Applying Predictive Analytics Algorithms to Support Sales Volume Forecasting

Jörg H. Mayer[1]([✉]), Milena Meinecke[1], Reiner Quick[1], Frank Kusterer[2], and Patrick Kessler[3]

[1] Darmstadt University of Technology, Hochschulstrasse 1, 64289 Darmstadt, Germany
jhmayer@t-online.de
[2] UE University of Europe for Applied Sciences, Reiterweg 26B, 58636 Iserlohn, Germany
[3] WILO SE, Wilopark 1, 44263 Dortmund, Germany

Abstract. Caompanies struggle with predictive analytics (PA), which aims to be a "modern" crystal ball. But how does one choose the "right" algorithms? Based on the findings from a sales volume forecasting case study, this article presents six design guidelines on how to apply PA algorithms properly: (1) When fixing the objective of your forecast, start with reflecting the available data. (2) Considering the available data and forecast horizon, develop a strategy for the training phase, ultimately the model's deployment. (3) Choose algorithms first that act as an orientation as well as a benchmark for more elaborated models. (4) Continue with time series algorithms such as (S)ARIMA and Holt-Winters. Take automated parameter setting into consideration. (5) Integrate additional input by applying ML-based algorithms such as LASSO Regression. (6) Besides accuracy, process efficiency and transparency determine the most suitable approaches.

Keywords: Forecasting · Predictive Analytics · Artificial Intelligence · Machine Learning · Design Science Research in Information Systems · Design Guidelines · Case Study: Manufacturing Industry

1 Introduction

Companies that rely solely on human judgment or simply extrapolate historical values neglect predictive analytics (PA) to improve their forecasting. Although studies have revealed that machine-learning-(ML)-based algorithms can increase the value of big data, practice falls short on applying such algorithms [1]. For example, we found only few case studies such as Wang [2], Gerritsen et al. [3], Sagaert et al. [4], and Concalves et al. [5] addressing the integration of external variables such as leading indicators into PA sales forecasts. We argue that practitioners often lack guidance when applying predictive analytics on big data.

Accordingly, the objective of this article is to present both findings from a sales volume forecasting for a reference company and – derived from these findings – associated design guidelines on how to apply PA algorithms properly. We take an international pump system provider as our reference aiming to increase forecast accuracy, whilst

also improving process efficiency and transparency compared to its current approach. Efficiency covers the reduction of manual actions whilst generating comparable outcomes. Transparency refers to the reproducibility and traceability of actions. We pose two research questions (RQ):

1. What constitutes design guidelines on how to apply PA algorithms?
2. Do the proposed design guidelines fulfill what are they meant to do (validity) and are they useful beyond the environment in which they were developed (utility and generalizability)?

To create things that serve human purposes [6], ultimately to build a better world [7], we follow Design Science Research (DSR) in IS [8, 9] for which the publication scheme from Gregor and Hevner [10] gave us direction. The lack of experience in applying PA algorithms for sales (volume) forecasting motivates this article *(introduction)*. Highlighting several research gaps, we contextualize our research questions *(literature review)*. Addressing these gaps, we adopt a single case study *(method)*. Emphasizing a staged research process with iterative "build" and "evaluate" activities [10], we come up with design guidelines *(artifact description,* related to RQ 1). Testing their validity and utility [11], we perform expert interviews within and beyond the reference company *(evaluation,* related to RQ 2). Comparing our results with prior work and examining how they relate to the article's objective and research questions, we close with a summary, limitations of our work, and avenues for future research *(discussion and conclusion)*[12].

2 Literature Review

Following Webster and Watson [13] as well as vom Brocke [14, 15], we identified relevant articles in a four-stage process. (1) We focused on leading IS journals [16] as well as selected business, management, and accounting journals [17], complemented by proceedings from major IS conferences [18] *(outlet search)*. For a practitioners' perspective, we considered journals such as MIS Quarterly Executive and Harvard Business Review. (2) Accessing the identified outlets, we used ScienceDirect, EBSCOhost, Springer Link, and AIS eLibrary *(database search)*. (3) Then, we searched for articles through their titles, abstracts, and keywords *(keyword search)* – limiting the results to the last ten years.

Integrating this strategy for our research, we combined the keywords "sales (volume)" or "demand" with "planning," "forecasting," or "predicting." To limit our findings to data-driven approaches, we then added the keywords "analytics" or "model." This search resulted in *31 relevant papers* that focus on sales or demand forecasting within a business setting applying data-driven methods. Finally, (4) we conducted a *backward and forward search* identifying another sixteen publications such as Verstraete et al. [19].

