

# Demand forecasting in the fashion business — an example of customized nearest neighbour and linear mixed model approaches

Joanna Henzel\*, Łukasz Wawrowski<sup>||</sup>, Anna Kubina<sup>||</sup>, Marek Sikora\*, Łukasz Wróbel\*

\* Silesian University of Technology  
Department of Computer Networks and Systems  
ul. Akademicka 2A, 44-100 Gliwice, Poland  
joanna.henzel@polsl.pl

<sup>||</sup> Łukasiewicz Research Network – Institute of Innovative Technologies EMAG  
ul. Leopolda 31, 40-189 Katowice, Poland  
lukasz.wawrowski@emag.lukasiewicz.gov.pl

**Abstract**—The fashion industry is characterised by the need to make demand forecasts in advance and for highly volatile products for which we often have no sales history at the time the forecasts are made. For this reason, it is necessary to propose forecast mechanisms that can cope with the given conditions. Such forecasts can be based on expert predictions for generalized product categories. In this case, the task of machine learning forecasting methods would be to divide the aggregate prediction into forecasts for individual products, in each colour and size. In the paper, we present several approaches to this specific task. We present the use of the naive method, custom nearest neighbour approach, parametric linear mixed model and an ensemble approach. Overall, the best results we obtained for the ensemble method. Our research was based on real data from fashion retail.

## I. INTRODUCTION

**D**EMAND forecasting in the fashion industry is characterised by specific conditions. In this industry, it is necessary to be able to forecast for a long horizon in time. This is because many products are ordered from other parts of the world, where they are produced on a large scale. Products must be ordered in advance so that the entire process of production, delivery, promotion and distribution to shops in different countries can take place on time.

On the other hand, fashion products are highly variable over time. Rarely are products sold for several seasons. Most of them appear on sale only in one sales season, which translates into a very short sales history for a particular product. This relates to natural seasonality due to changing seasons, but also to trends that can vary significantly from one year to the next.

The combination of the need to order products in advance and the volatility of the products being sold means, that we have to make predictions of demand for products that mostly have not previously been on sale. These are difficult conditions for making forecasts. Sales forecasts could be made by experts based on their domain knowledge, but such experts would

also have difficulty determining future sales of a particular product, in a particular size and colour, each week of the following season. Additionally, with many shops, this would be a very time-consuming process. In this case, a good idea is to use automated forecasting methods, which would be based on statistical models or machine learning models.

The research described in this article dealt with the problem of sales forecasting in the fashion industry. What is important, in our research we obtained sales forecasts for product categories and our task was to build on these general predictions the forecasts for individual products (described by specific product type, colour and size). Sales forecasts were made on weekly aggregates. Forecasts for individual categories were provided to us by a business partner and are proprietary. They are not the subject of this article. However, it should be noted that the forecast data could be provided by an expert. Making assumptions about aggregate sales within more general product categories should be a more manageable and less time-consuming task for an expert.

In this paper, we would like to present several approaches to making demand forecast for the specific products from fashion retail based on higher level forecasts. We present naive method, custom nearest neighbour approach, parametric linear mixed model and an ensemble approach.

## II. RELATED WORK

Demand forecasts are the basis of most decisions in supply chain management. Forecasting methods applied to this problem are based on both domain knowledge and historical data analysis. In the former approach, the retailers knowledge is utilised to develop demand prognosis. In the later approach statistical and machine learning based methods are intensively used [1]–[3].

Our research is focused on the data describing the demand in a fashion sector. In this area, research on machine

learning methods application were carried out. In the article [4] the authors present the use of deep neural networks for sales forecasting in fashion industry, especially forecasting the sales of new fashion products. They compare the deep learning approach with other algorithms e.g. Decision Trees, Random Forest, Support Vector Regression, Artificial Neural Networks and Linear Regression. The authors found deep learning models to have good performance, however for some metrics the models were not significantly better than some simpler techniques. It shows that in this sector the simpler and interpretable methods may obtain good results.

