets of all features. Random forests was selected in view of its capability in handling multicollinearity. The combination of the recursive feature elimination and random forests increases the performance of feature selection since recursive feature elimination attempts to decrease the importance of correlated variables and to eliminate dependencies when multicollinearity occurs. A tenfold cross-validation on the train data was used. There are many advantages of feature selection such as facilitating data understanding, reducing the measurement and storage requirements, reducing training and utilization times, coping with dimensional by selecting subsets of features that are useful to build a good predictor (Guyon and Elisseeff 2003). Feature engineering and selection process helped determine the proper features to be used in each step of the model which is explained in the following.

3.2 Three-step model

The three-step model consists of three separate partitions defined as (i) *clustering*, (ii) *classification*, and (iii) *prediction* (see Fig. 1). The main objective behind this model is to identify the look-alike group of products from the train set. Once these products are identified, their average sales can be used as a proxy to forecast the sales for brand-new products with both the validation and test sets. Due to the dissimilar purpose of each step, different combinations of data features were used across the whole methodology.

The *clustering* step grouped all the styles in clusters based on similarities across different features having continuous numeric values. This step included four sub-steps as “feature selection”, “data normalization” to be avoided of high level influence of some features by scaling, “high dimensionality reduction” by t-SNE algorithm (van der Maaten and Hinton 2008) and “*k*-means clustering” (James et al. 2014) to group the data records in *k* number of clusters. As a result of feature selection, the initial features used in clustering were “retail sales units”, “lifecycle length”, “MSRP”, “AUR”, and “store count”.

Next, the *classification* step was initiated to create a link between the styles with pre-assigned clusters from the train set and brand-new styles from the validation and test sets. It had three sub-steps as “feature selection”, “classification” and “accuracy evaluation”. Categorical features were used in the classification step, except the “sales units” feature. With an outcome of the *clustering* step, *cluster numbers* were assigned to the records of both train and validation sets as a new feature. In this step, classification trees (James et al. 2014), random forests (Breiman 2001), and SVM (Pillo et al. 2016) were utilized to see the best performing one.

Finally, the *prediction* step was executed to predict the future sales for the brand-new styles in the validation and test sets. This step had three sub-steps as “feature selection”, “prediction” and “test and score”. The used features were “cluster numbers”, “sales units”, “life cycle length”, and with an addition of new feature *average sales*. Data normalization was performed, algorithms of regression trees, *k*-NN, linear regression (James et al. 2014), random forests (Breiman 2001), and neural networks (Zhang 2003; Das and Chaudhury 2007) were applied in *prediction* step. Besides, ensemble methods taking the median and average of the outputs from the five individual methods were also considered and the performance was evaluated in terms of accuracy.

Briefly, the features of *product*, *calendar*, *store*, and *price and promotion* were shared by the company, whilst the features of *lifecycle*, *cluster numbers* and *average sales* were created in different steps through the three-step model process.

3.3 Performance measurement

A well-predicted value means a well-performed model. The term *accuracy* refers to how the model is able to fit the current observations and the term *bias* determines the direction of difference between observed and predicted values, whether it is over or under forecast. We measured forecast accuracy using *Weighted Mean Absolute Percentage Error* (WMAPE) and forecast bias using *Weighted Mean Percentage Error* (WMPE). The weighted average of the errors avoids the mistakes caused by both the magnitude and the granularity of the data. Equations 1–5 shows the calculation on accuracy and bias based forecast error.

$$Absolute\,forecast\;error = |forecasted\;sale - actual\;sales| \tag{1}$$

$$Absolute\,forecast\;percentage\;error\;(APE) = \frac{|forecasted\;sales - actual\;sales|}{actual\;sales} \tag{2}$$

$$Forecastaccuracy\;(WMAPE) = \frac{\sum_{i=1}^{n} APE \times actual\;sales}{\sum_{i=1}^{n} actual\;sales} \tag{3}$$

$$Forecast\;percentage\;error\;(PE) = \frac{forecasted\;sales - actual\;sales}{actual\;sales} \tag{4}$$

$$Forecast\;bias\;(WMPE) = \frac{\sum_{i=1}^{n} PE \times actual\;sales}{\sum_{i=1}^{n} actual\;sales}. \tag{5}$$

