Demand Forecasting Using Ensemble Learning for Effective Scheduling of Logistic Orders

Katharina Lingelbach[1(✉)], Yannick Lingelbach[2], Sebastian Otte[3], Michael Bui[1], Tobias Künzell[4], and Matthias Peissner[1]

[1] Fraunhofer IAO, Fraunhofer Institute for Industrial Engineering IAO, Stuttgart, Germany
{katharina.lingelbach,michael.bui,
matthias.peissner}@iao.fraunhofer.de
[2] Powertrain Solutions Division, Manufacturing Technology, Robert Bosch GmbH, Stuttgart, Germany
yannick.lingelbach@de.bosch.com
[3] Neuro-Cognitive Modeling, University of Tübingen, Tübingen, Germany
sebastian.otte@uni-tuebingen.de
[4] Zeitfracht GmbH, Stuttgart, Germany
tobias.kuenzell@knv-zeitfracht.de

Abstract. We present forecasting models based on extreme gradient boosting to predict demand using real-world data of a German intermediary company in the media sector. The data set comprised the daily demand of 196,767 products from three years (mid-2017 to mid-2020) and meta information for each product including product type affiliation. Models were trained separately for each product type either to predict the demand on group or product level. For the latter, training was in rolling format based on the last 12 weeks to then predict the product's short-term demand one-week ahead. Performance, evaluated via the coefficient of determination, is especially precise for specific product types. Engineered features consisting of seasonal information, statistical indices, and general performing indices obtained via fuzzy c-means clustering over time improved the prediction. Especially, predictions for the upcoming week on product level are challenging but of high value for future business decisions regarding inventory planning and purchase orders.

Keywords: Demand forecasting · Ensemble learning · Boosting trees · Inventory management · Logistic · Real-world dataset · Application

1 Introduction

Accurate demand forecasting ensures optimal supply availability and plays, therefore, an important role in the cost-benefit analysis of online intermediaries. Typically, the goal is to minimize storage and opportunity costs in terms of lost sales by optimizing stock levels and the effective scheduling of logistic orders. In the previous decade, a growing field of research investigated the potential of predictive big data analytics to

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2021
T. Z. Ahram et al. (Eds.): AHFE 2021, LNNS 271, pp. 313–321, 2021.
https://doi.org/10.1007/978-3-030-80624-8_39

improve demand forecasting, inventory planning, delivery capability, and supply chain management [1–8].

Traditional statistical forecasting approaches are mainly linear, univariate time series analyses including exponential smoothing, regressions, or ARIMA models [6–10]. However, there is an ongoing debate whether these traditional forecasting approaches might be inferior to new ML approaches since the latter have the ability to learn and integrate non-linear relationships[1] for a more accurate prediction and suitability for large-scale applications [4, 6, 9–11]. Predictive capabilities of various machine learning methods have been evaluated, including k-nearest neighbors, support vector machines, ensemble learners (e.g., random forests and gradient boosting) as well as (deep) neural networks [6–12]. These methods can be combined with unsupervised clustering to improve demand forecasting performance [13–15]. Such clustering analyses allow to partition elements of a data set into subgroups based on their similarities (e.g., seasonality or performing measures of product groups). Thereby, they can be used to generate informative features or new subsets of data.

In this paper, we present forecasting models based on ensemble learning combined with fuzzy c-means clustering to predict the daily and weekly demand using data of a German intermediary company in the media sector. The company supplies more than 7,000 stores daily with products covering 20% of the German logistics volume in the media sector.

2 Real-World Use Case on Demand Forecasting for Logistic Orders

The data set comprises 196,767 products with two categories of data available per product: (1) *meta information*, describing the type of product (e.g., fiction or non-fiction), subtype (e.g., print or audiobook), and price, as well as (2) the *historical demand*, that is the demand per day of three years from mid-2017 to mid-2020.

We first compared two unsupervised cluster algorithms that group the time series of demand with similar patterns: k-means for time series and fuzzy c-means. The derived cluster affiliation can be used (a) as a further informative feature capturing seasonality and performance over time, and (b) as an important indicator for the company's product selection in the inventory.