Pre-season forecasting in fashion retail was also discussed in [5]. The authors point out, that the typical time-series methods can be introduced for this problem, however in most cases the retailers need to forecast new products, so in historical data there is no time-series linked precisely to the forecast product. It means that the retailers need to base on their intuition and the historical data of similar products. The authors also highlight the need for creating explainable forecasting models. Because of explainability and interpretability, many stakeholders still use a very simple and naive approach for forecasting new products — averaging sales of similar products for the new product. If AI method could be used in this sector it is important to give explanation, how the decision was made.

The authors of [6] focus on fast fashion sales forecasting where the data is limited and within limited time. The authors propose a novel algorithm — Fast Fashion Forecasting (3F) — which combines extreme learning machine (ELM) and the grey model (GM). Extreme learning machine was also used for fashion retailing forecasting in the [7]. In the aspect of Fast-Moving Consumer Goods (FMCG), the authors of [8] showed benefits of applying Machine Learning methods in creating demand forecasting models.

### III. PROBLEM STATEMENT

In our research, we were working with the specific problem of forecasting demand for fashion products. From our business partner, we obtained their forecasts of sale for pre-defined categories of products. Our task was to divide these forecasts to the forecasts for each separate product that belongs to the category. A unique product is described by a unique combination of attributes: product type, colour and size. The most important requirement was to use provided forecasts, not to create from scratch forecasts for each separate product. The second important assumption was the horizon of the forecasts. We were working with long-term predictions — forecasts had to be made for 29 weeks ahead. Predictions were made for weekly sale aggregates.

In the obtained dataset, we had real historical sale data from a fashion brand with a shop chain. The dataset contained sale data from January 2016 to November 2021. For the period from April 2021 to November 2021, we got from our business partner not only predictions for the whole category, but also their predictions for unique products. Because of this, we considered this part of a dataset as our test dataset, on which the experiment's result will be calculated. The data before

were our train dataset. In the data, we could observed strong seasonality — big peaks of sales before summer. It should also be noted that data contains sales from COVID-19 pandemic and lock-downs connected with this phenomena. Figure 1 presents demand for an example product.

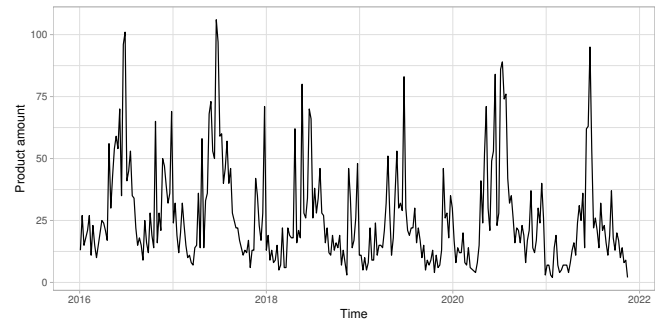


Fig. 1. Demand of selected product

In the data preparation process we created new attributes, that could be defined for new products, that would be in sale many weeks ahead. Because of forecasting demand with a long time-horizon, we couldn't use data about weather or information about sale from weeks just before the forecast week. In our dataset, each observation described one week of a sale of a specific product (a unique product described by a unique combination of attributes: product type, colour and size). Each week, for each product, was described by the attributes listed below.

- Attributes connected with a product in sale: product type, size, main colour and colour undertone.
- Attributes connected with time of a sale (describing the week of an observation): week number, season, quarter of a year, number of days to the closest event (Christmas, Easter, national holiday, Valentine's Day etc.), number of days that passed from the closest event, number of events/holidays that happened during the considered week, number of events/holidays that happened during three weeks — in the week under consideration, the preceding week and following week, binary feature that described if it is the last week of a year.
- Attributes connected with sale: number of sold units, number of sold units within category, fraction of the sale in a given category that account for the sale of the product concerned.
- Attributes connected with trend and seasonality: we used prophet tool [9] to decompose data from train dataset into trend and seasonality for each product. Obtained results were utilized for test dataset.
- ID attributes: product name, category name, date of a first day from a considered week.

In our dataset, we didn't have information about price or planned promotions.

