A Machine Learning Approach to Forecasting Demand in Fashion Industry

Michal Koren*
School of Industrial Engineering and Management
Shenkar—Engineering. Design. Art.
Ramat Gan, Israel
michal.koren@shenkar.ac.il
*Corresponding author

Noam Abulafia
Faculty of Engineering
Bar-Ilan University
Ramat Gan, Israel
noamabulafia@gmail.com

Lior Shviro
Faculty of Engineering
Bar-Ilan University
Ramat Gan, Israel
liorshviro@gmail.com

Noam Weiss
Faculty of Engineering
Bar-Ilan University
Ramat Gan, Israel
noweiss97@gmail.com

***Abstract*— Fashion apparel retail is lucrative, dynamic, and unpredictable, as it is influenced by a wide range of factors, including culture, economics, social media, and fashion trends. Due to these influences, coupled with the industry's inherent need to innovate, track customer preferences and global trends, and introduce new products, demand and sales forecasting are particularly challenging. Since the fashion apparel industry has a high degree of uncertainty, the methods available for demand forecasting in other industries are more difficult to apply. Data analysis and forecasting are becoming increasingly automated through the use of ML (Machine Learning) tools and algorithms. Our research suggests such tools and algorithms might be applied to fashion. This proposed model presents a forecasting method that can be used to develop a collection and manage production and supply chains by fashion retailers, based on their existing products as well as the model's assessment.**

Keywords- Fashion; Forecasing; Machine Learning; K-Nearest Neighbor; Classification;

I. INTRODUCTION

The fashion industry is a lucrative sector with a value of several trillion dollars worldwide. Moreover, the fashion industry generates thousands of tons of waste each year, making it one of the most polluting industries in the world [1, 2]. Financial losses can result from unsold clothes being thrown away after the season. Thus, developing models that reduce financial losses by making production quantities more accurate is a strong motivation, particularly in light of the fact that environmentalism is becoming more popular among major companies. In the fashion industry, accurate forecasting can reduce waste, save money, and give you an edge over your competitors [3]. Therefore, accurate forecasting is therefore crucial. Due to fashion's complexity and rapid changes, traditional models often fail to accurately forecast demand for inventory and supply chain operations [4, 5].

Color is one of the most powerful and emotive elements in fashion. Designers and brands use color as a primary tool for storytelling in collections, and consumers often make purchasing decisions based on their emotional reaction to a particular hue [6]. Consequently, understanding and predicting future color trends can give fashion brands a competitive edge. Fashion forecasting, particularly in terms of color, relies on a blend of creative intuition, data analysis, and societal observation [7]. Trend forecasting agencies like Pantone, WGSN, and Fashion Snoops dedicate entire divisions to color analysis, predicting what hues will dominate upcoming seasons [8, 9, 10].

As well as shaping the creative process, forecasting also impacts production, marketing, and retail strategies [11]. When a color forecast is established, it impacts fabric manufacturing and marketing campaigns. Retailers prepare their inventory to meet anticipated consumer demand, while suppliers source dyes and materials that match the forecasted shades [12]. Color decisions are often made 12 to 18 months in advance by brands, which is why accurate forecasting is critical. By correctly predicting color trends, brands can increase sales, improve brand visibility, and increase consumer loyalty. Conversely, if they miss a color trend, they can result in unsold inventory or missed market opportunities [13, 14].

Color forecasting is still not an exact science, despite advances in technology and data analytics [15, 16]. Regardless of how well-planned a forecast is, it can be disrupted by unforeseen events (such as sudden global shifts or changes in consumer mood). Additionally, the rapid turnover of trends in today's fast-fashion environment poses another challenge, as brands must balance long-term forecasting with the need to remain agile and responsive to immediate changes. In the age of online shopping, social media, and even search engines, data-driven forecasting is increasingly important [17]. ML algorithms can analyze vast amounts of consumer behavior data and track colors that are trending in real-time. Brands use this information to optimize their color palettes and design collections that align with consumer trends. ML-driven forecasting can also help companies anticipate customer needs and design products accordingly. As a result, they are able to stay ahead of the competition and remain relevant [18].

This paper proposes a forecasting model based on machine learning for forecasting in the fashion industry. Model contributions are as follows: (i) Short-term: supporting collection planning as much as possible, (ii) Long-term: training and verifying machine learning models with more data. Using this model based on machine learning to find and identify data patterns is a powerful way to support decision-making based on data itself.