4 Results

4.1 Clustering and classification results

Silhouette score is used to determine an appropriate number of clusters that fit the data used (Rousseeuw 1987). It is a method of interpretation and validation of consistency within clusters of data by measuring and comparing the mean intra-cluster distance to the mean nearest-cluster distance for each data point within a cluster. The silhouette score ranges between -1 as *wrong clustering*, $+1$ as *best value*, and 0 indicating *overlapping clusters*. Due to the difficulties in determining the number of clusters in a data set, in addition to the silhouette score assessment, we took another measure to verify what number of clusters was proper for this data set. Using the both train and validation sets, we ran a classification

Table 3 Comparison for the silhouette scores for different *k* number of clusters and overall classification accuracy by different algorithms

Comparison of silhouette scores		Comparison of overall accuracy by different algorithms		
Number of clusters	Silhouette score	SVM (%)	Classification trees (%)	Random forests (%)
2	0.378	85	83	88
3	0.422	88	75	84
4	0.481	88	86	88
5	0.469	93	69	89
6	0.487	91	83	87
7	0.489	86	76	76
8	0.463	78	71	73
9	0.441	80	71	75
10	0.427	78	72	70

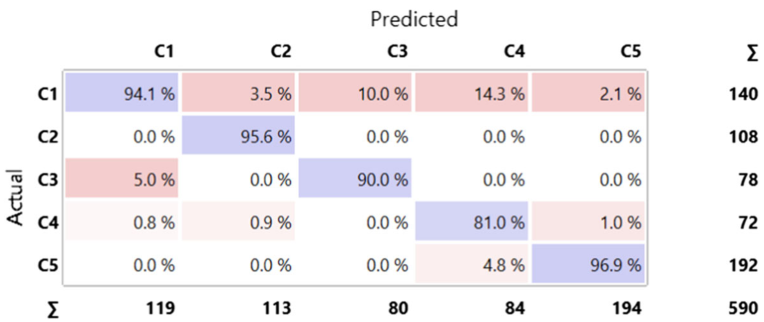


Fig. 2 Confusion matrix resulted from using five clusters and SVM

exercise for each *k* and then compared the clusters regarding the overall accuracy that resulted from different algorithms. Table 3 compares the silhouette scores of the different number of clusters the overall classification accuracies based on SVM, classification trees and random forest algorithms used. The number of clusters that revealed the highest silhouette score was seven as illustrated in Table 3, whereas the number of clusters that revealed the best classification match was five with an overall accuracy of 93% resulted in the best performing classification algorithm SVM. Considering the silhouette score for five versus seven clusters, the difference was minimal.

Figure 2 shows the classification results using five clusters and SVM algorithm. The vertical axis represents the number of records that were pre-assigned to each of the five clusters (*C1*–*C5*) based on clustering while the horizontal axis is the number of records allocated to each of the five clusters based on classification.

Figure 3 depicts data distributed in five clusters and Table 4 shows the detailed analysis which were performed on the validation set to better understand clusters' allocation based on *lifecycle length*, *sales volume*, *AUR*, *MSRP* and *store count* with the mean and standard deviation values.

Lifecycle length is a clear distinguishing driver among the clusters. Each cluster included the items with one lifecycle length, except *C4*, which had the styles with multiple lifecycles.