#### IV. METHODS

In this chapter, we want to present some selected approaches for breaking down the higher level forecasts (forecasts for one category of products) into lower level forecasts (per product forecasts). We present naive method, non-parametric custom nearest neighbour approach, parametric linear mixed model and an ensemble approach. The experiments were performed using R and Python languages.

##### A. Naive

Firstly, we proposed a naive method of splitting the forecast for the whole category into a forecast for the individual product. This involved finding the weight by which the category forecast was to be multiplied to produce the product forecast.

Suppose the forecast was made for the week  $n$  of year  $Y$  for a selected product  $p$ . The product belongs to category  $c$ . In such case, we looked up the sale of the product in week  $n$  in the previous year ( $Y - 1$ ). Then, we divided it by the total sales in the category  $c$  in week  $n$  in the previous year. The given fraction was the weight by which the forecast for the category for week  $n$  of year  $Y$  was multiplied. The result is the forecast for the product.

An extension of this proposed solution was to determine weight based not only on the week the year before, but also on its week preceding and the week following. This ensured that the weights were averaged and minimized the impact of outliers on the resulting forecast. The forecast for the product was calculated as follows:

$$y_{p,n,Y} = \frac{1}{3} \sum_{i=-1}^1 \frac{s_{p,n+i,Y-1}}{s_{c,n+i,Y-1}} y_{c,n,Y} \quad (1)$$

where  $s$  means real sale,  $y$  is a forecast,  $p$  is a product, for which we calculate the forecast,  $c$  is a category of a product,  $n$  is a week number,  $Y$  is a year. For example, if the year before the forecast week, sale of a product was 2% of total category sale, while the week before it was 6%, and the following week it was 7%, the final weight by which the forecast for the entire category was multiplied was the average of these values — 0.05. A naive method with averaged weights was considered in further stages of the work.

The presented naive approach has one big disadvantage — it works only on a product with history of sale. The naive method could be used for new products only with expert's help. The expert could indicate which product, with historical sales, the new product is similar to. Then, we would assume, that we could use the weight obtained from historical sale of a similar product, to get the forecast for a new product. However, we felt that it would be useful to focus on a solution that would allow us to forecast the demand for new products, with minimal expert involvement in the process.

##### B. Nearest neighbour (KNN)

As a more advanced forecasting procedure, we propose a nearest neighbour (KNN) approach. This is a non-parametric technique widely used in classification tasks, however we used

it as a method for finding similar observation from historical data. This algorithm, given an input vector, calculates distance (based on a chosen distance metric) to observations from a training dataset. The one observation, whose conditional attribute vector is most similar to the new feature vector, is considered to be its nearest neighbour.

In this approach, we used the dataset described in section III. In order to use some attributes in the KNN approach, additional input data had to be provided. This situation has occurred for "size" attribute. Our dataset included different fashion products, e.g. shirts, bras and socks. These products have different clothing sizes. In order to calculate distance between different observations using the attribute "size", we changed original sizes to numerical values based on our training set, i.e. historical data. We divided the historical dataset by size — a separate subset for each category of products, which have different clothing sizes. Then, for each of the sizes in the set, we determined the percentile values within those subsets. That is, if we forecasted sales for socks in size 36/38, we converted size 36/38 into the numerical value — percentile that this size represented in the historical data.

Additional input data was also provided for attribute "product type". This attribute had nominal values, that were identifiers of the fashion types — the values did not hold any additional meaning. In order to calculate distance between different product types, we proposed using *product types dissimilarity table*. In our *product types dissimilarity table*, we provided the distance between different product types. The distance was equal to 0, if we calculated the distance between products that had the same product type. The distance was equal to 0.5, if we calculated the distance between products that belonged to the same clothe category but are of different product type, e.g. both products were shirts, but they were of different type. The distance was equal to 1, if we calculated the distance between products that belonged to different clothe category, e.g. one product is a shirt and the other is socks.

In the KNN approach, we used the following attributes: week, product type, size, main colour, colour undertone, season, quarter of a year, number of days to the closest event, number of days that passed from the closest event, number of events/holidays that happened during the considered week, number of events/holidays that happened during the week under consideration, the preceding week and following week.

