Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 180 (2021) 40–49

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Industry 4.0 and Smart Manufacturing

Machine Learning and Statistics: A Study for assessing innovative Demand Forecasting Models

Nikolas Ulrich Moroff* [(a)], Ersin Kurt[(b)], Josef Kamphues[(c)]

[a,b,c] *Fraunhofer Institute for Material Flow and Logistics, Joseph-von-Fraunhofer-Str. 2-4, 44227 Dortmund, Germany*

Abstract

Besides increasing dynamics in market demands, companies strive to avoid short-term changes in their supply chain planning. Therefore, an essential lever to improve supply chain performance is the optimization of the demand forecast. In this regard, artificial intelligence is a widely adopted technique in Industry 4.0 that is associated with high expectations. Against this background, six different forecasting models from statistics and machine learning were evaluated in respect to forecast quality and effort for implementation. The results underline the potential of innovative forecasting models as well as the necessity for an intensive and application-specific evaluation of the advantages and disadvantages of the available approaches.

Keywords: Demand Forecast; Machine Learning; Statistical Methods; Deep Learning

1. Motivation

Services and products offered by companies have become mostly interchangeable. In order to stand out from competitors companies focus on flexible customer service, speed and adherence to delivery dates at reasonable prices [1]. Shortened product life cycles, fluctuating customer behavior and the necessity to react immediately to market fluctuations are just some of the challenges in this regard. In order to reduce short-term changes in the supply chain it is crucial to implement efficient forecasting models that allow companies to prepare for future situations in advance [2]. The available forecast algorithms in literature as well as in commercial (ERP) systems are constantly increasing in terms of quantity and complexity. In addition, computing power and storage capacities have become significantly

* Corresponding author. Tel.:+49 (0) 231-9743-290; fax: +49 (0) 231-9743-77-290
E-mail address: Nikolas.Moroff@iml.fraunhofer.de

1877-0509
10.1016/j.procs.2021.01.127

less expensive, which opens up new possibilities in forecasting for companies [3]. However, both quantitative and qualitative forecasting models are in some cases not suitable for producing a forecast of sufficient quality due to high market dynamics. Among other things, the increasing volume of data and an increasingly complex business environment are promoting uncertainties in the area of forecasting.

Uncertainties are partly sales-related and often the result of a lack of knowledge or incorrect information. Further, frequently changing product ranges and varying demand due to different interfering influencing factors (seasonal influences, price and assortment policies etc.) are only a few of many factors that make it difficult to use conventional statistical forecasting models. In addition, the actual correlation between the influencing factors is often difficult to grasp or not recognizable by humans, which leads to an increased desire for and rising usage of artificial intelligence methods in demand planning [4].

In order to obtain a precise forecast despite high market requirements, methods of machine learning (ML) and the subcategory deep learning (DL) are increasingly used [4, 5]. Both method types can be defined as sub-areas of artificial intelligence. The advantage of AI-based methods is an automatic analysis of patterns and dependencies in the input data, in order to use them for the subsequent forecast. As is the case with statistical methods, there is no generally valid AI-based method that generates an improved forecast for every situation. Rather, each method can be used to achieve different qualities depending on the application. Therefore, this article examines to what extent innovative AI-based methods offer added value, considering forecast quality, implementation effort and computing power.

Since the actual effect of machine learning methods strongly depends on the framework of the data set [6], the methods within this publication will be tested and validated using standardized demand patterns. In order to ensure comparability, all methods are only used as univariate methods. Thus, the only source of information is the historical demand of the respective product. The aim of the investigation is to elaborate whether innovative AI-based methods can lead to an improvement in forecast quality due to their extended possibilities of pattern recognition. The possibilities of integrating further influencing factors (e.g. information regarding marketing activities or other factors that may affect the customer behavior) are not considered in this paper. Both statistical and AI-based methods will be implemented and finally compared with the focus on forecast quality and implementation effort.

The structure of the paper is as follows: The following section deals with basic features of demand forecasting and available forecasting models (chapter 2). Afterwards, the procedure for evaluating different forecasting models is presented (chapter 3). The results are presented in the concluding chapter 4. The paper ends with an outlook on current challenges and next steps.