In the next step, we used gradient boosted decision trees trained with the XGBoost algorithm, which is particularly well-suited for sparse data sets [16] as it provides regularization, enhanced tree pruning, and other tweaks to foster generalization.

To predict the future demand, we follow two approaches: First, for long-term forecasting on group level, models were trained on data from 2017 and 2018 to predict the year 2019. Second, for short-term forecasting on product level, data was shaped in rolling format using the last 12 weeks to predict the products' demand one-week ahead. Finally, we compared the demand forecasting results of the two approaches and explored informative features by evaluation of the feature importance score derived from the regression models. Implementations and analyses were performed in Python (3.8.6). Clustering and

[1] Non-linear effects in the demand behavior are particularly challenging and can be explained by competition among suppliers, the bullwhip effect, and mismatch between supply and demand [7].

demand forecasting were performed separately for each product type group. There are eight product type groups available: Fiction (1), Children and Youth (2), Travel (3), Non-Fiction (4), Arts and Humanities (5), Mathematics and Natural Science (6), Social Sciences (7), and School and Learning (8).

2.1 Advanced Feature Engineering and Unsupervised Time Series Clustering

Time series of the historical demand were first scaled using the MinMaxScaler (scikit-learn 0.23.2; [17]). Since we aimed to identify low-, mid-, and high-performing products within each product type group in the data set, cluster size was set to three and clustering was performed separately for each product type. We compared fuzzy c-means (fuzzy-c-means 1.2.4, [18]) and TimeSeriesKMeans (tslearn 0.5.0.5, [19]) with the squared Euclidean distance as metric and max_iter = 5. Especially, fuzzy c-means found three clearly distinguishable and fairly balanced clusters with stably low-, mid-, and high-performing products (see Fig. 1, right, and Table 1). Clusters differed significantly in terms of demand behavior over time on a weekly basis (see Fig. 1, left). Results of the time series k-means clustering suggest two clusters as optimal size with only few low- and mid-performing products. For the demand forecasting, cluster labels of the fuzzy c-means were used as an additional feature to improve the prediction.

Table 1. Size of low-, mid-, and high-performing clusters for the two methods.

Cluster	Fuzzy c-means			k-means for time series		
	Low	Mid	High	Low	Mid	High
Size	4453142	17861349	15823780	1638480	3151858	33347933

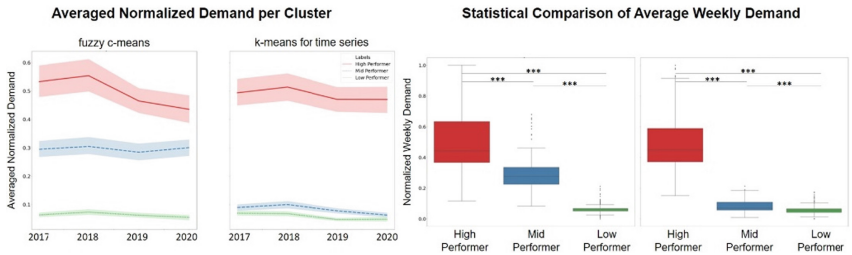


Fig. 1. Demand behavior of the clusters over time (left). Statistical cluster comparison of the weekly demand via the Kruskal-Wallis paired samples test with Bonferroni correction (right). Fuzzy c-means with High vs. Mid: $H = 148.6, p \leq .001$; Mid vs. Low: $H = 261.7, p \leq .001$; High vs. Low: $H = 264.1, p \leq .001$; k-means for time series: High vs. Mid: $H = 264.1, p \leq .001$; Mid vs. Low: $H = 50.2, p \leq .001$; High vs. Low: $H = 264.7, p \leq .001$.

