Abstract

We present Visuelle 2.0, the first dataset useful for facing diverse prediction problems that a fast-fashion company has to manage routinely. Furthermore, we demonstrate how the use of computer vision is substantial in this scenario. Visuelle 2.0 contains data for 6 seasons / 5355 clothing products of Nuna Lie\(^1\), a famous Italian company with hundreds of shops located in different areas within the country. In particular, we focus on a specific prediction problem, namely short-observation new product sale forecasting (SO-fore). SO-fore assumes that the season has started and a set of new products is on the shelves of the different stores. The goal is to forecast the sales for a particular horizon, given a short, available past (few weeks), since no earlier statistics are available. To be successful, SO-fore approaches should capture this short past and exploit other modalities or exogenous data. To these aims, Visuelle 2.0 is equipped with disaggregated data at the item-shop level and multi-modal information for each clothing item, allowing computer vision approaches to come into play. The main message that we deliver is that the use of image data with deep networks boosts performances obtained when using the time series in long-term forecasting scenarios, ameliorating the WAPE and MAE by up to 5.48% and 7% respectively compared to competitive baseline methods. The dataset is available at: https://humaticslab.github.io/forecasting/visuelle.

1. Introduction

Fashion forecasting has traditionally been studied in scientific sectors other than computer vision, such as operational research and logistics, with the primary aim of predicting trends \cite{16,22}, sales \cite{21,24,25}, and performing demand forecasting \cite{13,34}. In recent years this trend has faded, showing an increasing cross-fertilization with computer vision \cite{1,7,10,26,28}.

In this paper, we present Visuelle 2.0, which contains real data for 5355 clothing products of a retail fast-fashion Italian company, Nuna Lie. For the first time ever, a retail fast-fashion company has decided to share part of its data to provide a genuine benchmark for research and innovation purposes. Specifically, Visuelle 2.0 provides data from 6 fashion seasons (partitioned in Autumn-Winter and Spring-Summer) from 2016-2020, right before the Covid-19 pandemic\(^2\). Each product in our dataset is accompanied by an HD image, textual tags and more. The time series data are disaggregated at the shop level, and include the sales, in-

\(^{1}\)http://www.nunalie.it

\(^{2}\)The pandemic represented an unicum in the dynamics of the fast fashion companies, so it has not been included. The market effectively restarted in AW 21-22 which is an ongoing season at the time of writing.
Forecasting new product demand forecasting

With these data, Visuelle 2.0 allows to cope with several problems which characterize the activity of a fast fashion company: new product demand forecasting [13, 34], short-observation new product sales forecasting [21, 24, 25] and product recommendation [6].

In this paper, we focus on one of these problems: short-observation new product forecasting (SO-fore). SO-fore aims at predicting the future sales in the short term, having a past statistic given by the early sales of a given product (Fig. 2b provides a visual comparison with standard forecasting in Fig. 2a). In practice, after a few weeks from the delivery on market, one has to sense how well a clothing item has performed and forecast its behavior in the coming weeks. This is crucial to improve restocking policies [27]: a clothing item with a rapid selling rate should be restocked to avoid stockouts. Two particular cases of the SO-fore problem will be taken into account: SO-fore$_{2-10}$, in which the observed window is 2 weeks long and the forecasting horizon is 10 weeks long, required when a company wants to implement few restocks [14]; SO-fore$_{2-1}$, where the forecasting horizon changes to a single week, and is instead required when a company wants to take decisions on a weekly basis, as in the ultra-fast fashion supply chain [9, 39].

Our findings show that the usage of image data is crucial in absence of long term statistics which characterize the sales, because the pictorial content of a fast fashion product can be used to inherit long term statistics from visual similarity (see Fig. 2b): the images allow to refer with high precision to past data that are akin to the product of analysis, providing informative priors.

Visuelle 2.0 is a substantial extension of the unpublished Visuelle dataset [37], used only for new product demand forecasting, where less data (no weather conditions, no customer data) was furnished, aggregated per-product over all the retail stores (no geographical/store dimension).

2. The Dataset

Visuelle 2.0 describes the sales between Nov. 2016 and Dec. 2019 of 5355 different products across 110 different shops. For each product, multi-modal information is available, as described in the following.