The schema of the proposed procedure is presented in Figure 2. The KNN method is called with the parameter  $k=3$ . This parameter value was selected from the set of values  $k=\{1,2,3\}$ , as the value giving the best model results.

It can be noted that this method is based on a similar assumption as our naive method. In the naive approach, we looked for the weight in the history of a given product exactly one year before. In the KNN method, the weight is determined in the same way — it is a fraction of the product's sales to sales in the entire category. The difference is that in the KNN method, we refer to a week that did not necessarily occur exactly one year earlier. Additionally, the nearest neighbour (the nearest "week") may refer to historical data for a different

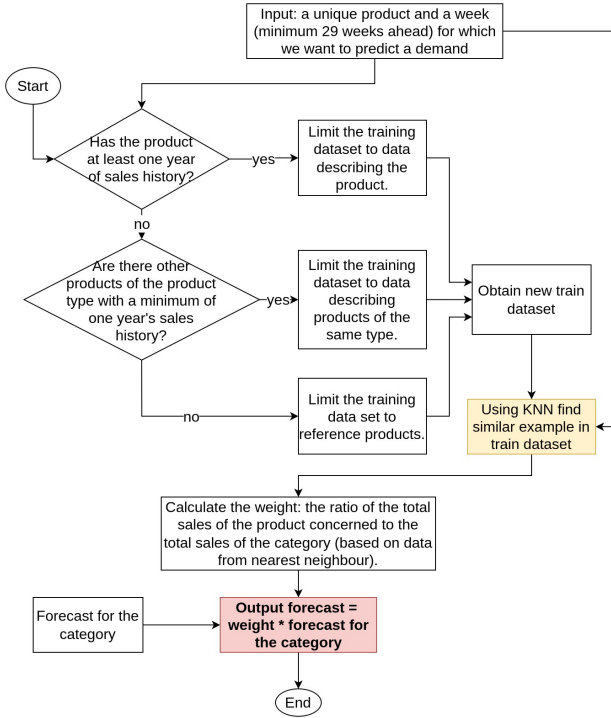


Fig. 2. The schema of the custom nearest neighbour procedure for obtaining the forecasted demand of a product.

product than the one for which the forecast is made.

### C. Linear mixed model (LMM)

Linear mixed models are an extension of simple linear regression models and can be used for data with a hierarchical structure which is observed in fashion. These models incorporate fixed and random effects:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \boldsymbol{\varepsilon} \quad (2)$$

where  $\mathbf{y}$  is a vector of outcome variable,  $\mathbf{X}$  is a matrix of predictors,  $\boldsymbol{\beta}$  is a vector of fixed-effects regression coefficient,  $\mathbf{Z}$  is a design matrix for random effects,  $\mathbf{u}$  is a vector of random effects and  $\boldsymbol{\varepsilon}$  is a vector of residuals [10].

In this approach, a model for each product group was estimated with random effects defined by product, size, and color. In groups where there was only one product, random effects were estimated only for size and color. Due to strong asymmetry of outcome variable (forecast weight) Box-Cox transformation [11] was applied.

We assure that there are significant differences between outcome variable and groups defined by product, size, and color using ANOVA. The significance of random effects was confirmed based on the permutation test.

As predictors in linear mixed models, we used attributes connected with product in the sale, time of a sale, trend, and seasonality (described in section III).

### D. Ensemble approach

In the ensemble approach, we combined the results from the non-parametric KNN method with the results from the parametric linear mixed model. The forecast returned was the average of forecasts provided by these two models.

## V. RESULTS

In this section, we present results obtained using methods described in section IV. We predicted demand for 7 categories of products and aimed for a reduction of mean absolute error (MAE) relative to baseline. Detailed results are presented in table I.