2.2 Long-Term Demand Forecasting on Group Level

For a long-term forecasting on group level, we first calculated the accumulated daily demand in each product type group. The data set was divided into a training and test set

with a 66:33 split using the demand from 2017 and 2018 to train the model and test it on the year 2019, separately for each product type group. For the meta information, averages and modes per day were extracted from those products with a demand greater than 0. Numerical features were scaled using the MinMaxScaler, while nominal features were dummy coded (e.g., subtype or cluster affiliation). Additional features were extracted including calendar week and statistical indices of the medium-term historical demand from the last four weeks (mean, standard deviation, maximum, minimum, kurtosis and skewness). Hyperparameters of XGBoost were optimized using only training data in a 5-fold cross-validated random grid search and evaluated via the mean squared error (*MSE*; see Table 2 for final hyperparameters). Next, we applied boosted trees to predict future demand. Performance of the regression models was evaluated using the coefficient of determination R^2. The relevance of each feature in the prediction was derived from the regression models (in XGBoost, this is accessible via the attribute *feature importance).*

2.3 Short-Term Demand Forecasting on Product Level

For the short-term forecasting on product level, data was in rolling format based on the last 12 weeks in order to predict each product's demand one-week ahead. Here, we reduced the level of granularity of the prediction to save computational costs, since it is suggested that a weekly basis is sufficient for such applications [6, 10]. Data from each product was divided into training and test samples with an 80:20 split. Data was scaled and dummy coded as described above. In addition to the seasonal information, statistical indices of medium-term demand, cluster affiliation, and daily demand of the last 12 weeks was extracted to include the time series information in the feature space (t−1, t−2, …, t−12). Models were trained and tested separately for each product type group using the hyperparameters derived from the random grid search of the group level prediction. To evaluate the models' performance, we used the R^2.

3 Results and Discussion

Table 3 provides an overview of the models' performance for the train and test set in the group and product level prediction. The results indicate that demand forecasting is especially precise and good for specific types of products. In general, long-term prediction on group level is much more accurate with most of the R^2 scores between 0.583 and 0.804 ($M = 0.722$; $SD = 0.066$) compared to short-term prediction on product level ($M = 0.287$; $SD = 0.082$). Only for the product type group *School and Learning* (8) a slight shift in the predicted peak season produced a noticeable error in the long-term group level forecasting. Figure 2 and Fig. 3 visualize the demand forecasting of the product type groups *Fiction* (1) and *School and Learning* (8). The short-term weekly demand forecasting on product level is of high value for future business decisions regarding the inventory planning and purchase orders. However, a precise prediction is much more challenging due to data sparsity, long tail distributions, insufficient information for new releases, and hard-to-predict external influences seriously affecting consumer demands such as political, social, medial, or promotional events. Exploring further approaches to optimize hyperparameters of XGBoost, especially for the short-term prediction, (e.g.,

Table 2. Overview of the size of product type group, average demand per day, number of features, and hyperparameters derived from the random grid search. Further fixed hyperparameters: booster = gbtree, num_boost_round = 999, early_stopping_rounds = 50, learning_rate = 0.01. For detailed explanation on the hyperparameters and their functionalities, we refer to [20].