**Time series data.** Given a product $i$ of size $s$ at a retail store $r$, we refer to its product sale signal as $S(i, s, r, t)$ where $t$ refers to the $t$-th week of market delivery, with $i = 1, \ldots, N$, $s = 1, \ldots, M$, $r = 1, \ldots, L$ and $t = 1, \ldots, K$. We also define the inventory position signal $I(i, s, r, t)$ indicating the inventory on-hand for that quadruplet $(i, s, r, t)$. Combining these data, we can individuate all those legit sales signals that do not involve a stockout until the $K$-th week. Formally, a legit sale signal is $S(i, s, r, 1), \ldots, S(i, s, r, t_{\text{legit}})$ where $t_{\text{legit}} + 1$ indicates the first week with a stockout $I(i, s, r, t_{\text{safe}} + 1) = 0$, and $t_{\text{safe}} > K$. Hence, we guarantee that a zero-sale legit signal (i.e. $S(i, s, r, t) = 0$) is provided only when nobody bought $i$ (even tough it was available at the shop) and not because of a stockout. These signals are important because they focus on the net performance of a product, independently on the inventory management. To make the signals denser, we aggregate the different sizes obtaining the final $\text{sale}(i, r, t) =$

---

3Prices have been normalized for the sake of confidentiality.
Multiple textual tags related to each product’s visual attributes are available. These tags have been extracted with diverse procedures or chosen by hand, carefully validated by the Nuna Lie team. The first tag is the category, taken from a vocabulary of 27 elements, visualized in Fig. 4a; the cardinality of the products shows large variability among categories overall, due to the fact that some categories (e.g. long sleeves) cost less and ensure higher earnings. The color tag (Fig. 4b) represents the most dominant color chosen among 10 manually detected colors. The fabric tag comes directly from the technical sheet of the products, chosen from a vocabulary of 59 elements. Finally, the release date for each product-shop pair is recorded as a textual string.

Exogenous time series Google Trends time series for each product are provided, based on the product’s three associated attributes: color, category, fabric. The trends are downloaded starting from 52 weeks before the product’s release date, essentially providing a popularity curve for each of the attributes. Google Trends’ efficacy for predictions on fast fashion problems has been demonstrated before in [35, 37]. Weather reports downloaded from IlMeteo4 are also supplied, containing the real weather conditions on a daily basis at the municipality level. The efficacy of weather reports for forecasting in fast fashion has been demonstrated in [3, 38].

---

4https://www.ilmeteo.it/portale/archivio-meteo/
3. Experiments

Here we show how Visuelle 2.0 is a genuine benchmark for two types of SO-fore: i) SO-fore_{2-10} and ii) SO-fore_{2-1}.

SO-fore_{2-10} allows the company to customize the re-stocking operations for each product on the basis of the early sales, minimizing the number of such operations. Firstly, the sales series are split into an observation window and a horizon window (i.e. the known past and the future to forecast), set here to 2 and 10 weeks respectively, covering the 12-week fast fashion life-cycle [41]. Formally, the goal is to perform a multi-step forecast of the sales of an item \(i\) in a store \(r\) \((\text{sale}(i, r, 3), ..., \text{sale}(i, r, 12))\), given the first two time-steps. Two weeks is a standard period to sufficiently understand the whereabouts of the fashion market and take decisions for the future [40, 45].

SO-fore_{2-1} serves in other contexts where a weekly re-stocking schedule is adopted [9, 39]. The idea is to estimate the time-point \(\text{sale}(i, r, t)\) given the previous two time-steps \(\text{sale}(i, r, t-1)\) and \(\text{sale}(i, r, t-2)\). Similarly to before, we set the initial observation window to 2, but make a unique prediction corresponding to each observed window.

In both cases we split the data into train and test sets, where the test set contains the 10\% most recent item-shop pairs, such that the items that are seen in training will have always been released before the ones we test on. We use the following three approaches as baselines:

- Classical forecasting methods, namely the Naive method [18] (using the last observed ground truth value as the forecast) and Simple Exponential Smoothing (SES) [18, 19];
- kNN [11, 29], which produces forecasts based on product similarity. This is done by finding k-Nearest Neighbors from the past that are similar to the input product. We set \(k = 11\) and we compute the similarity between products using the known time series or image features extracted by a pre-trained ResNet [15];
- An autoregressive, attention-based RNN architecture [11], where the different data modalities are first processed separately and then merged together through several attention modules [2].