2. State of the Art – Demand Forecasting

2.1. Introduction Demand Forecasting

In order to be able to classify the individual forecasting models, there are different categories that allow a classification according to logic and application [7]. Thus, forecasting models can be differentiated according to their methodological approach (qualitative and quantitative) and the forecast period (short, medium or long-term) [8]. The qualitative methods often have little or no quantitative information on the object of the forecast [9], so that expert knowledge and verbal statements are used (e.g. estimation of sales staff or customer surveys). Reasons for the application of qualitative methods are for example missing information from the past (e.g. when a new product is launched) or structural discontinuities (e.g. new technologies or changed political conditions).

This investigation implies that sufficient information is available to derive a forecast mathematically. Therefore, the paper focuses on quantitative forecasting models. In contrast to qualitative models, these models are based on mathematical rules and principles that calculate quantitative forecast results based on numerical historical values [9, 10]. In this regard, the fundamental assumption is that the demand of the past can be used to determine the demand of the future. The demand of the past is described by a time series, representing a chronological sequence of recorded observations. Considering the characteristic features of the time series, this sequence is projected into the future using appropriate forecasting models. The objective of the quantitative forecast can thus be defined as the updating of an existing time series [11, 12].

In quantitative forecasting, methods from statistics, machine learning and deep learning are used [12, 13, 14]. Usually the models are used for short and medium-term forecasting. Further, it is necessary to differentiate between univariate and multivariate forecasting models [15, 16]. With the latter, additional time-dependent, influencing factors (such as weather or economic data) and interrelations between different time series can be considered [11].

When analyzing a time series, Wold's decomposition theory is used, which states that each time series is composed of four components (trend component T, seasonal component S, cyclical component C and the residual component I) [5, 16]. In reality, an actual demand pattern often results from the interference of these different components. The combination of the individual components can be additive or multiplicative, so that different demand patterns are created based on the same component characteristics [16]. The following Figure 1 shows a short example of how trend and seasonal components can be combined with each other and the respective demand patterns.

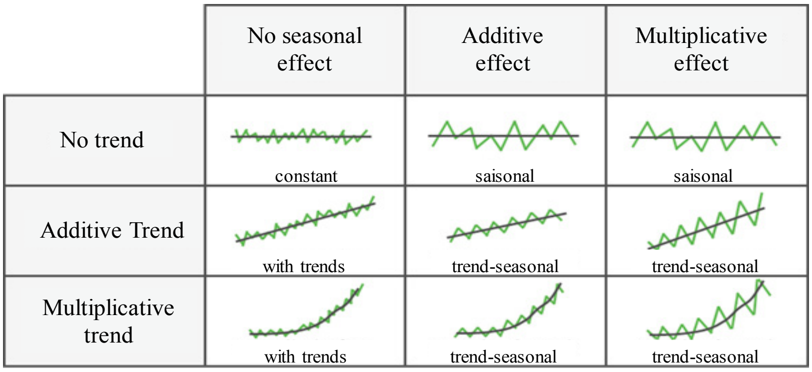


Fig. 1. Demand Patterns resulting from Trend and seasonal components [15].

Reflecting these different demand profiles, many quantitative models have been developed in the past. In order to allow a manageable comparison of different approaches, this paper focuses on a total number of six prevalent models in theory and practice. With all selected forecasting models, both seasonal and trend effects can be considered. In order to compare the advantages and disadvantages of individual models in relation to the representative demand patterns, two models each from the field of statistics, machine learning and deep learning were examined in regard to forecasting errors. The selected models are briefly described in the following.

2.2. Description of the chosen forecasting models

<u>Holt Winters - Triple exponential smoothing (ETS):</u> Holt Winters is a time series analysis method for short-term predictions from a sample of periodic historical data. Exponential smoothing gives these data a higher weighting as they become more current, so that more recent values have a greater influence on the forecast than the history. By setting parameters, factors such as seasonality and trends can be taken into account [17, 18].

<u>Seasonal Auto-Regressive Integrated Moving Average Extended (SARIMAX):</u> ARIMA models in general form a powerful model class for short-term prediction of time series used in many industrial applications. They are based on three components: an autoregressive part, a contribution from a moving average, and a part comprising the first derivative of the time series [17, 18]. The extended SARIMAX method additionally takes seasonal effects and exogenous variables (parallel time series variates) into account for preparing the actual forecast [19].