Type	Products	*M* (*SD*)/day	Features	Hyperparameter
1	49231	2.854 (14.395)	63	n_estimators = 400, max_depth = 5, gamma = 0.5, colsample_bytree = 0.9, min_child_weight = 1, subsample = 0.8, reg_lambda = 1.4, reg_alpha = 1.1
2	36359	2.965 (13.118)	63	n_estimators = 300, max_depth = 3, gamma = 0.5, colsample_bytree = 0.7, min_child_weight = 6, subsample = 0.8, reg_lambda = 1.4, reg_alpha = 1
3	9750	2.033 (4.961)	63	n_estimators = 400, max_depth = 3, gamma = 0.3, colsample_bytree = 0.7, min_child_weight = 3, subsample = 0.6, reg_lambda = 1.2, reg_alpha = 1.1
4	31691	2.875 (26.278)	62	n_estimators = 400, max_depth = 3, gamma = 0.3, colsample_bytree = 0.7, min_child_weight = 3, subsample = 0.6, reg_lambda = 1.4, reg_alpha = 1.2
5	22759	2.415 (7.483)	63	n_estimators = 500, max_depth = 3, gamma = 0.7, colsample_bytree = 0.8, min_child_weight = 1, subsample = 0.8, reg_lambda = 1.4, reg_alpha = 1.2
6	7724	2.686 (17.838)	63	n_estimators = 500, max_depth = 5, gamma = 0.7, colsample_bytree = 0.9, min_child_weight = 1, subsample = 0.7, reg_lambda = 1.2, reg_alpha = 1
7	14413	2.919 (11.843)	62	n_estimators = 500, max_depth = 3, gamma = 0.3, colsample_bytree = 0.7, min_child_weight = 3, subsample = 0.7, reg_lambda = 1.2, reg_alpha = 1.2
8	24840	5.108 (18.238)	62	n_estimators = 500, max_depth = 5, gamma = 1, colsample_bytree = 0.7, min_child_weight = 1, subsample = 0.6, reg_lambda = 1.2, reg_alpha = 1.1

via evolutionary algorithms), might potentially improve model performance. Moreover, additional features capturing seasonal information to improve the prediction of peak seasons should be explored.

To evaluate the informativeness of individual features for the demand forecasting, we investigated the tree's inherent importance attribute. We averaged the scores of each feature over the product type groups for the group and product level prediction. Figure 4

Table 3. Overview of the model performance of each product type group for long-term group level and short-term product level demand forecasting. The R^2 reveals the proportion of the variance in the outcome (here demand) explained by the variance in a set of independent features and ranges from 0 to 1. An $R^2 = 1$ means total label variance can be explained by the features.

Type	Group level		Product level	
	R^2 training	R^2 test	R^2 training	R^2 test
1	0.921	0.756	0.307	0.368
2	0.806	0.774	0.311	0.251
3	0.708	0.583	0.314	0.282
4	0.762	0.713	0.411	0.184
5	0.864	0.738	0.421	0.432
6	0.895	0.662	0.559	0.194
7	0.876	0.804	0.494	0.240
8	0.963	0.748	0.582	0.345
Mean (SD)	0.849 (0.080)	0.722 (0.066)	0.425 (0.104)	0.287 (0.082)

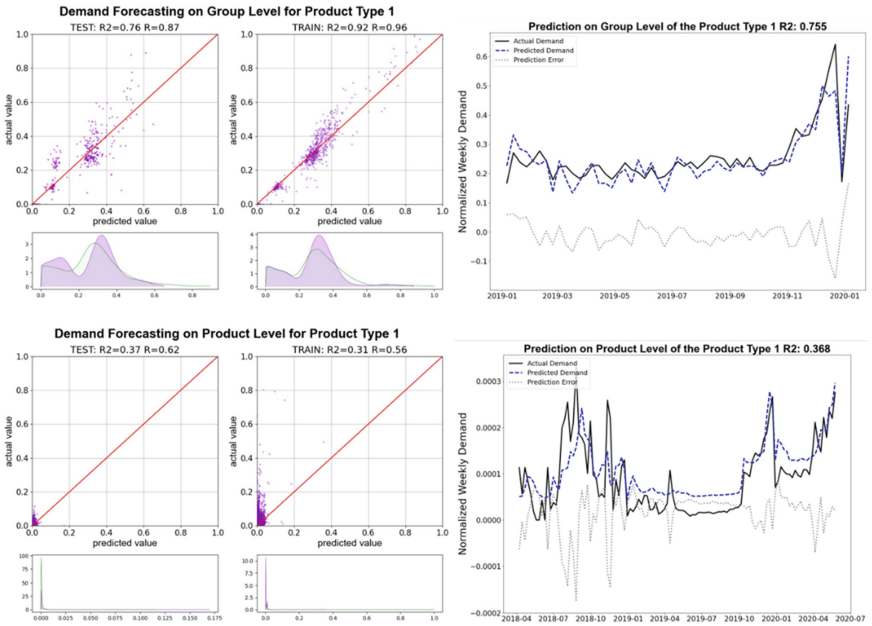


Fig. 2. Demand forecasting of the product type group *Fiction* (1).

reveals that statistical indices of the medium-term demand (four weeks) are identified as highly important features. Only in long-term group level prediction, models scored the cluster affiliation high in feature importance.