Results are displayed in Table 1 and Table 2, where all the listed approaches only use sales time series data as input, unless specified otherwise. Classical forecasting approaches tend to give poor performances due to the small number of observations [18]. The kNN-based methods show an improvement over the statistical forecasting baselines, demonstrating that inter-product similarity is important when predicting future sales. This is tied with the notion that new products will sell comparably to older, similar products. Trivially utilizing the images with kNN lowers the performances, but their contribution is highlighted when

| Method               | WAPE | MAE |
|----------------------|------|-----|
| Naive                | 118.176 | 1.31 |
| SES                  | 111.265 | 1.23 |
| kNN                  | 91.13 | 1.01 |
| kNN + image          | 97.97 | 1.06 |
| CrossAttnRNN         | 87.79 | 0.97 |
| CrossAttnRNN w/ image| 85.65 | 0.94 |

Table 1. Results for SO-fore_{2-10}, showing the Weighted Average Percentage Error (WAPE) and Mean Absolute Error (MAE) [37] for the different baselines. The lower the better for both metrics. In the first row we also report results for the demand forecasting of new products without past sales, demonstrating how much the knowledge of the initial sale dynamics improves forecasting.

| Method               | WAPE | MAE |
|----------------------|------|-----|
| Naive                | 101.922 | 1.13 |
| SES                  | 97.85 | 1.08 |
| kNN                  | 91.11 | 1.01 |
| kNN + image          | 92.97 | 1.02 |
| CrossAttnRNN         | 87.86 | 0.97 |
| CrossAttnRNN w/ images| 86.71 | 0.96 |

Table 2. Results for SO-fore_{2-1}, showing the Weighted Average Percentage Error (WAPE) and Mean Absolute Error (MAE) [37] for the different baselines. The lower the better for both metrics.

utilising a neural network architecture. Cross-Attention RNN [11] outperforms the others by a noticeable margin, because the model is able to learn non-linear, inter-product dependencies throughout the whole training set and also advanced temporal dynamics. It is worth noting that we reach the best performances in SO-fore_{2-10} by pairing each product’s time series input with its respective image. This shows that visual representations allow the model to better understand long term forecasting patterns. Another important takeaway is that the results for time series only methods are much better on SO-fore_{2-1}, due to the localized temporal information and shorter forecasting horizon, while the images become less important than in the previous case. Additionally, we tested Cross-Attention RNN in the Demand forecasting task, i.e., predicting the full sales series without having access to any previous observations, but only to the product image. We can observe that even in this more challenging setting, results are inferior to the SO-fore variants but better than the kNN approaches.

Our dataset and experiments provide a general overview of how problems in the fashion realm can be tackled and how the use of computer vision and multi-modal approaches is key to providing better solutions. The dataset page will contain further information regarding other, possible challenges and tasks for Visuelle 2.0.

Acknowledgements This work was partially supported by the Italian MIUR through PRIN 2017 - Project Grant 20172BH297; "I-MALL - improving the customer experience in stores by intelligent computer vision” and the MIUR project "Dipartimenti di Eccellenza 2018-2022".
4. Appendix

The main paper discusses two contributions: 1) the novel Visuelle 2.0 dataset and 2) the advantage of utilising a multi-modal approach for short-observation new product forecasting (SO-fore). This additional material adds to each one of these main topics, as follows:

**The Visuelle 2.0 dataset** The following topics will be faced in Sec. 4.1:

- How Visuelle 2.0 compares with respect to the current datasets of forecasting for fast-fashion, exploring the literature of time-series forecasting and computer vision in this area (Sec. 4.1.1);

- A showcase of the data contained in Visuelle 2.0, including examples of products, the associated time series data, image data and the text data. An excerpt of these data is reported in Fig. 7. Subsequently, we will show some examples of *exogenous data* i.e., the Google Trends associated to a given product, and the weather reports associated to a given shop; finally, customer purchase data will be presented (Sec. 4.1.2);

- A list of possible challenges that can be studied on Visuelle 2.0 will be presented, and specifically: more tasks related to new product short observation forecasting; the new product demand forecasting; the product recommendation (Sec. 4.1.3).