TABLE I  
MEAN ABSOLUTE ERROR OF PREDICTED DEMAND IN TEST DATASET

Product category	Baseline	Naive	KNN	LMM	Ensemble
Category 1	3.23	3.77	3.67	3.52	3.59
Category 2	9.68	11.61	11.01	9.29	9.72
Category 3	6.86	7.04	7.12	6.96	6.71
Category 4	6.44	8.43	8.64	7.74	8.12
Category 5	20.23	17.48	20.33	21.09	19.77
Category 6	11.55	11.51	11.48	11.05	10.70
Category 7	4.17	4.70	4.46	4.04	4.07

For the second and seventh categories the best results were obtained for LMM method. For categories third and sixth the lowest errors were obtained for ensemble approach. In the sixth category, we obtain gain in precision for all methods. In the case of the seventh category, both LMM and ensemble approaches have lower errors. For the fifth category also naive approach performed better than the baseline. The first and fourth categories were a little bit problematic because any of the proposed methods do not perform better than baseline. Overall, the biggest improvement is observed for the ensemble methods because for 4 from 7 categories this approach gave better results than the baseline.

In the next step, we tried to investigate what could be a possible reason for such results. Figure 3 presents differences between real and predicted demand for each method and category.

We observed that the fourth and fifth categories were characterized by a much lower number of observations compared to the rest categories. The small sample size could be one of the reasons why it was impossible to obtain a lower MAE than the baseline result for the fourth category. In the case of the fifth and sixth categories, we noticed a few outlier values and a flattened distribution of these errors compared to other categories. In the rest categories, the average of differences between true and predicted demand is close to 0.

The last part of the study was the analysis of stability over time. The Figure 4 shows differences between real demand (bold black line) and forecasts obtained with different approaches for one selected product.

In this particular case, the lowest MAE was obtained for the LMM approach. We can observe that almost all methods predicted lower than actual demand just before summer. The best estimation in this period was observed for KNN method,

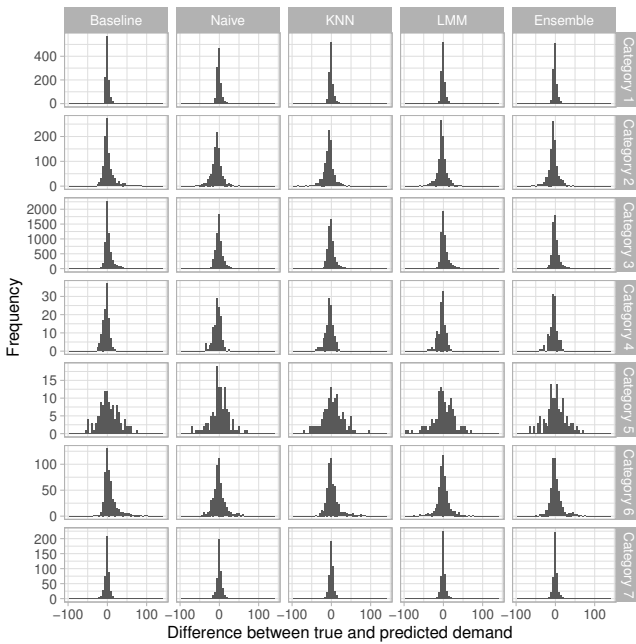


Fig. 3. Distribution of differences between true and predicted demand

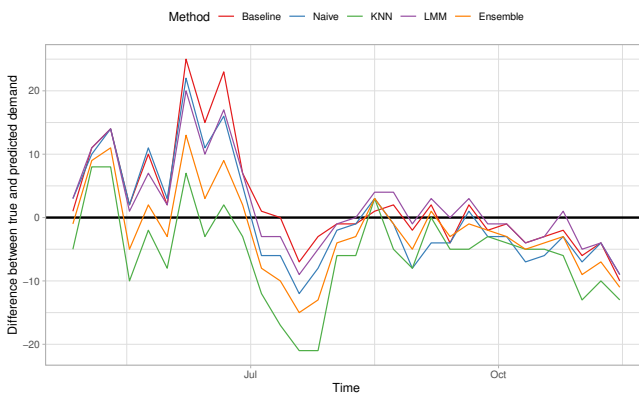


Fig. 4. Differences between true and predicted demand for selected product

however it gives the biggest error in summer. Generally, the proposed approaches underestimate seasonal peaks for all products in our dataset.