II. Model

We propose a model that leverages data provided by forecasting agencies, focusing on color trends, with the aim of identifying patterns that can be extended to future periods not currently forecasted. Forecasting agencies provide forecasts for colors relevant to the forecasted season by dividing them into shades. Each color is defined by a single value numerical code. The model is based on this data published by forecasting agencies.

Specifically, two stages are involved in the implementation: (1) ML Stage: A machine learning algorithm uses the agencies' data to forecast market color segmentation for future seasons and years and provides relative weights for various colors. Since forecasting agencies offer a lot of information, at this stage only the company-relevant data is processed and adjusted for its consumer audience. In this stage, a forecast vector is defined by using k-means algorithm to classify the forecasts, then a Naive Bayesian (NB) network is used to define the probability vectors for each color so that the final output of this stage is a forecast vector that includes weights for each color in the collection composition. In the next stage, this vector will be used to maximize profits. (2) Linear Programming (LP) Stage: An optimization model refines the general forecast, aligning it with the specific requirements and constraints of each individual company to maximize profitability. Consequently, the output details the optimal distribution of colors for production tailored to each product. Fig. 1 presents the stages of the model.

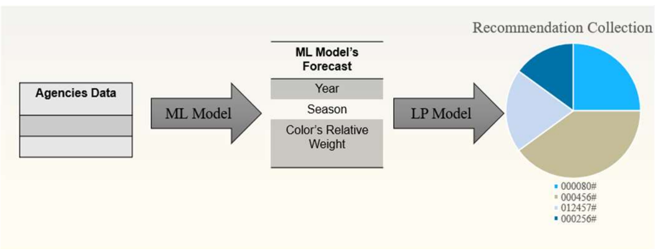

Fig. 1. Model's stages.

1) ML Stage: The focus of this stage is to develop a forecast vector for the specified season, based on forecast vectors as defined within the company. this stage includes three phases.

a) Clustering Phase – Visualization and labeling

Historical forecasts records from various forecasting agencies are used as input in the following structure: $< Year, Season, (W_1, W_2, W_3, \ldots W_n) >$

When W_i is the Weight of color i in a record, $i \in [1, \ldots, n]$.

In this stage, forecasts are plotted within a color space (corresponding to the number of colors defined by the manufacturer). Each axis represents the relative proportion of each color in the palette used for the prediction (the palette being the model's input). The predictions are then grouped into clusters using the k-means algorithm, which is applied based on the spatial proximity of the points. This clustering helps identify which forecasts are similar in terms of color composition and relative weight.

The k-means algorithm will be executed for varying values of K, and the optimal K will be determined using the Elbow Method criterion. The clusters will be mapped and labeled arbitrarily, with the centroid of each cluster being retained.

The collection of clusters will be denoted as $C = (C_1, C_2, C_3, \ldots, C_k)$.

In the output, forecasting agencies' records are grouped into clusters and their centroids are recorded (fig. 2)

b) Bayesian Network Phase

The Bayesian network model is applicable given that each season occurs annually, and all seasons are represented each year (assuming the manufacturer has adequately processed the data). This model utilizes the seasonal repetition pattern to estimate the probability of belonging to each cluster based on historical data from previous years. Specifically, for each season we aim to predict, we will use the historical data of previous seasons to inform our predictions.

The model will denote the target season as i and the cluster assignment for record j as $T(j)$.

The output will be a probability vector indicating the likelihood of a record belonging to each cluster. This probability vector will be used to categorize a new record within the appropriate cluster. In Fig. 3, the requested input is shown in red, while the probability vector output is shown in orange.