<u>Extreme Gradient Boosting (XGBoost):</u> XGBoost (extreme gradient boosting) is a machine learning technique for regression and classification problems that generates a prediction model in the form of a decision tree. It builds up the model step-by-step (general boosting concept) by allowing the optimization of any differentiable error function [20, 21].

Random Forest (RF): Random Forest is a machine learning process that consists of several uncorrelated decision trees. All decision trees have grown under a certain type of randomization during the learning process. For a classification, each node in the structure is allowed to make a decision and the class with the most votes decide the final classification (majority principle). Random Forests can also be used for regression analysis and problems [22].

Long-term short-term memory (LSTM): LSTM is an artificially recurrent neural network architecture (RNN) used in deep learning. In contrast to forward-directed neural networks, LSTM has feedback connections. It can process not only individual data points (e.g. images) but also entire data sequences and is therefore well suited for demand forecasting [23]

Multilayer Perceptron (MLP): A multilayer perceptron is an artificial neural network with forward coupling, characterized by multiple layers of neurons connected as a directed graph between the input and output layers. Each node, except the input nodes, has a nonlinear activation function. Since there are multiple layers of neurons, MLP is a technique for deep learning [24].

After the description of the methods, they are classified in the following Table 1 under Statistics, Machine Learning and the more advanced subcategory Deep Learning.

Table 1. Considered Forecasting Methods in Statistics, ML, and DL.

Statistics Methods	Machine Learning Methods	
		Deep Learning Methods
ETS: Holt Winters - Triple exponential smoothing	**XGBoost**: Extreme Gradient Boosting	**LSTM**: Long-term short-term memory
SARIMAX: Seasonal Auto-Regressive Integrated Moving Average Extended	**RF**: Random Forest	**MLP**: Multilayer Perceptron

3. Implementation of forecasting models

3.1. Selection of representative products

For evaluating the selected forecasting models, five typical demand patterns with interfering additive and multiplicative effects were selected. In the following Figure, the cases considered in this paper are highlighted.

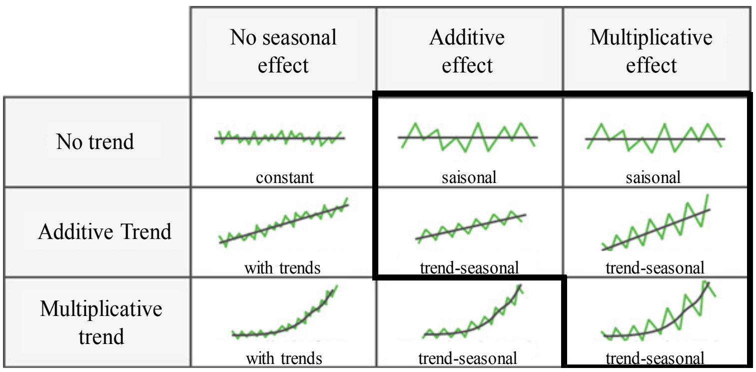
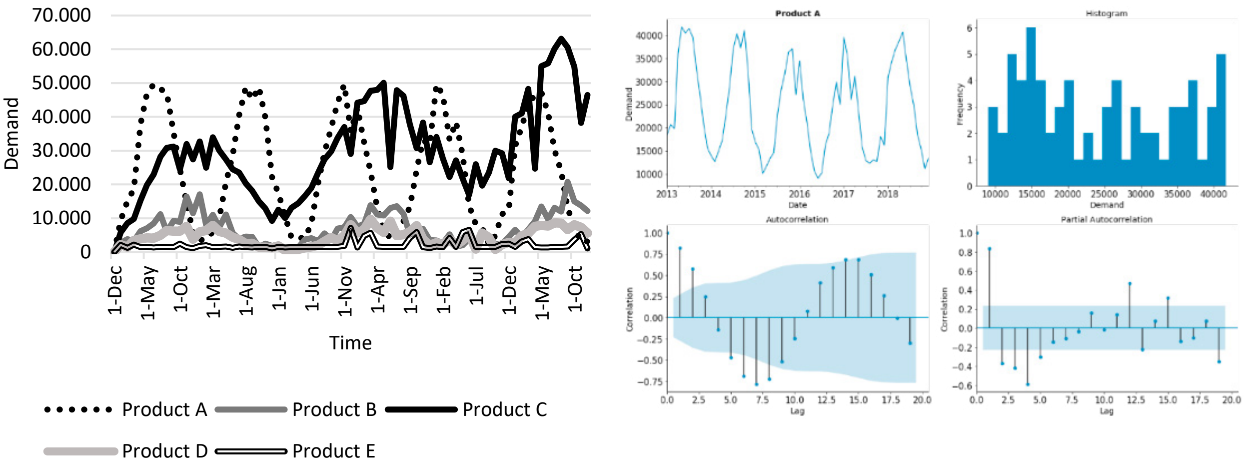


