New Product Performance Prediction in Fashion Retailing

Ruthvik Arepalle, Jeffrey Chittaranjan, Shruti Gupta,
Anjali Hegde, Gajender Saharan, Matthew A. Lanham
Purdue University, Department of Management, 403 W. State Street, West Lafayette, IN 47907
rarepall@purdue.edu; jchittar@purdue.edu; gupta592@purdue.edu;
hegde11@purdue.edu; gsaharan@purdue.edu; lanhamm@purdue.edu
March 5, 2020

ABSTRACT

Predicting the performance of a newly launched product in the fashion retail industry is one of the most crucial and challenging issues faced by retailers. Given the fast-changing customer preferences, high lead time, efficiencies in bulk orders and complexities in global supply chains, most retailers place one-time orders based on demand forecasts. Hence, retail category managers need to assess and re-evaluate the likelihood of success of newly launched products in the first few weeks after the product's launch. In our study, we have developed a two-step machine learning decision support system which helps in forecasting sales of new launches during the initial three months using first three weeks of sales using historical data of similar products. In the first step, we have used clustering algorithms to group patterns based on their characteristics and sales metrics in the first three weeks. Then, we have used the clusters as features along with other characteristics like new product's attributes (colour, patterns, size), sales across geographies in the first three weeks, merchandising features, cannibalisation features and sentiment scores to train our model. We have also compared the forecasts from different machine learning models like Linear Regression, Decision Trees, Random Forests and XGBoost. Our model can help category managers to take strategic decisions such as devising promotional strategies, achieving higher margins through better pricing, changing product allocations across geographies and channels to ensure the product's success and increase the bottom line.

Keywords: fashion retailer, machine learning, successful products, promotional changes, social media, new launch, sales forecasting, product performance

1. INTRODUCTION

Inventory Management plays a key role in the profitability of a fashion retail company. Companies conduct market research surveys to understand consumer buying behaviour and compare consumer preferences across similar products launched in the past to forecast the sales of a new product launch. Despite their best efforts, it is difficult to forecast 100% accurate results. The need to respond to fast fashion trends shortens the life cycle of retail fashion products. This phenomenon makes forecasting much more complex. Currently, most fashion retailers liquidate unsold products by offering huge discounts and sales promotions which can affect their brand image and hit their profit margins. According to a 2018 report by Select and Coresight Research, markdowns cost US non-grocery retailers $300 billion in revenue annually. According to a New York Times report, luxury retailers are under pressure to burn unsold inventory to protect their brand image.

These non-biodegradable substances lead to environmental pollution.

In fashion retail, most promotional offers and investments for a newly launched product are made well in advance. If a product does not perform as expected during the early weeks of launch, the methodology detailed in the study can help the category manager learn and implement promotional or location changes to help the product sell better in the remaining duration of its lifecycle. It can also help the product achieve its targeted sale by maintaining a balance between margins and volumes.

In collaboration with a fashion retailer, we have used point of sales (POS) and online (Web) transaction-level data, store information data and social media sentiment analysis scores to develop regression models which can identify whether a new product's success can be determined based on the first three weeks of sales. Since retailers launch multiple products in a year, similar products launched in the earlier time period can cannibalise sales of a new product. This event has also been factored as a feature by including the number of products and colours/patterns launched across months and seasons before the launch of the new product. A new product's success criteria have been defined as selling more units than expected by the third month of its lifecycle. On using our model, if the category manager estimates that a newly launched product will not sell as per the expectations, then decisions can be taken to either reduce margins or place inventory in locations where the product is successful.On the other hand, if the product is expected to be successful, then the product can be sold at the existing price without affecting the current margins. Since our data had hundreds of patterns and interpretability of the model was a key expectation, we also clustered patterns based on the first three weeks sales metrics and product characteristics using k-means clustering. The details of the product lifecycle are shown in Figure 1.

Our study is novel in the use of the machine and deep learning models in the fashion retail business. The business has historically been heavily domain knowledge-driven and has recently seen a massive investment into analytical decision-support, which includes integrating social media into predictive models that support category managers in decision making. We discuss feedback received from these decision-makers and how our solution will continue to evolve and support them.

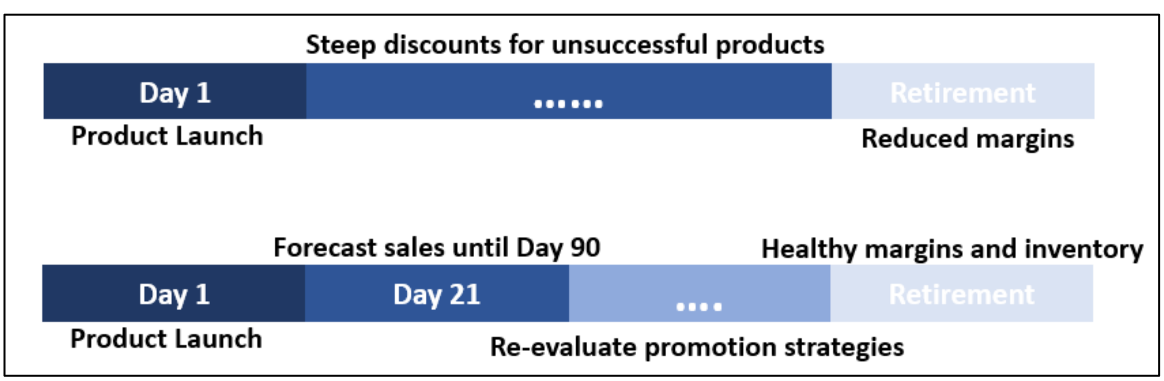


Figure 1: Current and proposed product lifecycle timelines

The remainder of this paper is structured as follows: **Section 2** provides a comprehensive literature review about this topic, **Section 3** describes a framework about data collection, preprocessing, methodology adopted to obtain sentiment score of the reviews, forecasting models and performance criteria. **Section 4** provides forecasting results and comparison of the models used. **Section 5** discusses the conclusions and limitations of this study and provides a further scope of this research.