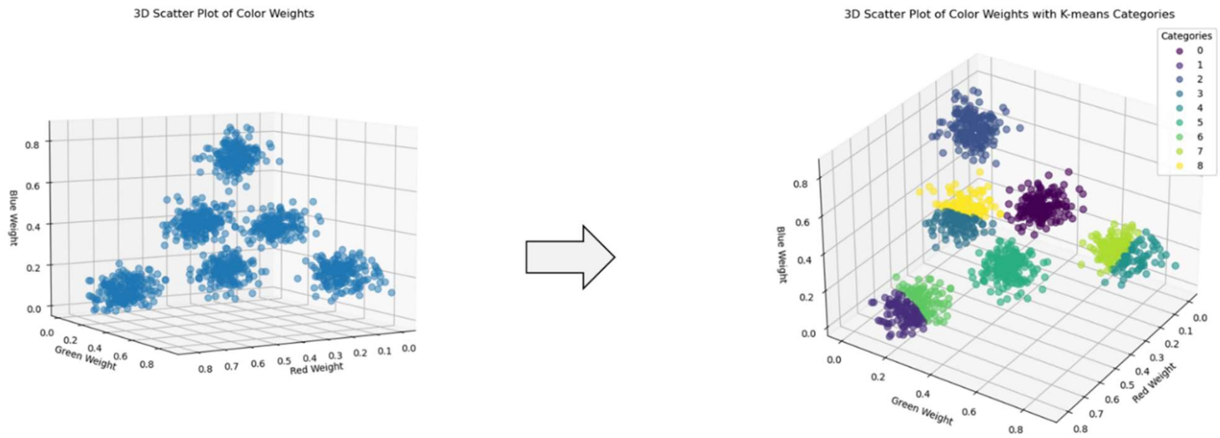

Fig. 2. Clustering Phase.

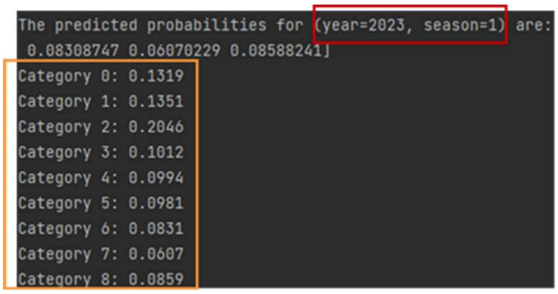

Fig. 3. Beyesian network phase

c) <u>Forecasting Phase</u>

A probability vector indicating the likelihood of belonging to each cluster is currently used as input.

Using the probability vector, the K-Nearest Neighbors (K-NN) Weighted Average method will be used to generate the forecast. In each cluster, the K closest records over the predicted period are identified. The average values of these records are multiplied by the input vector's probabilities, and the resulting weighted averages are then summed. Forecasts for the upcoming season are based on weighted contributions of different colors based on the selected forecasting method (see fig. 4).

2) **LP Stage**: It is at this stage that the problem is formulated as an optimization problem in order to achieve the best results. The objective is to maximize profits for a specific product by leveraging a general color forecast and detailed product data (including production and sales costs) while adhering to the organization's operational constraints in order to maximize profits.

The notations of the models are as follows:

$$X_i = weight\ of\ color\ i, i \in [1, n]$$
$$C_i = profit\ of\ item\ with\ color\ i$$
$$f_i = prediction\ weight\ of\ color\ i$$
$$\alpha = slack\ constant, \alpha \in (0,1)$$

An arbitrarily defined parameter, α, allows for deviations in the color weights from the forecasted values. This parameter introduces flexibility to accommodate minor variations and ensure practical applicability.

The objective function of the company's profit is formulated based on the color forecast and product-specific data:

$$Max\ Z = \sum_{i=1}^{n} C_i \cdot X_i$$

and the constraints are:

$$\sum_{i=1}^{n} X_i = 1$$
$$X_i \leq f_i + \alpha$$
$$X_i \geq 0, \forall i$$

Additionally, a company can add constraints based on its operational constraints, such as storage limits, supply chain costs, logistical considerations etc.

The optimization process determines which colors to use for a given product during a given season to maximize profits for the company whilst maintaining all constraints in place for the product. This X-vector identifies the optimal color weights for a given product during a given season.

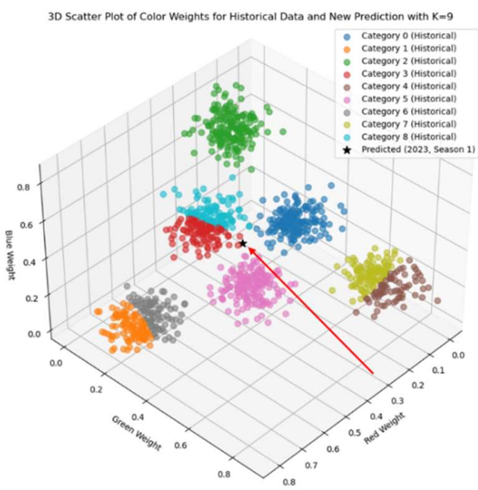

Fig. 4. Forecasting phase.