For our gap analysis, we structured the publications into three clusters: (1) The *application area* describes the context in which the existing PA models have been developed. Furthermore, we examined (2) *forecasting techniques.* (3) Lastly, we differentiated between *research methods*, addressing the way in which researchers collect and understand data [20]. Table 1 summarizes the take-aways of our literature review.

Table 1. Publications Shaping our Research Approach

Author	Research Gap	Take away for proposed model
(1) Application area: [1–5, 21–23]	Regarding RQ #1: Papers focusing on PA for manufacturing companies and their production planning are underrepresented	We contribute a use case focusing on forecasting for sales volumes at the product group level
(2) Forecasting techniques: [4, 19]	Regarding RQ #1: There is a lack of research on advanced forecasting approaches beyond judgmental and statistical forecasts	We propose to test a series of different forecast approaches including *ML-based algorithms,* to increase accuracy whilst improving process efficiency and transparency
(3) Research method: [22, 24]	Regarding RQ #2: We examined a lack of qualitative research within a company setting	With a *single case study*, we aim to expand the current body of know-ledge, focusing on a company setting, and present design guidelines for future discussions

Application Area. Most of the reviewed publications focused on PA models in the *retail industry*. Predicting online sales, we found articles about leveraging product reviews [25], social media posts, and other online activities via search engines [26]. Authors such as Abolghasemi et al. [27] considered the impact of promotions. To gain customer feedback, *sentiment analyses* were frequently conducted [28].

Focusing on production planning, we found few papers such as Fildes et al. [21] and Chen and Chien [22]. The examined forecasting horizon differed between short-term [23] and strategic planning [4]. With varying aggregations level, Wijnhoven and Plant [29] focus on car sales, Ma and Fildes [30] examined *forecasting at the product level.*

Forecasting Techniques. A growing number of articles examined *ML-based algorithms* such as artificial neural networks and support vector machines. They can model more complex relationships for leveraging big data [31]. *Big data* refers to the growing amount of diverse data being generated by different enterprise resource-planning processes, business intelligence systems, sensors such as smart meters, social media listening, and other sources [32]. However, research lacks deploying ML techniques, especially when adding leading indicators to potentially increase forecast accuracy. This is often done manually so that outcomes are susceptible to bias and lack transparency [33]. Based on this insight, we will apply different statistical and ML models to increase accuracy, whilst improving process efficiency and transparency compared to the judgmental forecast in place (Table 1).

Research Method. Although there are articles on sales volume prediction with a focus on quantitative research [21, 34], they are barely employed in practice [19, 22, 24, 35].

Research neglects qualitative research examining the integration of forecast solutions into day-to-day business. Thus, with a single *case study*, we aim to expand the current body of knowledge.

3   Method

Case studies bridge the gap between practice and academia when few activities on this research topic exist and practical insights are proposed as significant [36, 37]. Compared to surveys, they provide more substantial in-depth information and enable researchers to study their artifacts in a natural setting [38].

Compared to multiple case studies, which yield a multitude of results [39], *single case studies* are more suitable when the research topic is complex, and thus, relevant starting points for research are not easy to obtain [40]. Guided by our research questions (Sect. 1) and the findings from our literature review (Sect. 2), we decided to conduct a *single case study*. We took an international pump system provider (sales: 1.5 billion EUR; 8,000 employees) as our reference company. It produces pumps for various applications such as industrial production and building services and their sales are influenced by seasonal fluctuations and affected by continuous technological progress.

We focused on the sales volume forecasting for two product groups within Germany with a forecast horizon of twelve months. The pumps are installed in heating devices for residential and commercial buildings. Our project team consisted of the authors of this paper and members of the reference company's controlling department. Modeling our PA solution, we followed CRISP-DM [41] and gathered learnings throughout its different phases. We finally deducted design guidelines in an iterative manner by discussing these learnings within our project team and how they might be helpful for implementing PA forecasting solutions in similar use cases.

In order to gather feedback about their utility and validity, we present them in semi-structured expert itnerviews (RQ 2, Sect. 1) which took on average 45 min. These interviews combine both a comparable structure within a series of interviews whilst being flexible when interviewees want to share their individual way of thinking and any hidden facets [42]. We performed these interviews with the Head of Finance and three employees from Controlling and Supply Chain Planning of the reference company. Including an external perspective, we conducted interviews with two data science experts from the working group "Digital Finance" [43].