2. LITERATURE REVIEW

Over the years, researchers have developed various analytical techniques to forecast sales in the retail industry. Of all of them, time series forecasting methodologies like ARIMA, SARIMA, Exponential Smoothing and Box Jenkins have been amongst the most extensively used toolset (A.L.D. Louriero, 2018). These methods need aggregated time series data and are based on future projections of the past data. However, these do not apply to newly launched products due to lack of historical data, the shorter lifecycle of the products and irregular demands.

In one study, (Zhi-Ping Fan, 2017) used Bass/Norton model combined with sentiment analysis of online reviews to forecast the sales of products. Bass model is a diffusion-based model that considers all products having innovators and imitators. However, fashion products have a short lifecycle due to which such models cannot be applied. For fashion retail, (Pawan Kumar Singh, 2019) used tree-based and deep learning models due to their ability to model short-lived interactions and non-linear relationships between predictors and response variables. They used product attributes, temporal and engineered features to train the model. The features were engineered based on fashion, merchandising and derived factors like cannibalisation, the shelf life of a style, trends and seasonality. They observed long-tailed behaviour in the data, a characteristic of retail and, applied transformations at different scales to reduce the variance. They concluded that DNN models – LSTM and MLP, underperformed as compared to tree-based models. XGBoost, when used along with careful feature engineering, yielded good accuracies for demand forecasting of newly launched products.

In another study, conducted by (A.L.D. Louriero, 2018), deep learning model was trained using historical sales data of the products belonging to homologous seasons. Product characteristics corresponding to logistical and internal organisational aspects of the company, and opinion of domain experts were used as independent variables. DNN approach demonstrated promising results when there was a lack of historical data. However, the training process was relatively complex as compared to shallow techniques, specifically, Random Forest, which produced satisfactory predictive capability.

(Yanrong Ni, 2011) combined neural networks with autoregression technique in the form of a dynamic two-stage prediction model for fashion retail forecasting. They used neural networks to establish a multivariable error prediction model. The model developed the concept of "influence factors" and divided "impact factors" into two distinct stages (long and short term). The results showed that the multivariate error prediction model could produce good results for sale forecasting problems in fashion retail.

In a similar study published by (Asli Aksoy, 2012), in their research, they combined neural networks and fuzzy method to create a new model called 'fuzzy inference system' based on an adaptive network. The proposed model combined the advantages of both the techniques, namely generalisation ability of the fuzzy logic technique and the learning ability of neural networks, generating a powerful hybrid model. (Tsan-Ming Choi, 2012) applied a grey model (GM) and ANN-based hybrid model to forecast sales based on the base colour of products. They analysed the changing regime of ANN, GM, Markov, and GM + ANN hybrid models. They concluded that the GM and ANN hybrid model was the best for predicting colour fashion sales when historical data was small. Recently, (Yen-Sen Shih, 2019) proposed LSTM approach to model the forecast of short-term demand with a sentiment score of consumers from online comments, and reviews crawled from "taobao.com". To achieve high accuracy, the model was trained with different time series length and window size. The difference in accuracy was compared and tested for different time series length, window size and different weights of the sentiment scores.

In another study, (Ilan Alon, 2001) forecasted aggregate retail sales by comparing artificial neural networks and traditional methods including Winters exponential smoothing, Box Jenkins, ARIMA model, and multivariate regression. Like much other economic time series, US aggregate retail sales have shown strong trend and seasonal patterns. It is challenging to devise ways to model and forecast patterns in time-series analysis. The results indicated that on an average ANNs fared favourably compared to more traditional statistical methods, followed by the Box Jenkins model. Despite its simplicity, the Winters model was shown to be a viable method for multiple-step forecasting under relatively stable economic conditions. The derivative analysis showed that the neural network model could capture the dynamic non-linear trend and seasonal patterns as well as the interactions between them.

As discussed, in the textile industry, existing forecasting models have found to be generally unsuitable or unusable to address the constraints like a short lifetime and numerous launches of new items. So, (Sebastien Thomassey, 2005) proposed a forecasting system, based on clustering and classification tools, which performed mid-term forecasting. The proposed model was based on the existing clustering technique (k-means algorithm) and decision tree classifier to estimate sales profiles of new items which did not have historical sales data. The clustering procedure grouped similar historical items in term of sales profiles. Each cluster centre defined a sales profile called prototype. And, the decision tree linked the descriptive criteria of historical items with prototypes of sales profiles extracted from clusters. The decision tree associated future items with prototypes from their descriptive criteria. (Kejia Hu, 2017) presented an approach to forecasting ready-to-launch new products that were similar to past products. The approach fits the product life cycle (PLC) curves to historical customer order data, clustered the curves of similar products, and used the representative curve of the new product's cluster to generate its forecast. This was achieved using three families of curves to fit the PLC: bass diffusion curves, polynomial curves, and simple piecewise-linear curves (triangles and trapezoids). The goodness of fit and complexity for these families of curves were later compared. It was found that fourth-order polynomial curves provided the best in-sample fit with piece-wise linear curves a close second. The fitted PLC curves of similar products were clustered either by known product characteristics or by data-driven clustering. The key empirical finding for the large data set showed that data-driven clustering of simple triangles and trapezoids were simple to estimate and explain and perform best for forecasting. Consistent results were found with other data sets.