**The SO-fore problem** In Sec. 4.2, the Cross-Attention RNN will be detailed, with a graphical representation that illustrates its components (Sec. 4.2.1).

4.1. The Visuelle 2.0 dataset

4.1.1 Datasets for forecasting in fast fashion

The Visuelle 2.0 dataset can be related to two scientific fields: 1) the one of forecasting for (fast) fashion [9,30,39], with particular emphasis on those works that exploit deep learning techniques [11, 24]; 2) the recent field at the intersection between computer vision and fashion [8], with special emphasis on the task of *popularity prediction* and *fashion forecasting*. In both cases, Visuelle 2.0 innovates for specific reasons which will be detailed in the following, in two separate sections. In both the cases, Visuelle 2.0 has an unprecedented richness of data which other datasets do not possess, such as customer purchase information, and the different exogenous data.
Forecasting for (fast) fashion  The most used forecasting techniques for fast fashion incorporate classical ARIMA, SARIMA, exponential smoothing [5], regression [31], Box & Jenkins [4] and Holt Winters [43], as reported in the recent review of [24]: machine learning approaches (decision trees, random forests, SVMs, neural networks) are at their infancy on this topic and most importantly they are not considering multi-modal data, but only time-series. This has naturally created an abundance of datasets for time-series analysis for sale forecasting [33] and demand forecasting [32], and the absence of datasets with images included, which we are filling with Visuelle 2.0. As a notable exception, the work of [11] proposes a set of techniques for demand forecasting, in which images are taken into account by an Attention-based RNN framework, which we also utilise for SO-fore as explained in Sec. 4.2.1. Unfortunately, the dataset on which they perform their experiments is not publicly available, while Visuelle 2.0 will be made publicly available.

Intersection between computer vision and fashion  The peculiarity of this field is the exploration of multi-modal data (images, text, time-series) for prediction tasks. In general, computer vision approaches have been considered for the task of popularity prediction or fashion forecasting [8]. In both the cases, the ground truth signal is built on top the public ratings obtained on online platforms such as Chicopia.com [36, 44], Lookbook.nu [23] or Amazon [1], which consider outfits [23, 36, 36, 44], or several outfits exhibiting the same style [1]: an outfit or a style is popular if it receives a high rating in terms of number of “likes” or “stars”. In the case of Visuelle 2.0, we can assume that a product is more popular than others of the same category, if in the same season, it has sold more. In this setup, our dataset allows to be more fine-grained, since one can predict the popularity, in terms of sales, of a single product. Also, Visuelle 2.0 represents the very first dataset which permits multi-modal analysis on the data of a real fast fashion company, meaning that approaches which succeed on this benchmark can be directly applicable on the fast fashion market.

4.1.2 Visuelle 2.0: a showcase

Example of products  In Sec.2 of the main paper we have given some statistics about the products which are in the Visuelle 2.0 dataset. Here we will give some qualitative examples of their image data, text attributes and associated time series: product sales, inventory position, Restock flag, and discount, the latter omitted in the main paper due to the lack of space. Formally, following the notation of the main paper, given a product \( i \) at a retail store \( r \), we refer to its discount signal as \( D(i, r, t) \) where \( t \) refers to the \( t \)-th week of market delivery, with \( i = 1, \ldots, N, r = 1, \ldots, L \) and \( t = 1, \ldots, K \). The discount signal is expressed as a percentage, describing how much a particular item is discounted; for example, \( D(i, r, t) = 20 \) indicates that the initial price defined for a product \( i \) in the retail store \( r \) was discounted by a 20% at time \( t \). Fig. 7 showcases all of these data for three products of season AW19. Other figures can be found at the end of this additional material, reporting products of other seasons (Fig. 12, Fig. 13 and Fig. 14).