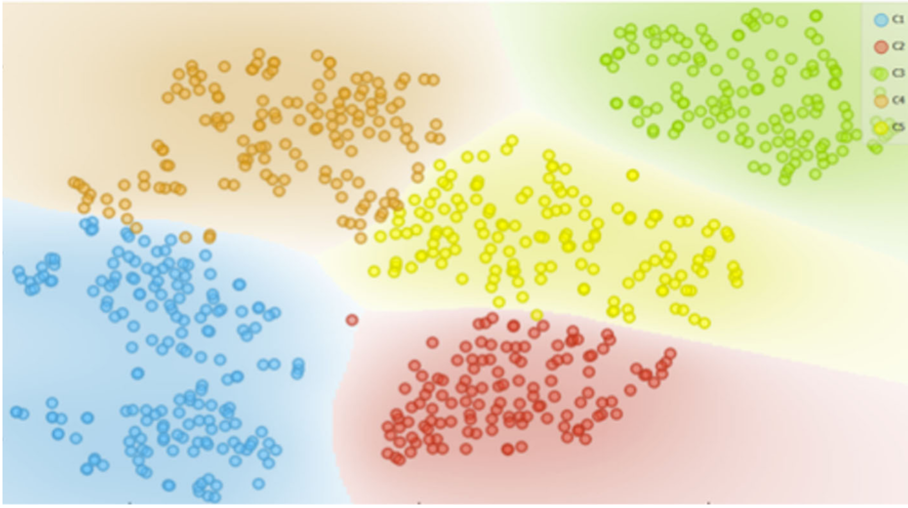


Fig. 3 Data distributed in five clusters

Table 4 Characteristics of clusters with mean and standard deviation values

Cluster	Lifecycle length (months)	Sales units (monthly)	AUR	MSRP	Store count
C1	3	1298 ± 864	\$37.69 ± 6.15	\$62.22 ± 8.49	88 ± 11
C2	2	953 ± 722	\$30.02 ± 8.38	\$49.86 ± 12.41	79 ± 21
C3	3	839 ± 684	\$23.79 ± 3.78	\$39.36 ± 5.39	83 ± 16
C4	2, 3, 4	463 ± 414	\$43.77 ± 10.14	\$67.88 ± 11.99	37 ± 30
C5	4	958 ± 640	\$31.05 ± 9.77	\$51.19 ± 15.21	82 ± 16

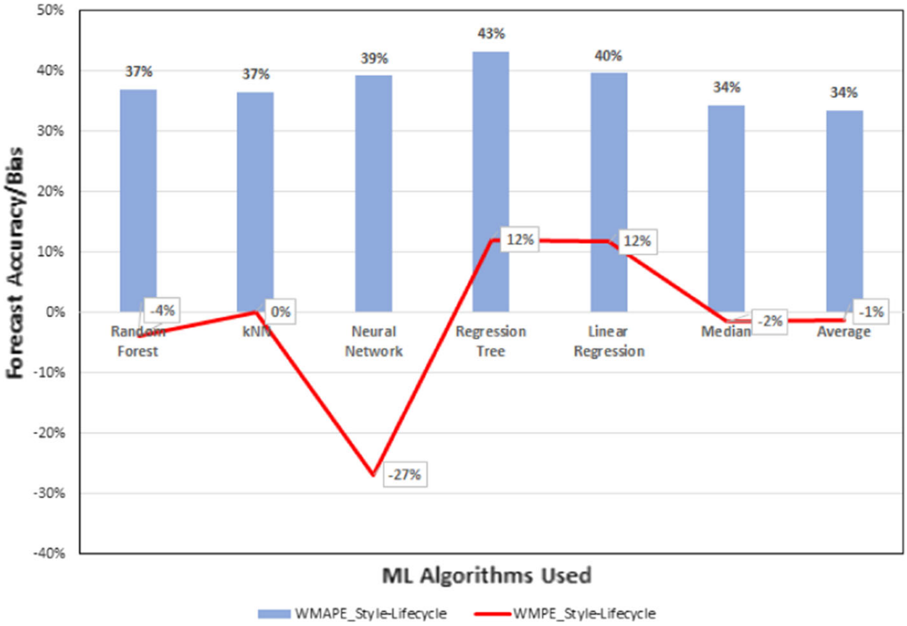
However, *C4* seemed to have relatively lower sales volume, lower store count, and higher AUR compared to the other clusters.

Both clusters *C1* and *C3* included the seasonal styles with a 3-months lifecycle length. However, *C3* had lower sales volume and lower AUR compared to *C1*.