A business needs to do proper forecasting before developing new products, product lines and prevent spending time and money for developing a product that fails in the marketplace. In a study by (Özlem İpek KALAOGLU, 2015), they used three different quantitative forecasting models: simple moving average model, weighted moving average model and linear trend model. The models were applied by using the past sales data of a well-known retailing brand in Turkey for forecasting sales. The weighted moving average model gave the best results with the least amount of deviation.

Our problem is based on the fashion retail industry and is novel as it forecasts 3-months sales of newly launched products based on their sales in the first 3-weeks. The aim is to identify the characteristics of high performing products and differentiate them from poor performing products. While many studies have been conducted for forecasting sales and demand of newly launched products in the fashion retail, however, they lack specifics related to our problem. We have used various tree-based regression techniques for training our model.

3. DATA

The data provided to us contain online and point of sales transactions collected from several points of sales channels of a fashion retailer. The data points collected range from February 2017 to October 2019 for a wide range of categories. Each category has items launched across different patterns/colours and sizes. The number of stock-keeping units (SKU's) was very large and it was challenging to forecast the units sold in 3-weeks sales for each SKU. So, the data has been aggregated at colour and category level for forecasting purpose. The data sets provided have web-sales data and pos-sales data at the transaction level and features of each SKU (e.g. category, pattern, colour, fabrication and other product attributes). The data sets have been merged, and the variables used for analysis have been shown in Table 3. The cumulative sentiment score for the first three weeks has been provided based on the online reviews and comments. The sentiment score ranged from -1 to 1 (negative sentiment to positive sentiment). The hierarchy of the items is shown in Figure 2.

3.1 Lifecycle of Items

The general life cycle of the items showed that the new products were launched across all the categories in four seasons (spring, summer, fall and winter). A typical product stayed on the floor for 3-4 months after which the product was sold at steep discounts and pushed to retirement. The items are "one-shot" items sold only during the given period and are not replenished due to high lead time of manufacturing. Hierarchy of product attributes is shown in Figure 3 and data is summarised in Table 1.