Fig. 2. Selected Demand Patterns resulting from Trend and Seasonality [15].

The five demand patterns were represented by 5 products. A demand history of six years was available on a daily basis. All demand values were aggregated monthly (see Figure 3). For all products only the time was considered as a variable - other possible influencing factors were not examined in this analysis. Thus, a data set of 72 demand values for each of the examined products was used.

As a necessary part of the data preparation and analysis, the frequency of occurrence, the autocorrelation (similarity of a time series, with a time series shifted by x months) and partial autocorrelation (similarity between the current and the previous value) of the time series were examined. Based on these analyses, the historical time series could be assigned to the corresponding demand patterns. The results of the data analysis for product A (no trend but additive seasonality) are shown in Figure 3. The autocorrelation clearly shows that it is a seasonal trend. The correlation becomes minimal with a time shift (lag) of 12 months. After the description of the data basis, the following section briefly describes the system architecture for calculating the forecast.



	Product A	Product B	Product C	Product D	Product E
Maximum	2074	1497	2229	779	797
Minimum	0	0	0	0	0
Average	804,2	234,3	991,2	150,0	70,9
Standard Deviation	606,3	363,4	532,7	173,4	97,5
Trend	none	additiv	multiplikativ	additiv	none
Seasonality	additiv	additiv	multiplikativ	multiplikativ	multiplikativ
Cycle [month]	14	27	27	29	8

Fig. 3. (left) Time series of the chosen products; (right) Exemplary Results of the Time Series Analysis for product A; (table) Overview of the key facts of the Time Series Analysis for all products

3.2. System architecture

The forecasting models were implemented with Python in a Visual Code development environment. The following Figure 4 briefly summarizes the implemented system architecture hierarchically. As a framework for accessing the program libraries of the different forecasting models Tensorflow was used. For the data preparation and subsequent visualization of the results, Pandas and Matplotlib as well as Seaborn libraries was used.

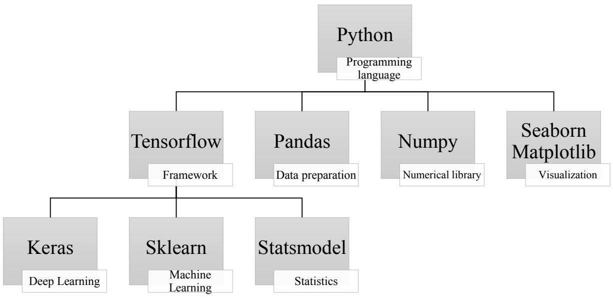


Fig. 4. System Architecture.

The individual forecasting models use different program packages from the libraries, which are listed briefly in the following Table 2, sorted by forecasting model. Since a good parameter setting is a prerequisite for comparing different forecasting models with each other, the following section will discuss the parameter adjustment of the individual models.

Table 2. Program Packages of the Forecasting Models.

Method	Packages
ETS	from statsmodels.tsa.holtwinters import ExponentialSmoothing
SARIMAX	from statsmodels.tsa.statespace.sarimax import SARIMAX
XGBoost	from xgboost import plot_importance, plot_tree
RF	from treeinterpreter import treeinterpreter from sklearn.ensemble import RandomForestRegressor
LSTM	from sklearn.preprocessing import MinMaxScaler from keras.models import Sequential from keras.layers import Dense from keras.layers import LSTM from keras.layers import Dropout
MLP	from keras.models import Sequential from keras.layers import Dense

3.3. Parameter settings

In order to be able to compare the individual forecasting models with each other, each model is implemented with optimal parameters. The statistical models have been analyzed with focus on trend and seasonality in order to optimally adjust the forecasting models. The ML and DL models were tested using the respective standard configuration of the above-mentioned libraries with only few exceptions. The exact parameter settings of the models are shown in Table 3.