4   Artifact Design

Emphasizing a staged research process with iterative "build" and "evaluate" activities, we started with the results from our literature review (Sect. 2) and followed CRISP-DM. We gained **(1) a business understanding** by examining the forecasting process of sales volumes currently in place at our reference company. The twelve-month rolling forecast is performed at the sub-group level and later aggregated to the group level. It is conducted as follows: (a) After cleansing the historical sales volume data from exceptional influences, (b) the reference company currently applies a Holt-Winters algorithm

within SAP Advanced Planning and Optimization (SAP APO, R. 4.1). (c) The results are then manually adjusted by local planners considering short-term influences such as sales promotions as well as general economic climate and technological trends.

The cleansing step is skipped, as data relating to exceptional events are often missing. The Holt-Winters parameters had not been altered in recent years. With the different sales regions in focus, we observed differences in the local planner's manual forecasting adjustments. Due to a series of unstructured adjustments, the current procedure is time-consuming and lacks transparency resulting in forecasts that are prone to bias with no monitoring or documentation of achieved accuracies in place.

We gathered possible influences on the sales volumes in seven expert interviews with local demand planners who have been involved in the forecast process and members of the sales department who are knowledgeable about the specific market conditions. We interviewed all available employees until saturation. Based on the qualitative content analysis of our interviews, we compiled a list of possible long- and short-term influences. Their effect on the sales volume was perceived as high by the interviewees, which motivated us to include them in the PA model.

Differentiating between influences and indicators, the latter being specific data that can be processed by the model, we devoted considerable effort to identifying influences on sales volumes and finding leading indicators representing them. We recommend limiting the effort of collecting and analyzing indicators, as their benefit for the forecast accuracy is often questionable as measurable data often do not accurately represent underlying constructs [44]. In our experience, starting the modeling with historical time series data only is a more efficient approach since those data are often readily available. Considering the overall objective of the forecast, budget, expertise, and available data, we recommend to iteratively specifying the scope of the forecast and then gradually collecting additional data when necessary.

> *Design guideline #1: When fixing the objective of your forecast, start with reflecting the available data.*

Within the **(2) data understanding** process, we examined the historical sales volume data in order to apprehend their trend and seasonality. We took nine years of monthly sales volume data without missing data points, but with structural changes caused by product changes. Both product groups are highly seasonal with a decreasing trend (Fig. 1). By decomposing the time series into its seasonal trend and residual component, this becomes apparent (Fig. 1). Furthermore, we analyzed indicators representing the identified influences from the expert interviews (Table 2). Public statistics databases acted as our source for indicators such as economic sentiment indices [45] and building permits [46]. We were unable to gather indicators reflecting some aspects mentioned by the interviewees such as regulatory requirements and technological shifts. Our search resulted in six indicators (Table 2).

During **(3) data preparation**, indicators and historical data were prepared according to the requirements of different forecasting algorithms. Handling seasonality was crucial to the model's success. Removing seasonality helped to detect underlying patterns within time series. Whether seasonality should be removed before modeling depends on various factors like the size of available datasets, the stability of seasonal patterns, and the
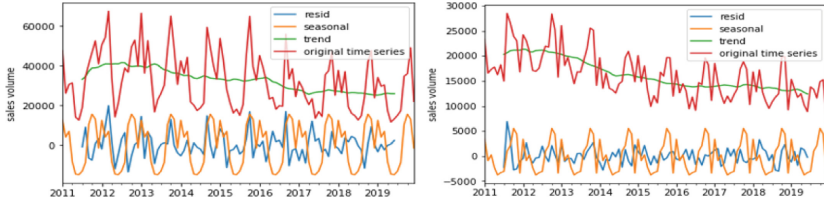
Fig. 1. Decomposition of Product Group 1(left) and 2 (right).

algorithm applied. We prepared two datasets, one with unadjusted data and the other one using seasonal differencing. The latter is subtracting the previous year's observation to remove seasonal fluctuations. In the case of our sentiment indicators, we used the seasonally adjusted data already available.

Table 2. Supposed Influences and Indicators of Sales Volume.

Influences	Leading indicators (not all influences represented)
General economic development	Economic sentiment index (Eurostats)
Sector-specific development	Confidence indicators for sectors industrial, construction,
Regulatory requirements	and retail (Eurostats)
Internal and external competition	Consumer confidence indicator (Eurostats)
	Building permits issued (Statistisches Bundesamt)

Additionally, we standardized the indicators by subtracting their mean and dividing the result by their standard deviation to remove differences in scale, which is necessary for multivariate models. Preparing the data is as important as choosing the right algorithm and having different test datasets is necessary for evaluating the performance of your algorithm accordingly.