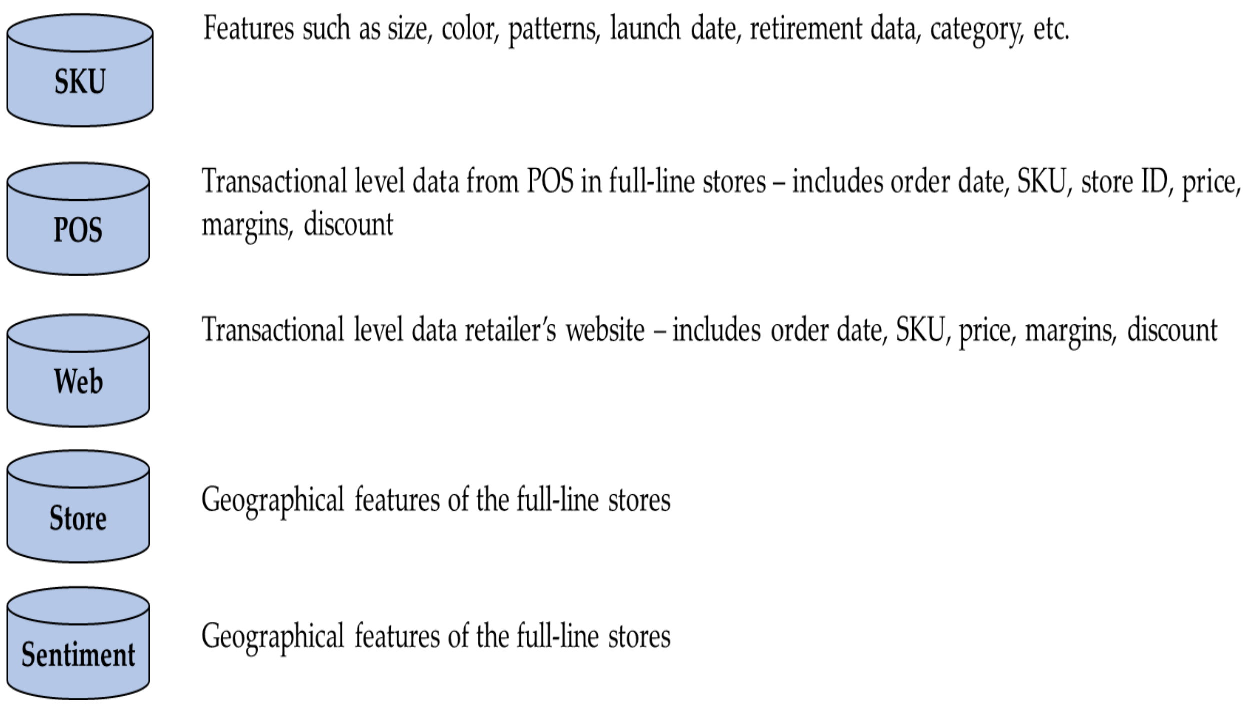


Figure 2: Schema diagram of the data

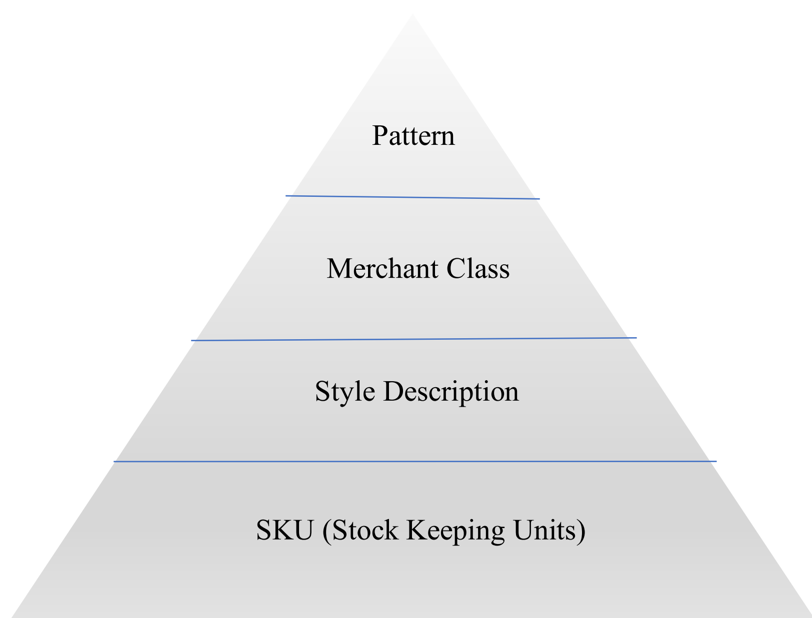


Figure 3: Hierarchy of the items

Table 1: Data used in the study

Variable Name	Variable Type	Description
# of Units Sold in 3 months	Numeric	Target Variable
Avg Price across 3 weeks	Numeric	Price ($)
Avg margins in 3 weeks	Numeric	The margin of the product ($)
Units sold in 3 weeks	Numeric	# Units sold
Cannibalisation	Numeric	The products launched in the past 3-months have some effect on the new products launched
Sentiment Scores	Numeric	Cumulative Sentiment scores in the first 3-weeks
Launch Season	Categorical	One-hot encoding (e.g. fall, winter)
Clusters	Categorical	One-hot encoding (feeding the clusters obtained from k-means clustering)
Channel	Categorical	One-hot encoding (e.g. WEB, POS)
Merchant Class	Categorical	One-hot Encoding of this variable (e.g. handbags, backpacks, winter accessories)
Geospatial Variables	Categorical	One-hot Encoding (location, e.g. MidWest, MidAtlantic)
Pattern/Color	Categorical	Patterns/Colors clustered into 3 groups
Relative Price of substitutable products	Numeric	The relative price of other products available in the same merchant class during that time

4. METHODOLOGY

Forecasts of unit sold in the first three-months are made based on the first 3-weeks of sales using Multiple Linear Regression. This benchmark model is compared against Decision Trees, Random Forest and XGBoost, which are the black box models expected to outperform the traditional methods. The data consisted of approximately 600 patterns over 30 product categories. A sample of 150 patterns has been taken for analysis that accounted for the majority of the sales. Subsample analysis has been carried out for the patterns where cumulative sentiment scores of first 3 weeks online reviews are provided.

In order to interpret the patterns easily, the patterns are clustered using the K-means clustering algorithm. Attributes such as the number of styles across which a pattern is launched, the number of SKUs, average price, total sales and number of units sold have been considered as inputs. The data points of the first three weeks sales are used for the clustering.

A 70/30 split of train and test data is used for building the model. The clusters obtained are fed into the model as a feature. The target variable is right-skewed, so logarithmic transformation has been performed as shown in Figure 5. Few of the predictor variables like units sold are also right-skewed, so they are log-transformed. The models are trained using 10-fold cross-validation in order to avoid overfitting. Although time series forecasting methods use MAPE as an evaluation metric, this problem doesn't fall under the category of traditional time series where the data is collected at regular intervals. So, the models are evaluated using Mean-Absolute-Error (MAE), R^2 and Root Mean Squared Error (RMSE).