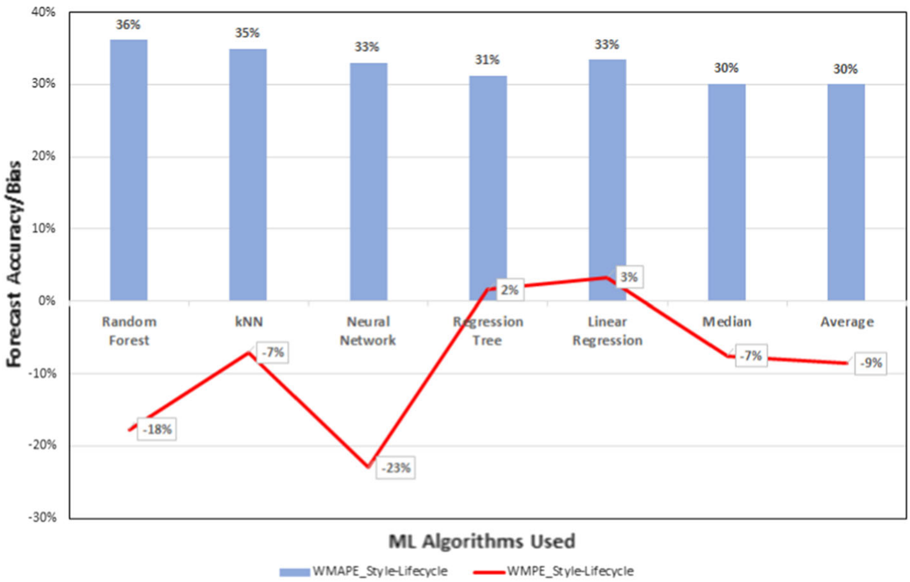
4.2 Prediction results

Running the prediction algorithms on both validation and test sets had relatively different results regarding the best performing algorithms. As seen throughout the study, lifecycle length is a distinguished feature for seasonal products. Generally the entire purchase quantity in the company is confirmed prior to the beginning of the season considering the product lifecycle length which is around 1–4 months. Therefore, we investigated the forecast accuracy for the whole season instead of individual month. However, the forecast accuracy of the ensemble methods for both data sets was much closer. Overall, the test set had slightly better forecast accuracy and worse forecast bias compared to the validation set. Figure 4 displays the accuracy and bias of forecast for both validation and test sets.

Prediction results are evaluated regarding to performance measures. At the graph of the validation set, both *k*-NN and random forests delivered the highest forecast accuracy on a *style-lifecycle length* level with a WMAPE of 37%. However, *k*-NN had no forecast bias



(a) Validation set



(b) Test set

Fig. 4 Forecast accuracy and bias of the proposed methodology for validation and test sets

Table 5 Prediction results on cluster-level

Cluster characteristics	Best performing algorithm	WMAPE	WMPE
Mono lifecycle-high average sale (C1)	*k*-NN, linear regression	28	+4/− 11
Mono lifecycle-medium average sales (C2, C3, C5)	Random forests	32/37	− 11/+6
Multiple lifecycle-low average sales (C4)	Regression trees	39/45	− 30/0

compared to 4% under-forecast bias by random forests. Neural networks were third-best in forecast accuracy on with 39% WMAPE and worst in under-forecast bias with 27% WMPE. Finally, regression trees had 43% of WMAPE and 12% over-forecast bias. The two ensemble methods, average and median, delivered the best overall results in forecast accuracy on *style-lifecycle* (34%) levels with a under-forecast bias around 1%. For the test set, regression trees delivered the best performance with a forecast accuracy of 31% and 2% of over-forecast bias regarding *style-lifecycle* level. Linear regression showed a very close performance with 31% of forecast accuracy and 3% of over-forecast bias. The neural networks' performance in the test set was slightly better compared to the validation set with a forecast accuracy of 33%. However, the neural networks' forecast bias 23% under-forecast was still relatively high compared to the other four algorithms. Random forests and *k*-NN had the lowest forecast accuracy around 36% while they had a forecast bias of 7% and 18% under-forecast, respectively. Like in the validation set, the ensemble methods also delivered the highest overall forecast accuracy with 30% WMAPE on *style-lifecycle* level. However, their forecast bias was relatively high around 8% under-forecast compared to the regression trees and linear regression algorithms.

In addition to the analysis above, we broke down into cluster level to determine which algorithm performs well with which cluster characteristics regarding *style-lifecycle* attribute. According to validation set, random forests was the best performer in clusters *C2*, *C3* and *C5* with a forecast accuracy of 36%, 37% and 34%, respectively. The forecast bias for random forests was − 11% (*C2*), − 1% (*C3*) and + 3% (*C5*). It is essential to note that these three clusters (*C2*, *C3* and *C5*) shared relatively similar store count and average monthly sales. The relatively bad performance in *C4* could be linked to the complexity of this cluster including multiple lifecycle lengths, low monthly sales volume and high AUR on average. The test set included three clusters which resulted from the classification stage: *C1*, *C3* and *C4*. Similar to the validation set, random forests and regression trees delivered the highest forecast accuracy and lowest forecast bias in clusters *C3* and cluster *C4*. In cluster *C1*, *k*-NN revealed the best forecast accuracy (28%) and lowest forecast bias (4% over-forecast). On a cluster level, the performance of the ensemble methods did not outperform the individual algorithms. While the individual ones as regression trees and linear regression models had the high forecast accuracy of > 35 % with lowest bias $< 5\%$, the ensemble methods obtained forecast accuracy of 30% with low forecast bias $< -10\%$. Table 5 shows a brief of all the cluster characteristics and best performing algorithms.