## VI. CONCLUSIONS AND FUTURE WORKS

Demand forecasting in the fashion business could be problematic due to short product history — in many cases, product is sold for one season. In our case we have access to longer time series nevertheless this data contains unexpected issues connected with COVID-19 such as lock-downs and changing habits of customers.

The proposed solution was based on the share of sale estimation and multiplication of the result by the provided forecast for the category. In almost all cases, utilized methods gave better results than the baseline provided by our business

partner. However, it should be noted that in such an approach we can highlight two sources of possible error — firstly at the method level connected with historical data and predictions and secondly at the category forecast level which differ from real category demand.

Future works explore further the topic of demand forecasting in various business cases: what to do in case of short historical data or how to forecast demand for a totally new products using existing data. We will also investigate other statistical methods designed for this purpose and optimize scope of conditional attributes used by techniques presented in this work.

## ACKNOWLEDGMENT

The work was carried out in part within the project co-financed by European Funds entitled “Decision Support and Knowledge Management System for the Retail Trade Industry (SensAI)” (POIR.01.01.01-00-0871/17-00). The research leading to these results received funding by Young Researchers funds of Department of Computer Networks and Systems, Faculty of Automatic Control, Electronics and Computer Science, Silesian University of Technology, Gliwice, Poland (project no.: 02/120/BKM22/0021). This work was partially supported by the European Union through the European Social Fund (grant POWR.03.05.00-00-Z305).

## REFERENCES

- [1] M. Z. Babai, J. E. Boylan, and B. Rostami-Tabar, “Demand forecasting in supply chains: a review of aggregation and hierarchical approaches,” *International Journal of Production Research*, vol. 60, no. 1, pp. 324–348, 2022.
- [2] E. Hofmann and E. Rutschmann, “Big data analytics and demand forecasting in supply chains: a conceptual analysis,” *The International Journal of Logistics Management*, 2018.
- [3] M. Seyedan and F. Mafakheri, “Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities,” *Journal of Big Data*, vol. 7, no. 1, pp. 1–22, 2020.
- [4] A. L. Loureiro, V. L. Miguéis, and L. F. da Silva, “Exploring the use of deep neural networks for sales forecasting in fashion retail,” *Decision Support Systems*, vol. 114, pp. 81–93, oct 2018. doi: 10.1016/j.dss.2018.08.010
- [5] S. Sajja, N. Aggarwal, S. Mukherjee, K. Manglik, S. Dwivedi, and V. Raykar, “Explainable AI based Interventions for Pre-season Decision Making in Fashion Retail,” in *ACM International Conference Proceeding Series*, 2020. doi: 10.1145/3430984.3430995. ISBN 9781450388177 pp. 281–289.
- [6] T. M. Choi, C. L. Hui, N. Liu, S. F. Ng, and Y. Yu, “Fast fashion sales forecasting with limited data and time,” *Decision Support Systems*, vol. 59, no. 1, pp. 84–92, mar 2014. doi: 10.1016/j.dss.2013.10.008
- [7] M. Xia, Y. Zhang, L. Weng, and X. Ye, “Fashion retailing forecasting based on extreme learning machine with adaptive metrics of inputs,” *Knowledge-Based Systems*, vol. 36, pp. 253–259, dec 2012. doi: 10.1016/j.knsys.2012.07.002
- [8] “Machine learning in predicting demand for fast-moving consumer goods: An exploratory research,” *IFAC-PapersOnLine*, vol. 52, no. 13, pp. 737–742, 2019. doi: 10.1016/j.ifacol.2019.11.203
- [9] S. J. Taylor and B. Letham, “Forecasting at scale,” *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [10] J. Fox, *Applied regression analysis and generalized linear models*. Sage Publications, 2015.
- [11] G. E. Box and D. R. Cox, “An analysis of transformations,” *Journal of the Royal Statistical Society: Series B (Methodological)*, vol. 26, no. 2, pp. 211–243, 1964.