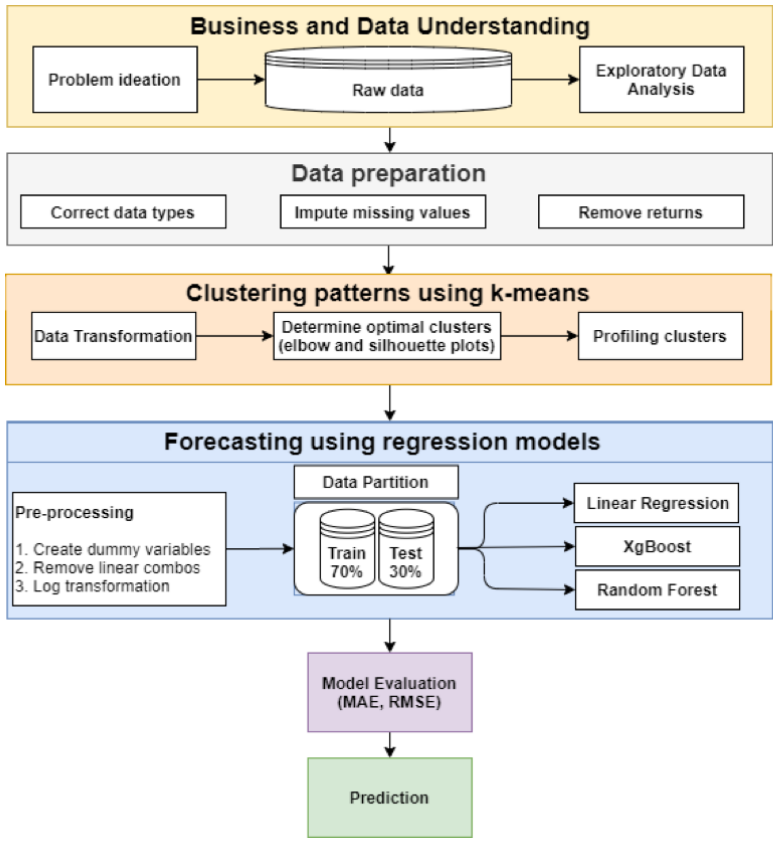


Figure 4: Analytics Workflow

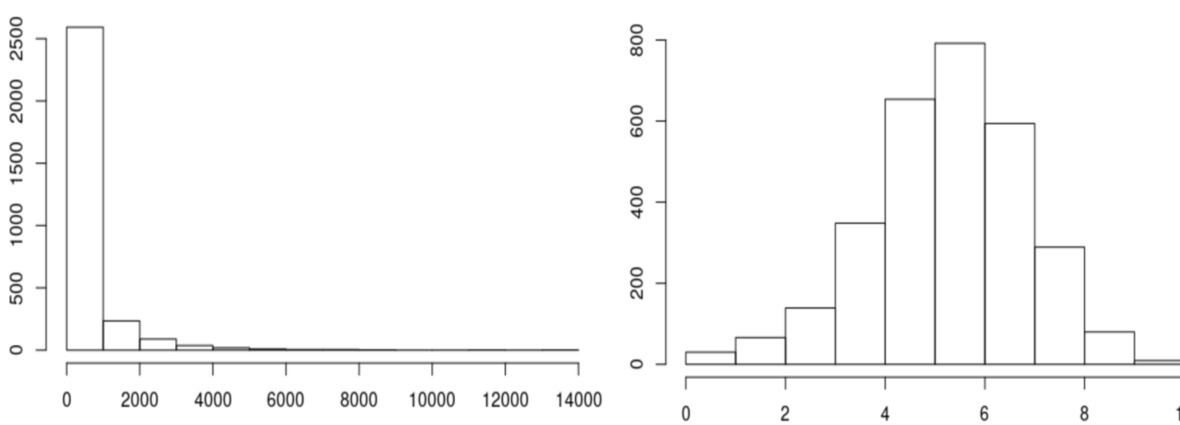


Figure 5: Logarithmic transformation of the target variable (Units sold in 3 months)

5. MODEL FRAMEWORKS

For the study, statistical techniques like Linear Regression, Random Forest, Decision Tree and XGBoost are used to train the model and perform hyperparameter tuning. The summary of all the models used is described below.

5.1 Linear Regression

The short-term sales are estimated using multiple linear regression. The linear regression coefficients are estimated by minimising the residual sum of squares. The below equation is used to estimate the mean of the dependent variable:

$$\bar{y} = \beta_0 + \sum_{i=0}^{n} \beta_i * x_i$$

where $\bar{y}$ is the expected value of the independent variable, which in our case is the sales of new product in the first three months, β_0 is the intercept, β_i is the coefficient of each of the feature x_i as described in sections 2 and 3.

5.2 Decision Trees (Bagging)

A Decision Tree is a non-parametric machine learning modelling technique which can be used for both classification and regression problems. It is based on the recursive binary splitting for partitioning the predictor feature space. The best predictors and best split are aimed at minimizing the RSS. The process is continued until a stopping criterion is achieved. Each tree is built on the bootstrapped sample and the predictions are averaged out to reduce the variance. We implemented the algorithm using "tree bag" function in the CARET package in R. The performance has been validated using the OOB (out-of-bag) samples. Tuning parameters such as the number of trees (nbag), complexity parameter (used to control the size of the trees) have been used to improve the model fit. The bagging prediction can be summarised as follows:

$$\hat{y}^{boot} = \frac{1}{B} \sum_{b=1}^{B} \hat{y}^{b}$$

where $\hat{y}^{b}$is the prediction of decision tree applied to bth bootstrap sample and, $\hat{y}^{boot}$ is the final prediction of the decision tree algorithm.

5.3 Random Forests

Random forests are an improvement over decision trees as they reduce correlation amongst the trees, thereby reducing the overall variance. The trees are grown in parallel, and a random sample of m predictors is chosen from p predictors at each split (m ~ p/3 for regression trees). Any large value for the number of trees to be considered can be chosen until the error has settled down. This has been built using "Ranger" function in CARET package in R. Tuning parameters such as the number of predictors, the number of trees and min node size are used for improving the model fit.

5.4 XGBoost

Unlike random forests, boosting trees are grown sequentially. Each tree learns from the error of the previous tree grown and tries to minimise error. Tuning parameters like learning rate, number of trees, maximum depth, minimum child weight, sub-sample are used to reduce the error.

6. RESULTS AND DISCUSSIONS

In this section, the results are summarised for the experiments conducted during the study.

6.1 Clustering Results

The patterns are clustered into three groups. The clusters are validated using the elbow plot and average width in the Silhouette plots as shown in Figure 6. The clusters are visualised using PCA as shown in Figure 7. The axis PC1 and PC2 explained 84% of the variance of the variables considered in clustering. The profile of each cluster is briefly summarised in Table 2.