4.3 Practical results

This study aims to answer the question of demand forecasting optimization for fashion retailers who lack an agile supply chain strategy that meets today's challenge. The proposed methodology leverage machine learning to introduce a scientific forecasting framework for newly launched products with improved prediction accuracy and mitigated bias using historical POS data while creating visibility for veiled features. A scenario analysis study with different percentages for both inventory leftover and inventory holding cost was run to compare the company's actual unit profitability to the expected unit profitability when the three-step model was used in order to place the optimal inventory orders. Our findings that help company reshape the strategic decisions are summarized as follows.

- The seasonal company products live in the market for 17 months, while the product lifecycle expectancy by the company is between 5 and 6 months due to the over-forecasting. The buyers tend to order 2 or 3 times more of what they can sell in one season including 3 months. This leads to 10% of inventory leftover by the end of the product lifecycle after the season is over with both an aggressive price and margin discount by 50% and 72%, respectively.
- The company can avoid over-forecasting and predict actual demand at an accuracy ranging between 65% and 70% with a forecast bias ranging between −5% and +5% via the three-step model. The use of this forecast accuracy along with an inventory policy, which takes into account the overestimation cost compared to the inventory shortage, should result in a 40–50% reduction in inventory as well as an additional 15–40% profit per unit mainly due to less inventory costs. More accurate forecast led to lower inventory level.
- The clustering process resulted in five unique clusters. By taking a closer look inside each of the clusters, we found out that lifecycle length was the main distinguishing factor from one cluster to another. The number of stores and retail prices also played a role in clustering. We also understand how the features we used differed by some factors, e.g., historical sales of similar products came first followed by the number of stores, color, gender, retail price, and material.
- The stakeholders can use the findings above to create the product and assortment segmentation strategies based on price elasticity and product feature popularity with some additional analysis.

5 Conclusions

Forecasting influences the whole supply chain operations in a company and it is affected by not only the sales pattern but also the various implicit features related to the products. This study proposes a data-driven methodology based on machine learning techniques to optimize demand forecasting process for newly launched products in fashion retailing. The three-step methodology involving *clustering*, *classification* and *prediction* enables further visualization of the relationship between significant features and allow for customization of the forecasting approaches accordingly.

As in all machine learning studies, data pre-processing of the proposed methodology is an important step in facilitating input formation. The feature engineering process helps create new features that bring additional value to demand interpretation, and the feature selection process allows us to gain insights into the importance of the different features and their influence on forecast accuracy. Another value proposition of this phase is the possibility of using, processing and delivering value out of the categorical features that have always been considered a challenge when it comes to forecasting demand in fashion supply chains. In evaluating the model performance for the sponsoring company, the ease of implementation was considered in addition to the models' predictive performances.

The three-step model provides visibility of underlying factors that affect demand through *clustering* and *classification*. With this model, *prediction* can be customized to give the best results based on product features such as lifecycle length, sales units, retail price, and store number. We recommend regression tree approach to apply to complex clusters with multiple lifecycle lengths. Random forests is the algorithm that we suggest using in clusters with mono-lifecycles, while k-NN and linear regression are the ones for similar clusters with higher sales volume and AUR.

Due to the limitation in the existing inventory data, the lost sales were not considered in building our prediction models. Inventory data was provided at the monthly and style level. As a result, we only had one snapshot of inventory level per month and a total level of all sizes of a style. Therefore, it is not possible to estimate lost sales required to reflect the actual demand not captured in POS data.

The proposed methodology is applicable and supportive for decision makers in demand forecasting of newly launched products without any historical data for fashion supply chains. However, still the features used and the machine learning techniques preferred are the open questions for future works. Surely, the opportunity of having and tracking of large and varied real-time data sets provides a different perspective for forecasting process. Therefore, this area is very promising and attractive in addition to many challenges.

References

Au, K. F., Choi, T. M., & Yu, Y. (2008). Fashion retail forecasting by evolutionary neural networks. *International Journal of Production Economics*, *114*(2), 615–630. https://doi.org/10.1016/j.ijpe.2007.06.013. http://www.sciencedirect.com/science/article/pii/S0925527308000443. special Section on Logistics Management in Fashion Retail Supply Chains.

Brahmadeep, & Thomassey, S. (2016). *Intelligent demand forecasting systems for fast fashion* (pp. 145–161). https://doi.org/10.1016/B978-0-08-100571-2.00008-7.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324.