Exogenous data  Exogenous data is often neglected as a resource within datasets, especially in forecasting. This is due to their nature, since they are, by definition, coming from an external phenomenon that is not directly related with the data being analysed. Nevertheless, adding exogenous variables such as weather data [3, 38] or popularity data [35, 37] to forecasting models has proven extremely beneficial in terms of forecasting performance. For this reason, we provide in Visuelle 2.0 multivariate exogenous data both for the weather and popularity, in the form of detailed weather reports (Fig. 8) and Google Trends (Fig. 9). An more profound explanation for both examples is provided in the respective captions.

Customer purchase data  Quoting Sec.2 and Fig.6 of the main paper, among the 667086 total registered users of Nuna Lie, 6k users have bought continuously a total of 25
Figure 9. Example of the exogenous Google Trends data available in Visuelle 2.0 for two completely different products. The signals can have regular trend and seasonality (left) or be stationary and seem noisy (right). This can prove helpful for forecasting models, because it allows to understand both global and cyclical popularity and therefore anticipate sales.

products over 4 seasons. In Fig. 10 we have a random excerpt of 10 of these users, with a random subset of 9 purchased items each (no cherry picking); in many cases, personal styles do emerge, showing systematic preferences on diverse attributes, as written in the caption of the figure.

4.1.3 Challenges on Visuelle 2.0

In this paper, we explored the problem of short-observation forecasting on Visuelle 2.0, with the precise focus of showing the benefit of the image data on this task. Obviously, we are far from saturating the performance, encouraging further improvements. These could be provided by diverse techniques (exploiting LSTM or transformer-based architectures), or including additional exogenous data, available in Visuelle 2.0. Google Trends and weather reports, in fact, are signals which have been shown elsewhere to be predictive [3, 35, 37, 38], so this should be a natural next step. Other challenges which can be experimented on Visuelle 2.0 are listed in the following.

• **Demand forecasting.** Forecasting demand is a crucial issue for driving efficient operations management plans [13, 30, 41]. This is especially the case in the fast fashion industry, where demand uncertainty, lack of historical data, variable ultra-fast life-cycle of a product and seasonal trends usually coexist [20, 27]. In rough terms demand forecasting outputs the amount of goods to buy from the suppliers. This amount is then distributed among the different retailers, with the aim of avoiding zero-stocks or excessive unsold inventory. In this paper we show a glimpse of demand forecasting on Visuelle 2.0, at the level of single shop (i.e. predicting how much a single shop will need during the next season), adopting the recent RNN-based approach of [11] on time series and time series + image. In [37] we report some results on the aggregated sig-
Figure 10. A random sampling of users/purchases. Personal styles do emerge: users 1 and 10 have no trousers in their logs, user 6 has bought almost short sleeves and no trousers, while user 7 seems to prefer long sleeves and several trousers; user 10 has a marked preference for light yellow-grayish colors.

The SO-fore problem

4.2. Cross-Attention RNN

The Cross-Attention RNN [11], can be described as an autoregressive, sequence-to-sequence neural network that
Figure 11. A visual description of the Cross-Attention RNN model, the neural network architecture used in our SO-fore experiments. The model diagram is taken from [11].

tries to understand the different, non-linear relationships in the various data modalities and then perform predictions by understanding which part of the data is most important for the forecasting task. The attention modules constitute a large part of the model and are exactly as in [2], where at each decoding step we try to attend to the encoder features based on the current decoder hidden state.

The encoder starts by embedding each input modality into a common feature space $R^D$. The input observation sales (time series) are first passed through an additional self-attention layer [42], differently from the original work in [11] and then projected through a fully connected layer. This helps filter out initial noise from past sales observations. The image and textual tags are embedded processed using a ResNet-101 [15] and learnable embedding layers respectively. The temporal features extracted from the product’s release date are also embedded using learnable embedding layers.

Cross-attention RNN works, by default, in an autoregressive manner, therefore at each decoding step three different additive attention modules are applied. These modules allow the decoder hidden state to attend to the time series embedding, the image embedding and most importantly to the concatenated, multi-modal embedding. A residual learning approach [15] is applied to allow the network to scale better with the number of hidden layers and also learn to ignore null contributions from the attention mechanism. After each decoding step, the GRU hidden state is updated based on the last processed input. For extensive details on this model, we refer to [11].