Table 3. Parameter setting.

	Product A	Product B	Product C	Product D	Product E
SARIMAX(p,d,q)x(P,D,Q,m)	(2,1,2)x(0,1,1,12)	(2,1,1)x(1,1,0,34)	(3,1,0)x(0,1,0,24)	(0,1,1)x(0,1,1,29)	(1,0,0)x(1,1,0,8)
ETS	α = 0.052	α = 0.053	α = 0.21	α = 0.47	α = 0.103
	β = 0.052	β = 0.053	β = 0.052	β = 0	β = 0
	γ = 0.631	γ = 0.95	γ = 0.79	γ = 0.53	γ = 0.537
	seasonal_period = 14	seasonal_period = 27	seasonal_period = 27	seasonal_period = 29	seasonal_period = 8
RF	Configuration: (bootstrap=True, criterion='mse', max_depth=2, max_leaf_nodes=None, min_samples_leaf=1, min_samples_split=2, n_estimators=100, random_state=0, warm_start=False)				
XGBoost	Configuration: (learning method = Information gain, n_estimators =1000, learning rate = 0.03, max_tree depth = 2, paralysing threads = 4, training ratio = 0.7)				
LSTM	Configuration: [n_input, n_nodes, n_epochs, n_batch]				
	[14,500,100,100]	[27,500,100,100]	[27,500,100,100]	[25,500,100,100]	[12,500,100,100]
MLP	Configuration: [n_input, n_nodes, n_epochs, n_batch]				
	[14,500,100,100]	[27,500,100,100]	[27,500,100,100]	[25,500,100,100]	[12,500,100,100]

4. Results

An initial assessment of the six different forecasting models was made on the basis of the calculated one-step forecasts. Therefore, the forecast errors of five out of ten products were compared with each other. As expected, the calculations show that all models lead to different forecast qualities for the products examined. The calculated error values (root-mean-square deviation, RMSE) are presented in Table 4. Highlighted values show the lowest error value for each product. The lowest error values could be achieved with the models ETS, XGBoost, LSTM and MLP.

Table 4. Forecasting Errors (RMSE) of Forecasting Models by Product.

Prod.	Stat. Models		ML Models		DL Models	
	SARIMAX	ETS	RF	XGBoost	LSTM	MLP
A	4.341	6.786	10.817	9.706	3.932	6.172
B	8.109	3.219	8.777	8.712	4.658	3.656
C	14.237	13.513	25.477	25.978	24.539	11.591
D	1.909	1.879	3.902	3.983	2.633	2.313
E	1.010	1.208	1.439	729	1.034	954

For a better comparability of the results, a performance score was determined, which represents the relative deviation of a forecast error from the best model of the respective product under investigation (see Table 5).

Table 5. Performance scores of Forecasting Models.

Prod.	Stat. Models		ML Models		DL Models	
	SARIMAX	ETS	RF	XGBoost	LSTM	MLP
A	91%	58%	36%	41%	100%	64%
B	40%	100%	37%	37%	69%	88%
C	81%	85%	45%	45%	47%	100%
D	98%	100%	48%	47%	71%	81%
E	72%	60%	51%	100%	70%	76%
Ø	76,4%	80,8%	43,5%	53,8%	71,6%	81,9%

The mean value of this performance score across all examined products allows a statement about the performance of the respective process for different product samples: The higher the overall performance score, the more suitable the model is for a wider range of products. In this investigation, the deep learning model MLP delivered the highest overall score of 81.9 %. The second-best value (80.8 %) was achieved by the triple exponential smoothing model from Holt and Winters (Table 5).