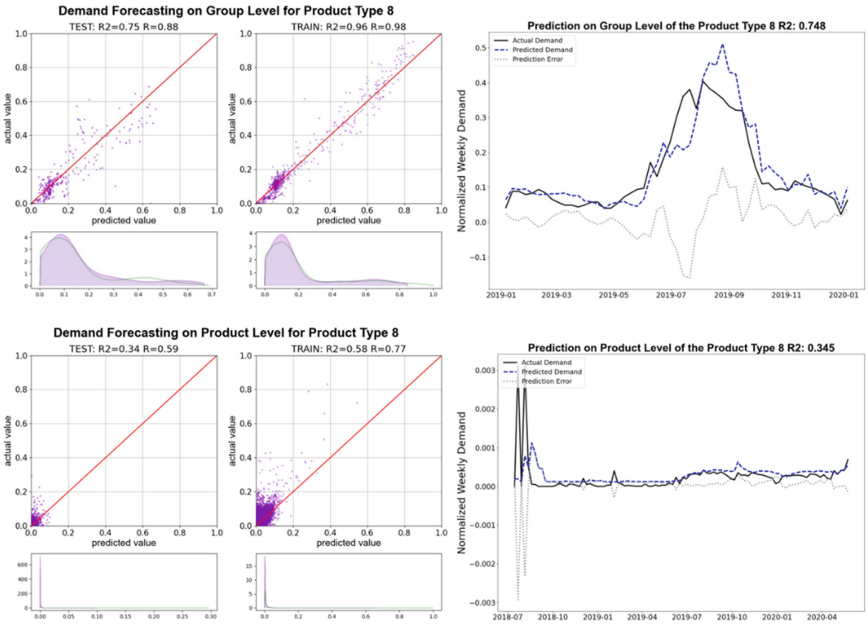


Fig. 3. Demand forecasting of the product type group *School and Learning* (8).

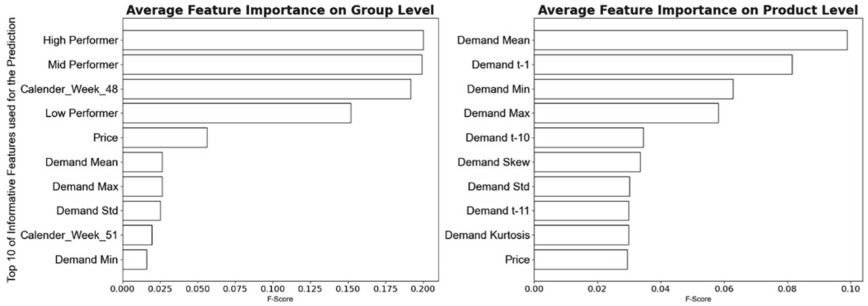


Fig. 4. Top 10 features ranked by their average importance score of the group and product level prediction.

4 Conclusion

Developing and providing accurate demand forecasting models helps intermediary companies to remain competitive and maintain or increase their profit margin by optimizing inventory management and making informed, data-driven decisions. Future research will investigate and compare the potential of autoregression and filtering methods as well as deep learning methods for time series prediction, such as Recurrent Neuronal Networks (RNNs) and Long Short-Term Memories (LSTMs) in particular (e.g., [10, 11, 21]), to improve the demand forecasting on product level. The latter typically requires less sophisticated data preprocessing and feature engineering. Furthermore, combining

different models and prediction levels (i.e., short- and long-term) might additionally improve the performance of the demand forecasting [11, 21].