The model is trained with a batch size of 128 and MSE (Mean Squared Error) loss function, using the Adafactor optimizer, on two NVIDIA RTX TITAN GPUs. During training, we apply dropout after each embedding module and also apply teacher forcing at random with a probability $p_{tf} = 0.5$. 
Figure 12. Examples of different products that have been sold in the shop SHOP3 during the SS17 season. The figure reports (from left to right): the product’s image; textual tags; sales time series; restocking information; inventory time series; and discount time series.

Figure 13. Examples of different products that have been sold in the shop SHOP28 during the AW17 season. The figure reports (from left to right): the product’s image; textual tags; sales time series; restocking information; inventory time series; and discount time series.
Figure 14. Examples of different products that have been sold in the shop SHOP51 during the SS18 season. The figure reports (from left to right): the product’s image; textual tags; sales time series; restocking information; inventory time series; and discount time series.
References

[1] Ziad Al-Halah, Rainer Stiefelhagen, and Kristen Grauman. Fashion forward: Forecasting visual style in fashion. In Proceedings of the IEEE international conference on computer vision, pages 388–397, 2017. 1, 6

[2] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. Neural machine translation by jointly learning to align and translate. Jan. 2015. 3rd International Conference on Learning Representations, ICLR 2015 ; Conference date: 07-05-2015 Through 09-05-2015. 4, 9

[3] Samaneh Beheshti-Kashi, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen, and Michael Teucke. A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering, 3(1):154–161, 2015. 3, 6, 7

[4] George EP Box, Gwilym M Jenkins, Gregory C Reinsel, and Greta M Ljung. Time series analysis: forecasting and control. John Wiley & Sons, 2015. 6

[5] Robert Goodell Brown. Smoothing, forecasting and prediction of discrete time series. Courier Corporation, 2004. 6

[6] Samanbeh Beheshti-Kashi, Hamid Reza Karimi, Klaus-Dieter Thoben, Michael Lütjen, and Michael Teucke. A survey on retail sales forecasting and prediction in fashion markets. Systems Science & Control Engineering, 3(1):154–161, 2015. 3, 6, 7

[7] Audrey Aurelia Chang, Jihad Fahri Ramadhan, Zainina Kyla Shafira Adnan, Bayu Kanigor, Edy Irwan, et al. Fashion trend forecasting using machine learning techniques: A review. In Informatics, volume 8, page 49. Multidisciplinary Digital Publishing Institute, 2021. 2

[8] Wen-Huang Cheng, Sijie Song, Chieh-Yun Chen, Shintami Chusnul Hidayati, and Jiayi Liu. Fashion meets computer vision: A survey. ACM Computing Surveys (CSUR), 54(4):1–41, 2021. 5, 6

[9] Tsan-Ming Choi, Chi-Leung Hui, Na Liu, Sau-Fun Ng, and Yong Yu. Fast fashion sales forecasting with limited data and time. Decision Support Systems, 59:84–92, 2014. 2, 4, 5

[10] Yuyuan Ding, Yunshan Ma, Lizi Liao, Qimin Peng, Xiaohui Hu, and Rong Yang. Fashion trend forecasting: Principles and Practice. OTexts, Australia, 2nd edition, 2018. 4

[11] Majid Kharfan, Vicky Wing Kei Chan, and Tugba Fir dolas Efendigil. A data-driven forecasting approach for newly launched seasonal products by leveraging machine-learning approaches. Annals of Operations Research, 303(1):159–174, 2021. 1, 2

[12] Eundeok Kim, Ann Marie Fiore, Alice Payne, and Hyejeong Kim. Fashion trends: Analysis and forecasting. Bloomsbury Publishing, 2021. 1

[13] Ling Lo, Chia-Lin Liu, Rong-An Lin, Bo Wu, Hong-Han Shuai, and Wen-Huang Cheng. Dressing for attention: Outfit based fashion popularity prediction. In 2019 IEEE International Conference on Image Processing (ICIP), pages 3222–3226. IEEE, 2019. 6

[14] Ana LD Loureiro, Vera LM Miguelés, and Lucas FM da Silva. Exploring the use of deep neural networks for sales forecasting in fashion retail. Decision Support Systems, 114:81–93, 2018. 1, 2, 5, 6