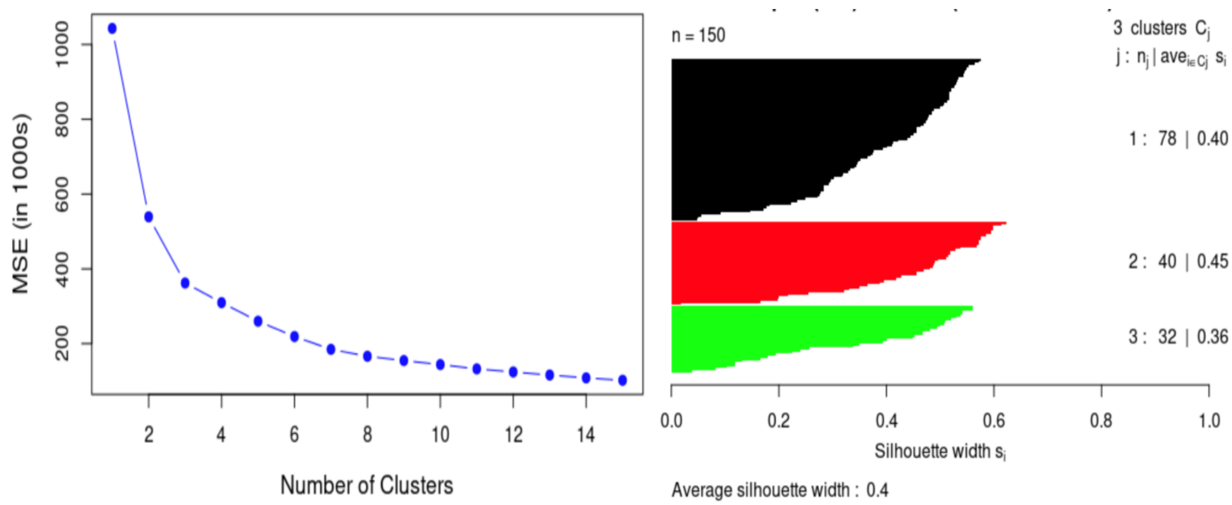


Figure 6: Elbow plot and Silhouette plot

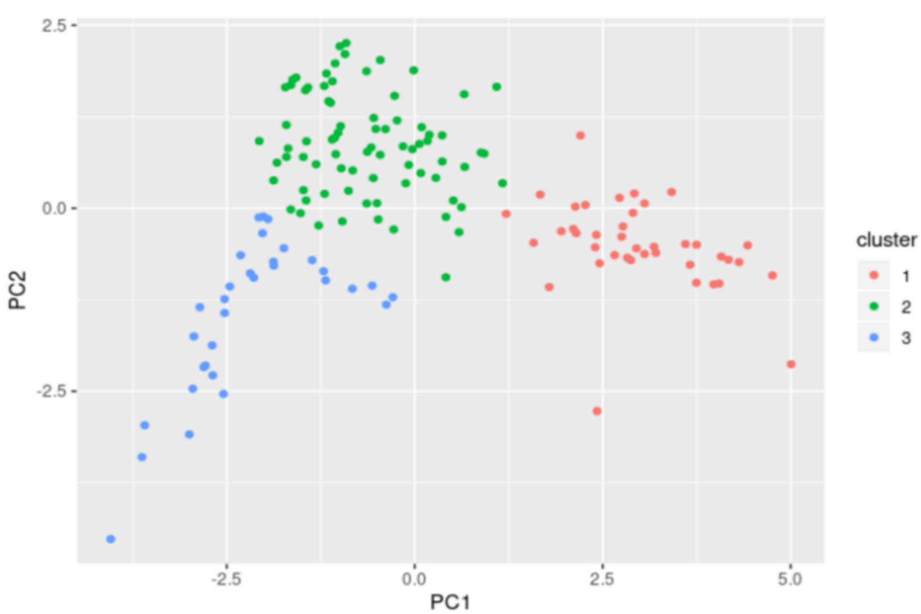


Figure 7: Cluster Profiles using PCA

Table 2: Cluster profiles

Cluster	Description
Cluster 1	Launched across many merchant classes, Highest Revenue and highest units sold
Cluster 2	Launched only in few merchant classes, Higher Revenue and higher units sold
Cluster 3	Less Revenue and lesser units sold, High Priced items

6.2 Forecasting Results

The forecasting results are compared using R^2, MAE and RMSE metrics for Linear Regression, Bagging Trees, Random Forest and XGBoost, as shown in Figure 8. Linear Regression has performed better across train and test sets. Overall, tree-based models have performed better compared to linear regression, but they are overfitting on the training data set.

The actual vs prediction graph for the three clusters on the test dataset is shown in Figure 9. The MAE for Cluster-2 and Cluster-3 is lower than the overall test MAE. However, Cluster-1 has shown significantly higher MAE, as can be seen in Figure 10.