Comparing the accuracy of various approaches, the best error measure depends on the individual use case [47]. We used the mean absolute percentage error (*MAPE*, Table 3) because it is easily comprehensible for business users. We performed stepwise *cross-validation*, splitting our dataset into overlapping time periods with three years of training data, one year representing the forecast horizon, and using the consecutive month as the test point. This approach validates the model efficiently while always considering the time restrictions of the data to generate valid results.

After the deployment of the model, the accuracy must be continuously monitored and documented, as this was a major hurdle for the use case, having only data from the past three years from the current forecast.

Design guideline #2: Considering the available data and forecast horizon, develop a strategy for the training phase, ultimately the model's deployment.

For **(4) modeling**, we adopted seasonal naïve, taking the unaltered observation of the previous year as our prediction. Understanding the time series' composition (step 2)

helped to choose a simple algorithm as a starting point and reference for other modeling approaches. Depending on the use case, even simple algorithms such as naïve methods can help companies improve their existing forecasting processes [48]. These models do not require extensive data preparation and therefore are easy to implement.

Design guideline #3: Choose algorithms first that act as both an orientation as well as a benchmark for more elaborated models.

The seasonal naïve approach could significantly lower the MAPE compared to the automated SAP APO forecast and even outperformed the manually adjusted forecast, in doing so, setting the seasonal naïve as the benchmark to beat (Table 3). Moving on to time series algorithms, we applied Seasonal Autoregressive Integrated Moving Average (SARIMA) and Holt-Winters. Besides dealing with potential outliers and missing data, the predefined algorithms do not require additional data preparation. Performing a stepwise forecast, we used the Python packages *pmdarima* and *statsmodels* instead of manual parameter setting, which is often challenging. We recommend opting for popular methods, as the available packages are usually well-maintained, and resources on their application are widely available.

Table 3. Results of several Forecast Algorithms

Method	MAPE (Product group 1)	MAPE (Product group 2)
SAP APO (Holt-Winters)	38.76%	23.61%
Local planner	22.18%	16.67%
Seasonal naïve	21.5%	14.68%
SARIMA	37.26%	27.75%
Holt-Winters (dynamic parameters)	22.05%	16.44%
LASSO linear regression	17.49%	14.96%

Although we could not achieve an improvement compared to seasonal naïve, we improved the accuracy of the automated SAP APO forecast using the Holt-Winters algorithm (Table 3). Using automated parameter setting allows to continuously update the algorithm's parameters, leading to better results.

Design guideline #4: Continue with time series algorithms such as (S)ARIMA and Holt-Winters. Take automated parameter setting into consideration.

Due to the assumption that external factors significantly affect sales volumes, we tried to further improve the forecasting accuracy by integrating *leading indicators*. Performing a correlation analysis, we tried to identify influential variables and their associated time lag. Exogenous variables can easily be added to the SARIMA or Holt-Winters algorithms used in our case study or to regression models. However, the procedure of manually selecting and integrating indicators is complex, especially when dealing with multiple indicators and different time lags.

In our case, we could not determine any unambiguous relation between sales volume and our indicators. Thus, instead of selecting the indicators manually, we used the LASSO algorithm with a linear regression model to select the indicators automatically. The algorithm consists of an additional penalization term that prevents overfitting and selects the best fitting variables. The predefined Python function we used can be found in the scikit-learn library. By adding the six indicators as well as the historical time series data and their respective 12-to-24-month lags, the LASSO algorithm could automatically select the best input variables with the optimal time shift.

For our two product groups, we could only slightly improve the forecast accuracy of the benchmark by including additional indicators with LASSO regression. The indicators chosen by the LASSO algorithm were mostly historical sales volume data, and other indicators were not consistent over the years. This could be because our time series are not significantly influenced by external indicators, or because our selection of indicators might not represent the actual influences on sales volume. Due to the nature and restricted size of our available datasets, we did not include other ML algorithms such as neural networks and vector autoregression.

Design guideline #5: Integrate additional input automatically by applying ML-based algorithms such as LASSO Regression.

