

Forecasting Seasonal Footwear Demand Using Machine Learning

By: Majd Kharfan and Vicky Wing Kei Chan

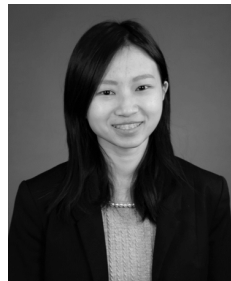
Advisor: Tugba Efendigil

Topic Areas: Forecasting, Demand Planning, Machine Learning

Summary: This project leverages machine learning to recommend two demand forecasting frameworks that help fashion retailers in forecasting new seasonal products' demand. The point-of-sale (POS) data of a U.S.-based footwear retailer was analyzed to identify significant predictor variables and to build two prediction models. Look-alike products were also identified using clustering and classification to be used in prediction. The machine learning-based forecasting models offer better forecast accuracy versus the current performance. The proposed methodology provides visibility into the underlying factors influencing demand and forecast accuracy. Finally, the project results demonstrate the value of forecast customization based on product characteristics.



Prior to MIT, Majd worked as a Strategic Sourcing Analyst at Nike with a focus on Converse's indirect procurement. Majd holds a Bachelor of Economics from Damascus University and will be rejoining Nike/Converse upon graduation.



Prior to MIT, Vicky worked as a Process Analyst at Canadian Tire with a focus on offshore vendor management. Vicky holds a Bachelor of Business Administration from the Hong Kong Polytechnic University.

KEY INSIGHTS

1. The application of machine learning offers visibility into the underlying factors that impact demand with an illustration of their significance.
2. Machine learning allows using, processing and delivering value out of categorical variables, which represent a significant chunk of the fashion industry's data.
3. Customizing the demand forecasting process based on product characteristics deliver better results in comparison to a general forecasting approach.

Introduction

Like other fashion retailers, the sponsoring company is at the crossroads of two key macro shifts: the "Buy Now/Wear Now" consumer mentality influenced by social media and the love of personalization, and the economic challenges facing the retail industry in the form of declining mall traffic and the obsolescence of the

traditional retail calendar. With that in mind, the company is reworking its strategy to improve its position in the marketplace by becoming closer to consumers and quicker in responding accurately to demand signals. This will consequently bring to the company operational efficiencies in the form of minimized order cancellation rates and healthier levels of inventory in the marketplace, which will be translated into cost savings and additional revenues.

Through carrying out this research project we aim to recommend solutions to the sponsoring company that will improve the demand forecasting capabilities and prediction accuracy. Applying machine learning will maximize the utilization of the point-of-sale (POS) data and help uncover new insights to be used in developing a demand forecasting framework that meets the company's strategic objectives.

Methodology

The proposed methodology can be broken down into the following main buckets: data pre-processing, models building, and performance measurement.

Data Pre-Processing: Two types of data were collected from the sponsoring company: sell-in (shipment) and sell-through (POS) data. The POS data collected covered 115 U.S. retail outlet stores and approximately four and half years of records. The objective was to find out how the data can be leveraged to improve the demand forecasting capability, especially for seasonal products without sales history.

Data Aggregation and Filtering: As the focus is to support the decision of how much to order from each style per season, the data were aggregated across all stores to the monthly level (the buy decision's level). Additionally, clearance (discounted) products and products with a lifecycle beyond 4 months were filtered out as the company is only interested in forecasting brand-new products with 2-4 months of lifecycle.

Feature Engineering: Some variables were modified or added as necessary in preparation for building the model. For example, colors were aggregated into groups based on similarities. Also, store count was added as a candidate variable and as a reference to the number of stores selling a product. As seasonal styles are launched at different times of the year with short lifecycles, their sales are believed to be dependent on lifecycle attributes in addition to the calendar attributes and therefore three lifecycle variables were created.

Feature Selection: Recursive feature elimination, a backward feature selection method, was used to eliminate features based on their contribution to improving forecast accuracy. A random forests algorithm was used on each iteration to evaluate the model with different subsets of variables.

The general model: Utilizing twelve different attributes selected from the feature selection process, Table 1 lists these attributes ranked by their significance, a general forecasting model was built.

Table 1. List of Attributes Selected for Model Building

Importance Rank	Attribute	Attribute Description
1	Store Count	Number of stores
2	Month	Fiscal month
3	Lifecycle Month	Number of months since product launch
4	Gender	Gender or age group
5	Average Unit Retail (AUR)	Actual selling price
6	Year	Fiscal year
7	Basic Material	Type of material
8	Manufacturer's Suggested Retail Price	Ticket price
9	Color Group	Color code
10	Lifecycle	Number of months in the lifecycle of a style
11	Cut Description	Ankle height
12	Product Class	Product main feature

The algorithms that were explored in this model included regression trees, random forests, k -nearest neighbor (k -NN) and neural networks. In addition, ensemble methods taking the median and average of the outputs from the four individual methods were also considered.