[15] Shaohui Ma and Robert Fildes. Retail sales forecasting with meta-learning. European Journal of Operational Research, 288(1):111–128, 2021. 1, 2

[16] Yuyuan Ding, Yunshan Ma, Lizi Liao, Wei Kung Wong, and Tat-Seng Chua. Knowledge enhanced neural fashion trend forecasting. In Proceedings of the 2020 International Conference on Multimedia Retrieval, pages 82–90, 2020. 1

[17] Dennis Maaß, Marco Spruit, and Peter de Waal. Improving short-term demand forecasting for short-lifecycle consumer products with data mining techniques. Decision analytics, 1(1):1–17, 2014. 2, 7

[18] Utkarsh Mall, Kevin Matzen, Bharath Hariharan, Noah Snavely, and Kavita Bala. Geostyle: Discovering fashion trends and events. In ICCV, 2019. 1
[29] Francisco Martínez, M.P. Frías, María Pérez, and Antonio Rivera Rivas. A methodology for applying k-nearest neighbor to time series forecasting. *Artificial Intelligence Review*, 52, 10, 2019. 4

[30] María Elena Nenni, Luca Giustiniano, and Luca Pirolo. Demand forecasting in the fashion industry: a review. *International Journal of Engineering Business Management*, 5:37, 2013. 5, 7

[31] Alex D Papalexopoulos and Timothy C Hesterberg. A regression-based approach to short-term system load forecasting. *IEEE Transactions on power systems*, 5(4):1535–1547, 1990. 6

[32] Bohdan Pavlyshenko. Using stacking approaches for machine learning models. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pages 255–258. IEEE, 2018. 6

[33] Bohdan M Pavlyshenko. Linear, machine learning and probabilistic approaches for time series analysis. In *2016 IEEE First International Conference on Data Stream Mining & Processing (DSMP)*, pages 377–381. IEEE, 2016. 6

[34] Shravan Sajja, Nupur Aggarwal, Sumanta Mukherjee, Kushagra Manglik, Satyam Dwivedi, and Vikas Raykar. Explainable ai based interventions for pre-season decision making in fashion retail. In *8th ACM IKDD CODS and 26th COMAD*, pages 281–289. 2021. 1, 2

[35] Emmanuel Sirimal Silva, Hossein Hassani, Dag Øivind Madsen, and Liz Gee. Googling fashion: forecasting fashion consumer behaviour using google trends. *Social Sciences*, 8(4):111, 2019. 3, 6, 7

[36] Edgar Simo-Serra, Sanja Fidler, Francesc Moreno-Noguer, and Raquel Urtasun. Neuroaesthetics in fashion: Modeling the perception of fashionability. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 869–877, 2015. 6

[37] Geri Skenderi, Christian Joppi, Matteo Denitto, and Marco Cristani. Well googled is half done: Multimodal forecasting of new fashion product sales with image-based google trends. *arXiv preprint arXiv:2109.09824*, 2021. 2, 3, 4, 6, 7

[38] Donald Sull and Stefano Turconi. Fast fashion lessons. *Business Strategy Review*, 19(2):4–11, 2008. 3, 6, 7

[39] Ian Malcolm Taplin. Global commodity chains and fast fashion: How the apparel industry continues to re-invent itself. *Competition & Change*, 18(3):246–264, 2014. 2, 4, 5

[40] Sébastien Thomassey. Sales forecasting in apparel and fashion industry: A review. *Intelligent fashion forecasting systems: Models and applications*, pages 9–27, 2014. 4

[41] S Thomassey et al. Intelligent demand forecasting systems for fast fashion. In *Information systems for the fashion and apparel industry*, pages 145–161. Elsevier, 2016. 4, 7

[42] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. 9

[43] Peter R Winters. Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342, 1960. 6

[44] Kota Yamaguchi, Tamara L Berg, and Luis E Ortiz. Chic or social: Visual popularity analysis in online fashion networks. In *Proceedings of the 22nd ACM international conference on Multimedia*, pages 773–776, 2014. 6

[45] Yong Yu, Tsan-Ming Choi, Kin-Fan Au, and Chui-Yan Kwan. A web-based system for fashion sales forecasting. *Research Journal of Textile and Apparel*, 2008. 4