(5) Evaluating the results of our models (Table 3), for our time series, we could not significantly improve our long-term forecasts (twelve months) over our simple benchmark (seasonal naïve, PG1: 21.5%, PG2: 14.68%) by adding additional data or using more complex methods. Compared to the automatic forecast currently in place (SAP APO, PG1: 38.76%, PG2: 23.61%), we could achieve a significant improvement in this use case. Even the manually adjusted forecast currently in place did not yield better results for twelve-month forecasts (PG1: 22.18%, PG2: 16.67%). This indicates that the current method is not adequate for our examined products and aggregation level.

From the algorithms that we considered, seasonal naïve yielded the highest accuracy except for PG1, which had better accuracy with the LASSO regression (17.49%). Considering the inconsistent coefficients generated by the algorithm, the improvement does not seem significant. When including other evaluation criteria transparency and process efficiency, the seasonal naïve has the advantage of entailing minimum effort for collecting data and generating the forecast, while being transparent and easy to understand for the recipients of the forecast. Process efficiency can encompass a variety of factors such as computational resources as well as human resources for developing and maintaining the forecast. Especially, when manual reviews and adjustments are included in the process, transparency is an important criterion when selecting a forecasting approach. The acceptance and abilities of the involved employees must be considered.

Therefore, we propose forecasting the sales volumes of our examined product groups by applying seasonal naïve for twelve-month forecasts. By reducing the need for manual adjustments for those product groups, the planners' expertise could be targeted toward more challenging forecasting tasks. Identifying adequate methods by systematically testing different approaches and continuously monitoring them can lead to more accurate forecasts, even when using simple data-driven approaches. Robust and tangible results

are to be preferred over complex and costly approaches, as accuracy is not the only criterion that should be considered.

Design guideline #6: Besides accuracy, process efficiency and transparency determine the most suitable approaches.

For the final CRISP-DM step **(6) deployment** of our prototype, we consider the insights from our use case. The forecast of the product groups can be improved by adapting the current SAP APO forecast to a seasonal naïve approach. Alternatively, the dynamic adaption of Holt-Winters parameters could improve the forecast in comparison to the current static approach. Concerning the manual adjustment by local planners, the consideration of long-term influences such as economic indicators might not improve the forecast and the effort of manually adjusting forecasts could be decreased challenging some of the assumptions about exogenous influences.

In addition to the new forecasting approach, the forecast process must be monitored continuously, updating applied methods when necessary and setting the basis for more advanced algorithms in the future. The technical deployment of such a forecasting and monitoring tool at the reference company is still in an initial stage.

5   Evaluation

Starting with RQ 1, we began our evaluation interviews with a question about the *understandability* of our design guidelines at hand, followed by their potential for implementation. All participants agreed that the design guidelines are easy to understand. The participants were constituted in a *mixed team* of business users and IS experts.

Another aspect that we reflected on intensively is the fact that extensive projects typically aim at implementing advanced forecasting solutions right from the beginning of the project (Sect. 2). We argue for an *iterative approach* including starting with straightforward methods and continuing with more advanced PA models as needed. We learned in our case study that this is especially true when companies lack the necessary data and data science expertise. Simpler solutions can already be beneficial often being more effective than complex ML approaches (Sect. 4).

Following RQ 2 and incorporating the criteria proposed by Gregor and Hevner [10], we structured our evaluation interviews as follows. **Validity:** When asked if the design guidelines would help in applying PA methods, the participants from Controlling and Supply Chain Planning stated: "The systematic approach helps to structure the selection of an appropriate forecasting method with the presented algorithms being a helpful guidance." One of the external experts added that such design guidelines would support a *situational artifact design*, managing "pure" standardization vs "pure" individualization of future forecasting applications.

Compared to the current forecasting of the reference company, we asked the respondents whether the design guidelines could improve the forecast *accuracy*. The answers varied. Half of the interviewees agreed, whereas others were skeptical as the "expectations are very high" and only use cases that are "inadequately maintained" can achieve

significant improvement. The training of existing staff and additional hiring is necessary but difficult, as these profiles are barely available, especially for medium sized companies, stated the Head of Finance.

Regarding process *efficiency*, all participants agreed that the design guidelines can help to reduce manual effort, especially for the *operational controllers*, however their implementation requires adequate infrastructure that allows the convenient testing and monitoring of data models.

Regarding *transparency* the Head of Finance stated that he is not really interested in the "technical" details of different algorithms, but rather in the results of the process. Nonetheless, he was interested in the kind of data influencing the forecast, especially when it comes to *leading indicators*. One of the experts agreed with this answer, in that knowing what data influences is valuable for a greater acceptance of algorithms, whereas understanding the individual algorithm in detail was not deemed necessary.