For the evaluation of the forecasting models, also the effort for parameterization and calculation of the models must be considered. In this regard the ML-based models RF and XGBoost proved to be particularly low in effort. While the additional effort for the statistical methods results from a more complex data preparation and parameterization, the DL models proved to be more effortful due to a high computational effort. It is therefore always crucial to consider the application case when selecting the forecast model in order to effectively use AI methods in the Industry 4.0 environment. In our application case we cannot recommend the low-cost RF/XGBoost methods for our products, as they achieve too low forecast quality. The deep learning procedure MLP can be recommended to a company equipped with the appropriate software/hardware to perform the high computational effort, because it has the highest forecast quality and could be identified as the most robust procedure. It is therefore suitable for the largest quantities of different products. The conventional ETS method can be suggested to companies that do not have the computing power and expertise in AI but still want to achieve good quality forecast results.

5. Conclusion

In order to compare forecasting models with each other, two factors must be considered: On the one hand, the models used should achieve the lowest possible forecast errors. On the other hand, the implementation effort of the methods is relevant with regard to their applicability in practice.

The investigation carried out shows that the forecasting models have different forecast qualities depending on the underlying demand patterns. Although the DL method MLP showed the best overall performance in the study, it became clear that conventional statistical methods (in this case: ETS) are still justified for certain demand patterns. Overall, it could not be proven that ML or DL methods by itself lead to better forecast results. Rather, it became clear that a product-specific analysis and an ex-post evaluation is of great importance in order to be able to identify the most suitable forecasting model.

With regard to the implementation effort, the ML methods RF and XGBoost proved to be particularly low in effort. While the additional effort for the statistical models results from a more complex data preparation, the DL methods proved to be more effortful due to the increased computational effort.

In order to evaluate the potential of ML and DL methods in more detail, further investigations are necessary. It can be assumed that these methods can achieve lower forecasting errors, especially if other parameters that correlate with the demand are included in the analysis (e.g. information about the weather, key figures of economic development or marketing activities). Against this background, an expansion of the calculations presented here to include multivariate methods seems to be expedient for elaborating the full potential of AI-based forecasting methods.

References

[1] Wassermann, Otto (2013) “Das intelligente Unternehmen: Mit der Wassermann Supply Chain Idee den globalen Wettbewerb gewinnen.“ Berlin, Heidelberg, Springer-Verlag.

[2] Erben, Roland Franz, and Romeike, Frank (2003) “Komplexität als Ursache steigender Risiken in Industrie und Handel.“ in: *Erfolgsfaktor Risiko-Management.* Berlin, Heidelberg, Springer-Verlag: 43–63.

[3] Welsch, A., Eitle, V., & Buxmann, P. (2018). “Maschinelles Lernen.“ in: *HMD Praxis der Wirtschaftsinformatik*, *55*(2), 366-382.

[4] Döbel, Inga, Leis, Miriam, Vogelsang, Manuel Molina, Neustroev, Dmitry, Petzka, Henning, and Riemer, Annamaria et al. (2018) “Maschinelles Lernen. Eine Analyse zu Kompetenzen, Forschung und Anwendung.“ München, Fraunhofer-Gesellschaft.

[5] Chase, Charles W. (2016) “Next generation demand management: People, process, analytics, and technology.“ New York, John Wiley & Sons.

[6] Makridakis, Spyros, Spiliotis, Evangelos, and Assimakopoulos, Vassilios (2018) “Statistical and Machine Learning forecasting methods: Concerns and ways forward.” in *PLOS ONE* 13 (3).

[7] Hansmann, Karl-Werner (1983) “Kurzlehrbuch Prognoseverfahren: Mit Aufgaben und Lösungen.“ Wiesbaden, Gabler Verlag.

[8] Meffert, Heribert, Burmann, Christoph, Kirchgeorg, Manfred, and Eisenbeiß, Maik (2018) “Marketing: Grundlagen marktorientierter Unternehmensführung Konzepte-Instrumente-Praxisbeispiele.“ Berlin, Heidelberg, Springer-Verlag.

[9] Makridakis, Spyros G., Wheelwright, Steven C., and Hyndman, Rob J. (1998) “Forecasting. Methods and applications.” New York, John Wiley & Sons.

[10] Koch, Jörg (2004) „Marktforschung: Begriffe und Methoden.“ München, Oldenbourg.

[11] Box, George E. P., Jenkins, Gwilym M., Reinsel, Gregory C., and Ljung, Greta M. (2015) “Time series analysis: forecasting and control.” New York, John Wiley & Sons.