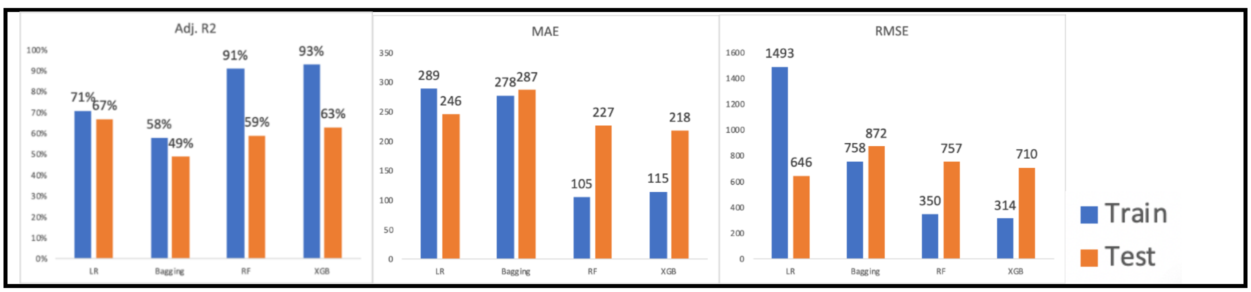
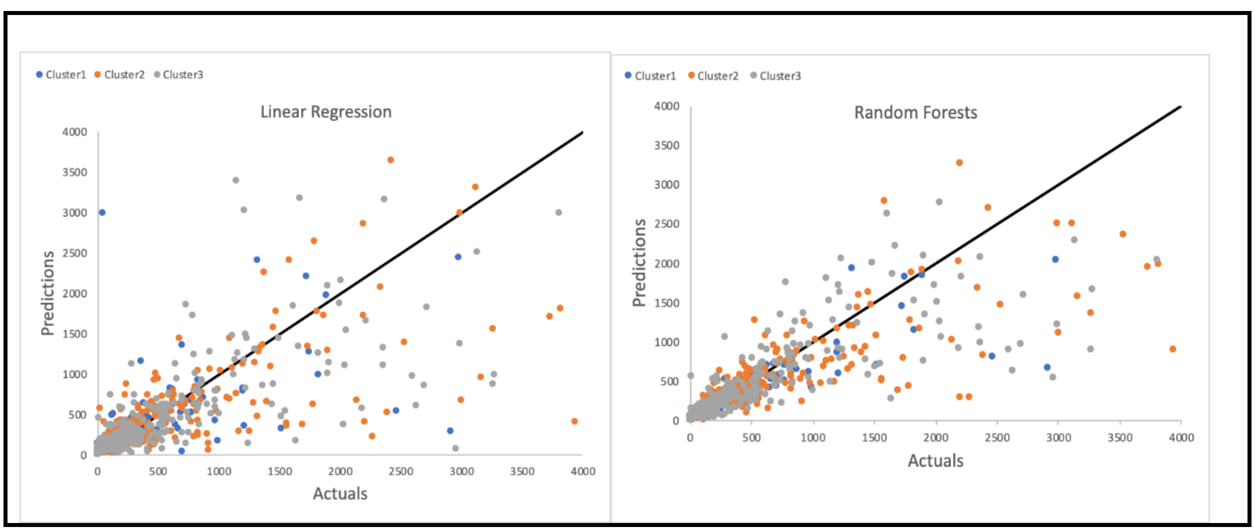