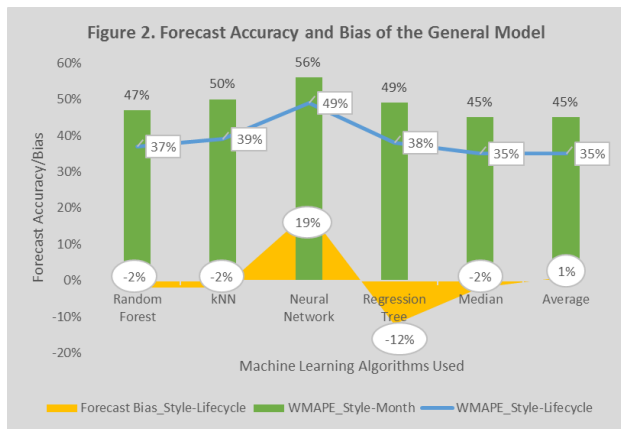
The three-step model: The three-step model can be distinguished from the general model that it consists of three separate stages: (i) clustering, (ii) classification, and (iii) prediction. The main objective behind this model is to identify look-alike group of products from the historical data. Once these products are identified, their average sales can be used as a proxy to forecast the sales for brand-new products.

Performance Measurement: Two performance metrics were used in the forecasting models: forecast accuracy and bias. Forecast accuracy was measured using Weighted Mean Absolute

Percentage Error (WMAPE) and bias was measured using Weighted Mean Percentage Error (WMPE).

Results

General Model: The results in terms of forecast accuracy of the four individual models using regression trees, random forests, *k*-NN and neural networks and the two ensemble models using the median and average of individual outputs are shown in Figure 2.



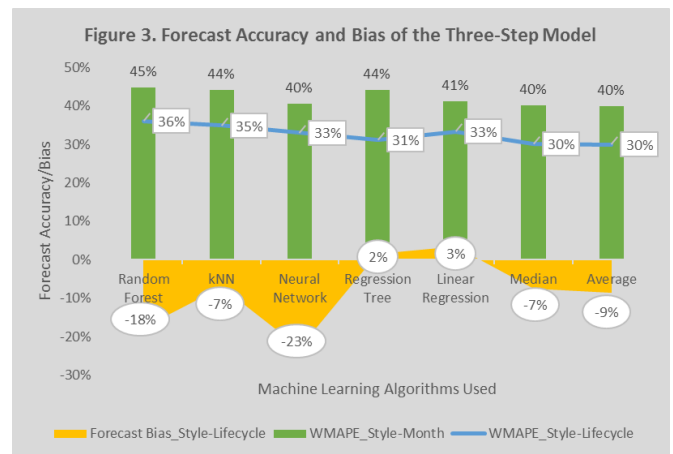
Considering the individual models, random forests gives the best predictive performance with the highest accuracy and lowest bias. It achieved 37% WMAPE on the style-lifecycle level and 47% on the style-month level with a negative bias of 2%. *k*-NN also gives reasonably good results in terms of forecast accuracy.

As for the ensemble methods, taking the median and average of the individual model outputs yields a better forecast accuracy with a WMAPE of 35%. Neural networks show the worst performance in terms of both accuracy and bias, with a 49% WMAPE and a positive bias of 19%, indicating a tendency to over-forecast.

Three-Step Model: The clustering and classification analyses shows that lifecycle length is a clear distinguishing driver in clustering. The number of clusters that seemed to represent the data and which gave the best classification results was five. Each of the five clusters included one lifecycle length, except one

cluster, which included styles with mixed lifecycles. However, the styles that were included in this cluster seemed to have relatively smaller sales volume, smaller store count and higher average unit retail price compared to the other clusters.

When it comes to prediction, regression trees delivered the best performance with a forecast accuracy of 31% (style-lifecycle level) and 2% of forecast bias (over-forecast). Linear regression showed a very close performance with 31% of forecast accuracy (style-lifecycle level) and 3% of forecast bias (over-forecast). Although the neural networks delivered a forecast accuracy of 33% (style-lifecycle level), it's 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. Overall, the ensemble methods delivered the highest forecast accuracy with 30% WMAPE on a style-lifecycle level. However, their forecast bias was relatively high (around 8% under-forecast) compared to the regression trees and linear regression algorithms. See Figure 3.



Clicking down into cluster level, the machine learning algorithms performed differently from one cluster to another. *k*-NN and linear regression were best performers in the clusters with a mono lifecycle and high average sales volume. Their WMAPE was 28% while

WMPE ranged from -11% to +4%. Random forests was the best performer in the clusters with mono lifecycle and medium average sales volume with a WMAPE ranges from 32% to 37% and WMPE ranges from -11% to +6%. Finally, regression trees was best performer in the cluster with multiple lifecycles and low average sales volume. WMAPE ranged from 39% to 45% and WMPE ranged from -30% to 0%.

Finally, the project opens doors for further research that possibly cover store inventory allocation, size curve analysis and price optimization.

Conclusions

Our research project proposed a methodology that offers two different forecasting models based on machine learning techniques. These models will enable the sponsoring company to achieve better forecast accuracy compared to the current performance by considering store count, lifecycle, calendar and product attributes simultaneously.

The data pre-processing phase of the proposed methodology is an important stage that facilitates the formation of the inputs to the models. The feature engineering process helps create new variables that bring additional value to demand interpretation. The feature selection process allows us to gain insights into the importance of the different predictor variables 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 variables that have always been considered a challenge when it comes to forecasting demand in the fashion industry.

When it comes to the models, the general model serves as a starting point for easy implementation of the machine learning forecasting framework. The three-step model involving clustering, classification and prediction enables the company further to visualize the relationship between predictor variables and customize the forecasting approaches accordingly.