[12] Schröder, Michael (2012) “Einführung in die kurzfristige Zeitreihenprognose und Vergleich der einzelnen Verfahren.“ in Peter Mertens, Susanne Rässler (eds) *Prognoserechnung*, Berlin, Heidelberg, Physica-Verlag, 11–45.

[13] Crone, Sven F. (2010) „Neuronale Netze zur Prognose und Disposition im Handel.“ Wiesbaden, Gabler Verlag.

[14] Herrmann, Frank (2011) „Operative Planung in IT-Systemen für die Produktionsplanung und -steuerung: Wirkung, Auswahl und Einstellhinweise von Verfahren und Parametern.“ Wiesbaden, Vieweg+Teubner Verlag / Springer Fachmedien.

[15] Scheer, August-Wilhelm (1983) “Absatzprognosen.“ Berlin, Heidelberg, Springer-Verlag.

[16] Treyer, Oscar A. G. (2010) “Business Forecasting: Anwendungsorientierte Theorie quantitativer Prognoseverfahren.“ Bern, Haupt.

[17] Makatjane, Katleho, and Moroke, Ntebogang (2016) “Comparative study of holt-winters triple exponential smoothing and seasonal arima: forecasting short term seasonal car sales in South Africa.” in *Risk governance & control: financial markets & institutions* 6 (1): 71–82.

[18] Omane-Adjepong, Maurice, Oduro, Francis T., and Oduro, Samuel Dua (2013) “Determining the better approach for short-term forecasting of ghana’s inflation: Seasonal ARIMA Vs holt-winters.” in *International Journal of Business, Humanities and Technology* 3 (1): 69–79.

[19] Papaioannou, George P., Dikaiakos, Christos, Dramountanis, Anargyros, and Papaioannou, Panagiotis G. (2016) “Analysis and modeling for short-to medium-term load forecasting using a hybrid manifold learning principal component model and comparison with classical statistical models (SARIMAX, Exponential Smoothing) and artificial intelligence models (ANN, SVM): The case of Greek electricity market.” in *Energies* 9 (8): 635.

[20] Chen, Tianqi, and Guestrin, Carlos (2016) “Xgboost: A scalable tree boosting system.” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*: 785–794.

[21] Gumus, Mesut, and Kiran, Mustafa S. (2017) “Crude oil price forecasting using XGBoost.” in *2017 International Conference on Computer Science and Engineering (UBMK)*. IEEE: 1100–1103.

[22] Lauer, Tim, Legner, Sarah, and Henke, Michael (2019) “Application of machine learning on plan instability in master production planning of a semiconductor supply chain.” in *IFAC-PapersOnLine* 52 (13): 1248–1253.

[23] Sak, H., Senior, A. W., & Beaufays, F. (2014). Long short-term memory recurrent neural network architectures for large scale acoustic modeling.

[24] Hamedmoghadam, H., Joorabloo, N., & Jalili, M. (2018). “Australia's long-term electricity demand forecasting using deep neural networks.” arXiv preprint arXiv:1801.02148.

AUTHORS BIOGRAPHY

Nikolas Ulrich Moroff (M.Sc.) graduated in 2018 with a master's degree in mechanical engineering from the Ruhr University of Bochum. Since 2 years he is working at the Fraunhofer Institute for Material Flow and Logistics IML and has focused on the research areas artificial intelligence in demand forecasting and material flow simulations in logistics.

Ersin Kurt (M.Sc.) graduated in 2020 with a master's degree in logistics with a focus on IT and data science from Technical University of Dortmund. During his time at the Fraunhofer Institute for Material Flow and Logistics, he was especially involved in the technical implementation of innovative algorithms in the field of demand forecasting.

Josef Kamphues (Dipl.-Ing.) graduated in 2012 with a diploma in mechanical engineering from the Leibniz University of Hannover. From 2012 to 2018 he worked as a research assistant in the Supply Chain Engineering department at the Fraunhofer Institute for Material Flow and Logistics. Since 2018 he is team leader of the research group SCM at Fraunhofer IML. His work focuses on design and planning of supply chains and logistics processes as well as simulation of procurement, production and distribution processes.