III. Discussion and Conclusion

It is presented in this study that a linear programming and machine learning model can be employed to improve demand forecasting in the fashion apparel industry. As a result of the dynamic nature of the industry and the difficulties associated with predicting consumer preferences, the integration of data from forecasting agencies into a two-stage advanced process offers a promising solution in this regard. It follows from the ML stage that the forecast is refined in order to meet the company's specific constraints and profitability goals based on the results of the LP stage.

As a result, fashion retailers are able to better respond to market changes by aligning color trends with production planning. Using the proposed model, inventory management, collection development, and supply chain optimization can be accomplished with greater precision. Therefore, the model illustrates how data-driven strategies, based on machine learning and optimization techniques, can be effectively employed to enhance decision-making in the fashion apparel industry, allowing businesses to navigate uncertainty while maximizing profits in order to meet customer demands effectively.

References

[1] Niinimäki, K., Peters, G., Dahlbo, H., Perry, P., Rissanen, T., & Gwilt, A. (2020). The environmental price of fast fashion. *Nature Reviews Earth & Environment*, *1*(4), 189-200.

[2] Gupta, R., Kushwaha, A., Dave, D., & Mahanta, N. R. (2022). Waste management in fashion and textile industry: Recent advances and trends, life-cycle assessment, and circular economy. *Emerging trends to approaching zero waste*, 215-242.

[3] Koren, M., & Shnaiderman, M. (2023). Forecasting in the fashion industry: a model for minimising supply-chain costs. *International Journal of Fashion Design, Technology and Education*, *16*(3), 308-318.

[4] Fildes, R., Ma, S., & Kolassa, S. (2022). Retail forecasting: Research and practice. *International Journal of Forecasting*, *38*(4), 1283-1318.

[5] Koren, M., Perlman, Y., & Shnaiderman, M. (2024). Inventory management for stockout-based substitutable products under centralised and competitive settings. *International Journal of Production Research*, *62*(9), 3176-3192.

[6] Ren, S., Chan, H. L., & Siqin, T. (2020). Demand forecasting in retail operations for fashionable products: methods, practices, and real case study. *Annals of Operations Research*, *291*, 761-777.

[7] Li, S., Chen, R., Yang, L., Huang, D., & Huang, S. (2020). Predictive modeling of consumer color preference: Using retail data and merchandise images. *Journal of Forecasting*, *39*(8), 1305-1323.

[8] DuBreuil, M., & Lu, S. (2020). Traditional vs. big-data fashion trend forecasting: an examination using WGSN and EDITED. *International Journal of Fashion Design, Technology and Education*, *13*(1), 68-77.

[9] Koren, M., Perlman, Y., & Shnaiderman, M. (2022). Inventory management model for stockout based substitutable products. *IFAC-PapersOnLine*, *55*(10), 613-618.

[10] Fu, Y., & Fisher, M. (2023). The value of social media data in fashion forecasting. *Manufacturing & Service Operations Management*, *25*(3), 1136-1154.

[11] Lai, P., & Westland, S. (2022). On the relationship between colour search trends, economic indicators and colour forecasting. *Journal of the International Colour Association*, *30*, 24-34.

[12] Yu, Y., Hui, C. L., & Choi, T. M. (2012). An empirical study of intelligent expert systems on forecasting of fashion color trend. *Expert Systems with Applications*, *39*(4), 4383-4389.

[13] Chen, D., Liang, W., Zhou, K., & Liu, F. (2022). Sales Forecasting for Fashion Products Considering Lost Sales. *Applied Sciences*, *12*(14), 7081.

[14] Pedro, Y., Koren, M., Peretz, O., Fisher, N., & Yazdi, O. (2024, July). FASHION'S COLOR PALETTE IN TIMES OF CRISIS. In *Global Fashion Management Conference* (pp. 492-495).

[15] Gladys, A. O., & Olalekan, A. S. (2021). A machine learning model for predicting colour trends in the textile fashion industry in south-west Nigeria. *International Journal on Integrated Education*, *4*(2), 174-188.

[16] Kharfan, M., Chan, V. W. K., & Firdolas Efendigil, T. (2021). A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. *Annals of Operations Research*, *303*(1), 159-174.

[17] Han, A., Kim, J., & Ahn, J. (2022). Color trend analysis using machine learning with fashion collection images. *Clothing and Textiles Research Journal*, *40*(4), 308-324.

[18] Shi, M., Chussid, C., Yang, P., Jia, M., Dyk Lewis, V., & Cao, W. (2021). The exploration of artificial intelligence application in fashion trend forecasting. *Textile Research Journal*, *91*(19-20), 2357-2386.