Summarizing the *validity* of our design guidelines, the Head of Finance said: "For a wider application of PA, statistical knowledge among employees and the variety of products, aggregation levels, and influences within the reference company, all pose a challenge." Therefore, a partially automated application that allows the easy comparison of different algorithms for specific use cases is key to enabling more fact-driven decision-making, whilst reducing human bias and improving efficiency.

Utility and Generalizability: Moving on to the utility and generalizability of our design guidelines, we asked the participants how they could be useful for other product groups and applications beyond the sales-volume domain at our reference company. All interviewees stated that the design guidelines would *support more fact-driven decisions* within the reference company – probably in other companies as well. One of the employees from the Controlling and Supply Chain Planning department explained that the design guidelines are applicable in sales volume forecasting beyond the examined product groups as they provide a general starting point for challenging existing forecast routines.

When it comes to more granular or high-frequency data – for example for detailed warehouse planning – our approach may not fit. When *scaling our model* to establish a cohesive approach across the company, one employee from the Controlling and Supply Chain Planning department emphasized the issue of data availability and inconsistencies between different departments. This impedes the usage of more advanced methods such as ML algorithms. Consequently, providing a solid data architecture is a crucial step for effectively implementing PA.

When asked about the value of our design guidelines beyond *improving forecasting performance*, the Head of Finance mentioned that dealing with PA with a more hands-on approach has helped him to gain a better understanding.

6 Discussion and Conclusion

As their current forecast process requires substantial manual effort and lacks transparency, this article laid out an improved sales volume forecasting for an international pump system provider. The associated design guidelines – derived from these findings – should help to apply PA algorithms properly.

Based on an examined lack of statistical knowledge among the employees of the reference company (Sect. 5) and a large variety of products, aggregation levels, influences, and potential leading indicators, easy-to-handle PA applications were fundamental to enabling business users to drive their forecasting. Thus, for *practice*, our approach should help to get started more effectively with PA – without having extensive data science knowledge.

For *research purposes*, we updated current PA research in a new *industry setting*. Starting with a literature review, we laid out the role of PA in business forecasting. Besides forecast accuracy, we have argued that additional evaluation criteria such as process efficiency and transparency should determine the most suitable approaches as well. Considering the perspective of business users that are hesitant to apply modern forecasting techniques, we detailed the process of applying PA algorithms for practice.

Although business users showed great interest in including leading indicators and thereby making the model "more explainable," the benefit of leading indicators as input for predictive models is questionable. In fact, leading indicators might be better suited to modeling *causal relations* in explanatory models.

Even though ML-based algorithms can outperform simpler algorithms, their limiting factor is often the availability of data. However, our use case showed that even simple models with limited data can improve the forecasting performance.

Newer studies often present specialized solutions neglecting quick win use cases [49]. Our research aims at closing this gap by providing comprehensive design guidelines with a focus on *business users.*

However, our research inevitably reveals certain *limitations.* Although single case studies offer a broad range of advantages, their limited generalizability requires complementary use cases. Furthermore, the lack of data availability prevented us from using more complex methods. Focusing on the most common algorithms kept the project tractable. For future research, we will examine how the software providers keep responding to the PA trend, offering *Auto ML solutions* that aim at enabling business users. The feedback from our interviewees has shown that the interest in understanding the influences on a forecast is high, but the understanding of algorithms itself is not deemed necessary (Sect. 5). This shows that a *black-box approach* does not conform to the expectations of a "modern" PA model. Responding to that, *explainable artificial intelligence* (XAI) is certainly worth being considered in greater detail [50]. Accordingly, a final topic of research could be how the algorithm understandability correlates with its *acceptance* by business users.