References

1. Guo, Z.X., Wong, W.K., Li, M.: A multivariate intelligent decision-making model for retail sales forecasting. Decis. Supp. Syst. **55**(1), 247–255 (2013)
2. You, Z., Si, Y.-W., Zhang, D., et al.: A decision-making framework for precision marketing. Expert Syst. Appl. **42**(7), 3357–3367 (2015)
3. Blackburn, R., Lurz, K., Priese, B., et al.: A predictive analytics approach for demand forecasting in the process industry. Int. Trans. Oper. Res. **22**(3), 407–428 (2015)
4. Hofmann, E., Rutschmann, E.: Big data analytics and demand forecasting in supply chains: a conceptual analysis. IJLM **29**(2), 739–766 (2018)
5. Nguyen, T., Zhou, L., Spiegler, V., et al.: Big data analytics in supply chain management: a state-of-the-art literature review. Comput. Oper. Res. **98**, 254–264 (2018)
6. Fildes, R., Ma, S., Kolassa, S.: Retail forecasting: research and practice. Int. J. Forecast. (2019, in press)
7. Seyedan, M., Mafakheri, F.: Predictive big data analytics for supply chain demand forecasting: methods, applications, and research opportunities. J. Big Data **7**(1), 1–22 (2020)
8. Wang, G., Gunasekaran, A., Ngai, E.W., et al.: Big data analytics in logistics and supply chain management: certain investigations for research and applications. Int. J. Prod. Econ. **176**, 98–110 (2016)
9. Makridakis, S., Spiliotis, E., Assimakopoulos, V.: Statistical and machine learning forecasting methods: concerns and ways forward. PLoS One **13**(3), e0194889 (2018)
10. Huber, J., Stuckenschmidt, H.: Daily retail demand forecasting using machine learning with emphasis on calendric special days. Int. J. Forecast. **36**(4), 1420–1438 (2020)
11. Punia, S., Nikolopoulos, K., Singh, S.P., et al.: Deep learning with long short-term memory networks and random forests for demand forecasting in multi-channel retail. Int. J. Prod. Res. **58**(16), 4964–4979 (2020)
12. Vairagade, N., Logofatu, D., Leon, F., et al.: Demand forecasting using random forest and artificial neural network for supply chain management. In: Nguyen, N.T., Chbeir, R., Exposito, E., et al. (eds.) Computational Collective Intelligence, pp. 328–339. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-28377-3_27
13. Boylan, J.E., Chen, H., Mohammadipour, M., et al.: Formation of seasonal groups and application of seasonal indices. J. Oper. Res. Soc. **65**(2), 227–241 (2014)
14. Chen, I.-F., Lu, C.-J.: Sales forecasting by combining clustering and machine-learning techniques for computer retailing. Neural Comput. Appl. **28**(9), 2633–2647 (2016). https://doi.org/10.1007/s00521-016-2215-x
15. Thomassey, S., Fiordaliso, A.: A hybrid sales forecasting system based on clustering and decision trees. Decis. Supp. Syst. **42**(1), 408–421 (2006)
16. Chen, T., Guestrin, C.: XGBoost. In: Krishnapuram, B., Shah, M., Smola, A., et al. (eds.) Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 785–794. ACM, New York (2016)
17. Pedregosa, F., Varoquaux, G., Gramfort, A., et al.: Scikit-learn: machine learning in python. J. Mach. Learn. Res. **12**, 2825–2830 (2011)
18. Nwadiugwu, M.C.: Gene-based clustering algorithms: comparison between denclue, fuzzy-c, and BIRCH. Bioinform. Biol. Insights **14**, 1–6 (2020)
19. Tavenard, R., Faouzi, J., Vandewiele, G., et al.: Tslearn, a machine learning toolkit for time series data. J. Mach. Learn. Res. **21**(118), 1–6 (2020)

20. https://xgboost.readthedocs.io/en/latest/parameter.html. Accessed 10 Mar 2021
21. Kilimci, Z.H., Akyuz, A.O., Uysal, M., et al.: An improved demand forecasting model using deep learning approach and proposed decision integration strategy for supply chain. Complexity **2019**, 1–15 (2019)