Figure 8: Comparison of R2, MAE, RMSE



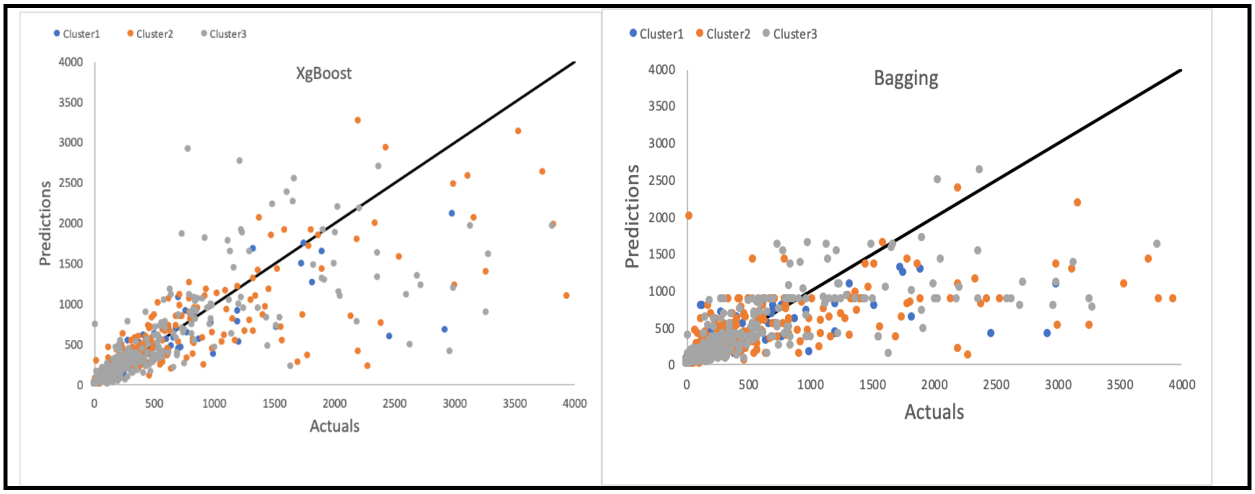


Figure 9: Actual vs Prediction for different models

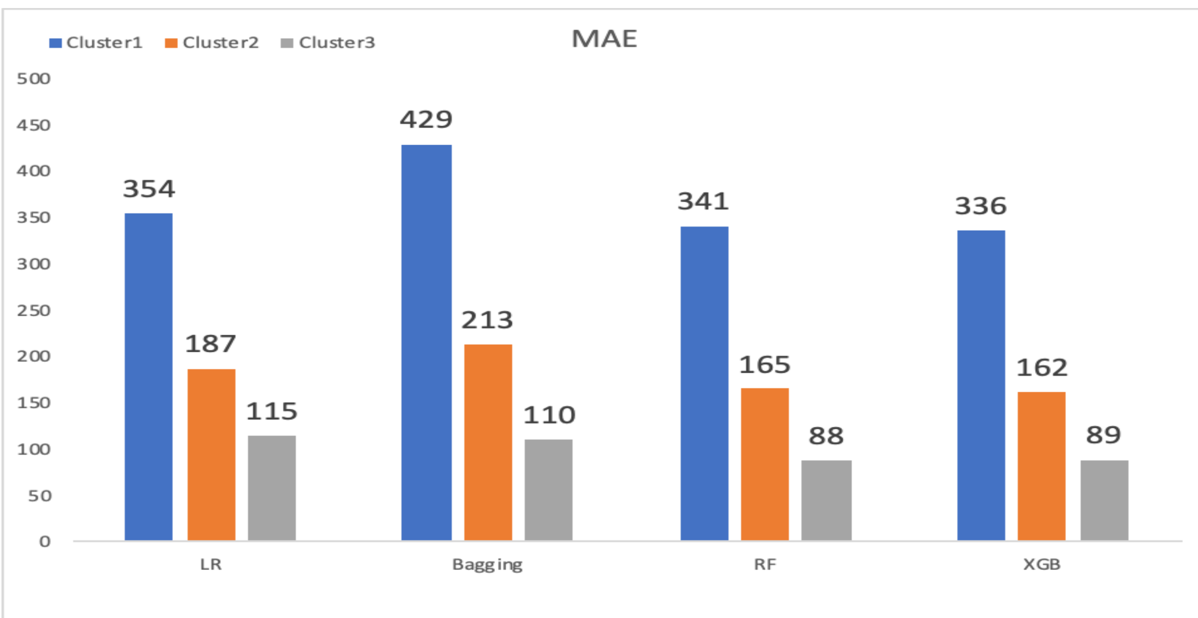


Figure 10: MAE for different clusters

6.3 Effect of Sentiment scores

The model is built using the aggregated sentiment score and volume of comments collected for the pattern in the first three weeks. A sub-sample analysis is conducted for the patterns where sentiment scores have been available. While sentiment scores and volume of comments are found to be significant, the exact impact is not quantifiable as the sample size is very small. The model can be further improved using a larger sample size.

7. CONCLUSIONS AND FUTURE SCOPE

Due to fast-changing fashion trends and shorter lifecycle of the items, forecasting the third-month sales of new products using the initial three weeks of data alone do not yield satisfactory results. However, combining the clustered historical data of previous launches along with the initial three weeks of data of new launches to predict the sales in the first three months have yielded promising results.

Linear Regression has yielded great outcomes with the lowest MAEs as compared to other sophisticated models like Decision Trees, Random Forest and XGBoost. One probable reason could be that the advanced model works best when the size of data is huge, which is a limitation in our case. Moreover, Linear Regression is interpretable and intuitive, which makes it a preferred model if MAE of the other models is similar.

In line with our initial expectations, cannibalization affects the sale of new launches during the initial weeks. Specifically, the patterns launched two months prior to the current launch cannibalize the sales of new launches more as compared to the ones that were launched more than two months prior to the current pattern.

In the future, advanced techniques like neural networks can be employed if we have a larger dataset. We can also try other clustering techniques like hierarchical clustering and partition around medoids (PAM) using existing features and image specific features to get better results and differentiation in the clusters. The current analysis can be extended to optimise the prices of new products across various merchant classes and tailor promotional strategies accordingly to increase their bottom line.

REFERENCES

Özlem İpek Kalaoglu, E. S. (2015). Retail Demand Forecasting In Clothing Industry. *Konfeksiyon Endüstrisinde Perakende Talep Tahminlemesi*.

A.L.D. Louriero, V. M. (2018). Exploring the use of neural networks for sales forecasting in fashion retail. *Elsevier*.

Asli Aksoy, N. O. (2012). A decision support system for demand forecasting in the clothing industry. *International Journal of Clothing Science and Technology, Vol. 24*, 221-236.

Ilan Alon, M. Q. (2001). Forecasting aggregate retail sales: a comparison of arti"cial neural networks and traditional methods.

Kejia Hu, J. A. (2017). Forecasting New Product Life Cycle Curves: Practical Approach and Empirical Analysis.

Pawan Kumar Singh, N. J. (2019). Fashion Retail: Forecasting Deman for New Items. *KDD 2019 Workshop.* Alaska.

Sebastien Thomassey, A. F. (2005). A hybrid sales forecasting system based on clustering and decision trees.

Tsan-Ming Choi, P. C.-L.-F. (2012). Color Trend Forecasting of Fashionable Products with Very Few Historical Data. *IEEE Transactions on Systems Man and Cybernetics Part C (Applications and Reviews)*, 1003-1010.

Yanrong Ni, F. F. (2011). A two-stage dynamic sales forecasting model for the fashion retail. *Expert Systems with Applications*, 1529-1536.

Yen-Sen Shih, M.-H. L. (2019). A LSTM Approach for Sales Forecasting of Goods with Short-Term Demands in E-Commerce. *Intelligent Information and Database Systems*, 244-256.

Zhi-Ping Fan, Y. J.-Y. (2017). Product sales forecasting using online reviews and historical sales data: A method of combining the Bass model and sentiment analysis. *Journal of Business Research*.