References

1. Gilliland, M.: The value added by machine learning approaches in forecasting. Int. J. Forecast. **1**(36), 161–166 (2020)
2. Wang, Ch.-H.: Considering economic indicators and dynamic channel interactions to conduct sales forecasting for retail sectors. Comput. Ind. Eng. **165**, 107965 (2022)
3. Gerritsen, D., Reshadat, V.: Identifying leading indicators for tactical truck parts' sales predictions using LASSO. In: Arai, K. (ed.) IntelliSys 2021. LNNS, vol. 295, pp. 518–535. Springer, Cham (2022). https://doi.org/10.1007/978-3-030-82196-8_38

4. Sagaert, Y.R., Aghezzaf, E.-H., Kourentzes, N., Desmet, B.: Temporal big data for tactical sales forecasting in the tire industry. Interfaces **2**(48), 121–129 (2018)
5. Gonçalves, J.N.C., Cortez, P., Sameiro Carvalho, M., Frazão, N.M.: A multivariate approach for multi-step demand forecasting in assembly industries: empirical evidence from an automotive supply chain. Dec. Support Syst. **142**, 113452 (2021)
6. Simon, H.A.: The Science of the Artificial. MIT Press, Cambridge, Massachusetts (1996)
7. Walls, J.G., Widmeyer, G.R., El Sawy, O.A.: Building an information system design theory for Vigilant EIS. Inf. Syst. Res. **3**(1), 36–59 (1992)
8. Hevner, A.R., March, S.T., Park, J., Ram, S.: Design science in information systems research. MIS Q. **28**(1), 75–105 (2004)
9. Vom Brocke, J., Winter, R., Hevner, A., Maedche, A.: Accumulation and evolution of design knowledge in design science research - a journey through time and space. J. Assoc. Inf. Syst. **21**(3), 520–544 (2020)
10. Gregor, S., Hevner, A.R.: Positioning and presenting design science research for maximum impact. MIS Q. **37**(2), 337–355 (2013)
11. Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S.: A design science research methodology for information systems research. J. Manag. Inf. Syst. **24**(3), 45–77 (2007)
12. Esswein, M., Mayer, J.H., Stoffel, S., Quick, R.: Predictive analytics – a modern crystal ball? answers from a cash flow case study. In: Proceedings of the 27th European Conference on Information Systems, pp. 1–16 (2019)
13. Webster, J., Watson, R.T.: Analyzing the past to prepare for the future: writing a literature review. MIS Q. **2**(26), 13–23 (2002)
14. Vom Brocke, J., Simons, A., Niehaves, B., Riemer, K., Plattfaut, R., Cleven, A.: Reconstructing the giant: on the importance of rigor in documenting the literature search process. In: Newell, S., Whitley, E.A., Pouloudi, N., Wareham, J., Mathiassen, L. (eds.) Proceedings of the 17th European Conference on Information Systems (2009)
15. vom Brocke, J., Simons, A., Riemer, K., Niehaves, B., Plattfaut, R., Cleven, A.: Standing on the shoulders of giants: challenges and recommendations of literature search in information systems research. Commun. Assoc. Inform. Syst. **37**, 205–224 (2015). https://doi.org/10.17705/1CAIS.03709
16. AIS Senior Scholar's Basket of Journals: https://aisnet.org/page/SeniorScholarBasket/. Last accessed 30 Nov 2022
17. Scimago Journal & Country Rank, Business, Management, and Accounting: https://www.scimagojr.com/journalrank.php?area=1400. Last accessed 30 Nov 2022
18. AIS Conferences: https://aisnet.org/page/Conferences/. Last accessed 30 Nov 2022
19. Verstraete, G., Aghezzaf, E.-H., Desmet, B.: A leading macroeconomic indicators' based framework to automatically generate tactical sales forecasts. Comput. Ind. Eng. **139**(1), 1–10 (2020)
20. Myers, M.D.: Qualitative research in information systems. MIS Q. **21**(2), 241 (1997)
21. Fildes, R., Goodwin, P., Lawrence, M., Nikolopoulos, K.: Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. Int. J. Forecast. **25**(1), 3–23 (2009)
22. Chen, Y.-J., Chien, C.-F.: An empirical study of demand forecasting of non-volatile memory for smart production of semiconductor manufacturing. Int. J. Prod. Res. **56**(13), 4629–4643 (2018)
23. Wang, C.-H., Yun, Y.: Demand planning and sales forecasting for motherboard manufacturers considering dynamic interactions of computer products. Comput. Ind. Eng. **149**, 1–8 (2020)
24. Wu, S.D., Kempf, K.G., Atan, M.O., Aytac, B., Shirodkar, S.A., Mishra, A.: Improving new-product forecasting at intel corporation. Interfaces **40**(5), 385–396 (2010)
25. Liu, Y., Feng, J., Liao, X.: When online reviews meet sales volume information: is more or accurate information always better? Inf. Syst. Res. **28**(4), 723–743 (2017)