Carbonneau, R., Laframboise, K., & Vahidov, R. (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, *184*(3), 1140–1154. https://doi.org/10.1016/j.ejor.2006.12.004.

Chase, C. W, Jr. (2016). Machine learning is changing demand forecasting. *Journal of Business Forecasting*, *35*(4), 43–45.

Choi, T. M., Hui, C. L., & Yu, Y. (2011). Intelligent time series fast forecasting for fashion sales: A research agenda. In *International conference on machine learning and cybernetics, ICMLC 2011, Guilin, China, July 10–13, 2011, Proceedings* (pp. 1010–1014). https://doi.org/10.1109/ICMLC.2011.6016870.

Choi, T. M., Hui, C. L., Liu, N., Ng, S. F., & Yu, Y. (2014). Fast fashion sales forecasting with limited data and time. *Decision Support Systems*, *59*, 84–92. https://doi.org/10.1016/j.dss.2013.10.008.

Das, P., & Chaudhury, S. (2007). Prediction of retail sales of footwear using feedforward and recurrent neural networks. *Neural Computing and Applications*, *16*(4), 491–502. https://doi.org/10.1007/s00521-006-0077-3.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of Machine Learning Research*, *3*, 1157–1182.

Hui, P., & Choi, T. M. (2016). 5-Using artificial neural networks to improve decision making in apparel supply chain systems. In T.M. Choi (Ed.) *Information systems for the fashion and apparel industry. Woodhead publishing series in textiles* (pp. 97–107). New York: Woodhead Publishing. https://doi.org/10.1016/B978-0-08-100571-2.00005-1. http://www.sciencedirect.com/science/article/pii/B9780081005712000051.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2014). *An introduction to statistical learning: With applications in R*. Berlin: Springer.

Kaya, M., Yeşil, E., Dodurka, M. F., & Sıradağ, S. (2014). *Fuzzy forecast combining for apparel demand forecasting* (pp. 123–146). Berlin: Springer. https://doi.org/10.1007/978-3-642-39869-8_7.

Kogan, K., & Herbon, A. (2008). Production under periodic demand update prior to a single selling season: A decomposition approach. *European Journal of Operational Research*, *184*(1), 133–146. https://doi.org/10.1016/j.ejor.2006.11.009.

Loureiro, A., Miguéis, V., & da Silva, L. F. (2018). Exploring the use of deep neural networks for sales forecasting in fashion retail. *Decision Support Systems*, *114*, 81–93. https://doi.org/10.1016/j.dss.2018.08.010.

Lu, C. J. (2014). Sales forecasting of computer products based on variable selection scheme and support vector regression. *Neurocomputing*, *128*, 491–499. https://doi.org/10.1016/j.neucom.2013.08.012.

van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Mostard, J., Teunter, R., & de Koster, R. (2011). Forecasting demand for single-period products: A case study in the apparel industry. *European Journal of Operational Research*, *211*(1), 139–147. https://doi.org/10.1016/j.ejor.2010.11.001.

Pillo, G. D., Latorre, V., Lucidi, S., & Procacci, E. (2016). An application of support vector machines to sales forecasting under promotions. *4OR*, *14*, 309–325.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, *20*, 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sun, Z. L., Choi, T. M., Au, K. F., & Yu, Y. (2008). Sales forecasting using extreme learning machine with applications in fashion retailing. *Decision Support Systems*, *46*(1), 411–419. https://doi.org/10.1016/j.dss.2008.07.009.

Thomassey, S. (2010). Sales forecasts in clothing industry: The key success factor of the supply chain management. *International Journal of Production Economics*, *128*(2), 470–483. https://doi.org/10.1016/j.ijpe.2010.07.018.

Thomassey, S., Happiette, M., & Castelain, J. M. (2005). A short and mean-term automatic forecasting system-application to textile logistics. *European Journal of Operational Research*, *161*(1), 275–284. https://doi.org/10.1016/j.ejor.2002.09.001.

Wong, W., & Guo, Z. (2010). A hybrid intelligent model for medium-term sales forecasting in fashion retail supply chains using extreme learning machine and harmony search algorithm. *International Journal of Production Economics*, *128*(2), 614–624. https://doi.org/10.1016/j.ijpe.2010.07.008.

Xia, M., Zhang, Y., Weng, L., & Ye, X. (2012). Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs. *Knowledge-Based Systems*, *36*, 253–259. https://doi.org/10.1016/j.knosys.2012.07.002.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175.