26. Geva, T., Oestreicher-Singer, G., Efron, N., Shimshoni, Y.: Using forum and search data for sales predictions of high-involvement products. MIS Q. **41**(1), 65–82 (2017)
27. Abolghasemi, M., Hurley, J., Eshragh, A., Fahimnia, B.: Demand forecasting in the presence of systematic events: cases in capturing sales promotions. Int. J. Prod. Econ. **230**, 1–28 (2020)
28. Qiu, J.: A predictive model for customer purchase behavior in e-commerce context. In: Proceeding of the 19th Pacific Asia Conference on Information Systems, p. 369. Chengdu, China (2014)
29. Wijnhoven, F., Plant, O.: Sentiment analysis and google trends data for predicting car sales. In: Proceedings of the 38th International Conference on Information Systems, pp. 1–16 (2017)
30. Ma, S., Fildes, R.: Retail sales forecasting with meta-learning. Eur. J. Oper. Res. **288**(1), 111–128 (2021)
31. Tsoumakas, G.: A survey of machine learning techniques for food sales prediction. Artif. Intell. Rev. **52**(1), 441–447 (2018)
32. Grover, V., Chiang, R.H., Liang, T.-P., Zhang, D.: Creating strategic business value from big data analytics: a research framework. J. Manag. Inf. Syst. **35**(2), 388–423 (2018)
33. Benthaus, J., Skodda, C.: Investigating consumer information search behavior and consumer emotions to improve sales forecasting. In: Proceedings of the 21st Americas Conference on Information Systems, pp. 1–12 (2015)
34. Chong, A.Y.L., Li, B., Ngai, E.W., Ch'ng, E., Lee, F.: Predicting online product sales via online reviews, sentiments, and promotion strategies. Int. J. Oper. Prod. Manag. **36**(4), 358–383 (2016)
35. Blackburn, R., Lurz, K., Priese, B., Göb, R., Darkow, I.-L.: A predictive analytics approach for demand forecasting in the process industry. Intl. Trans. in Op. Res. **22**(3), 407–428 (2015)
36. Flyvbjerg, B.: Case study. In: Denzin, N.K., Lincoln, Y.S. (eds.) The SAGE Handbook of Qualitative Research, pp. 301–316. SAGE, Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne (2018)
37. Benbasat, I., Goldstein, D.K., Mead, M.: The Case Research Strategy in Studies of Information Systems. MIS Q. **11**(3), 369–386 (1987)
38. Dul, J., Hak, T.: Case study Methodology in Business Research. Butterworth-Heinemann, Amsterdam (2007)
39. Yin, R.K.: The case study crisis: some answers. Adm. Sci. Q. **26**(1), 58–65 (1981)
40. Gustafsson, J.: Single case studies vs. multiple case studies: A comparative study (2017)
41. Chapman, P., et al.: CRISP-DM 1.0: Step-by-step data mining guide (2000)
42. Qu, S.Q., Dumay, J.: The qualitative research interview. Qual. Res. Account. Manag. **8**(3), 238–264 (2011)
43. Working Group "Digital Finance" Schmalenbach-Gesellschaft: https://www.schmalenbach.org/index.php/arbeitskreise/finanz-und-rechnungswesen-steuern/digital-finance. Last accessed 29 Nov 2022
44. Shmueli, G.: To explain or to predict? Stat. Sci. **25**(3), 289–310 (2010)
45. Eurostat: Confidence Indicators: https://ec.europa.eu/eurostat/databrowser/view/teibs020/default/table?lang=en (2021). Last accessed 20 Apr 2021
46. Statistisches Bundesamt: Monthly issued building permits for Germany. https://www-genesis.destatis.de/genesis//online?operation=table&code=31111-0002&bypass=true&levelindex=0&levelid=1620400463162#abreadcrumb (2021). Last accessed 20 Apr 2021
47. Kumar, A., Shankar, R., Aljohani, N.R.: A big data driven framework for demand-driven forecasting with effects of marketing-mix variables. Ind. Mark. Manage. **90**, 493–507 (2020)
48. Tibshirani, R.: Regression shrinkage and selection via the lasso. J. Roy. Stat. Soc.: Ser. B (Methodol.) **58**(1), 267–288 (1996)

49. Schröer, C., Kruse, F., Gómez, J.M.: A Systematic literature review on applying CRISP-DM process model. Procedia Comput. Sci. **181**, 526–534 (2021)
50. Barredo Arrieta, A., et al.: Explainable artificial intelligence (XAI): concepts, taxonomies, opportunities and challenges toward responsible AI. Inform. Fusion **58**, 82–115 (2020)