

6-18-2022

## Explanation Interfaces for Sales Forecasting

Tobias Benjamin Fahse

*University of St. Gallen, tobias.fahse@unisg.ch*

Ivo Blohm

*University of St. Gallen, ivo.blohm@unisg.ch*

Richard Hruby

*University of St. Gallen, richard.hruby@student.unisg.ch*

Benjamin van Giffen

*University of St. Gallen, benjamin.vangiffen@unisg.ch*

Follow this and additional works at: [https://aisel.aisnet.org/ecis2022\\_rip](https://aisel.aisnet.org/ecis2022_rip)

---

### Recommended Citation

Fahse, Tobias Benjamin; Blohm, Ivo; Hruby, Richard; and van Giffen, Benjamin, "Explanation Interfaces for Sales Forecasting" (2022). *ECIS 2022 Research-in-Progress Papers*. 38.

[https://aisel.aisnet.org/ecis2022\\_rip/38](https://aisel.aisnet.org/ecis2022_rip/38)

This material is brought to you by the ECIS 2022 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2022 Research-in-Progress Papers by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact [elibrary@aisnet.org](mailto:elibrary@aisnet.org).

# EXPLANATION INTERFACES FOR SALES FORECASTING

*Research in Progress*

Tobias Fahse, University of St.Gallen, St. Gallen, Switzerland, tobias.fahse@unisg.ch

Ivo Blohm, University of St.Gallen, St. Gallen, Switzerland, ivo.blohm@unisg.ch

Richard Hruby, University of St.Gallen, St. Gallen, Switzerland,  
richard.hruby@student.unisg.ch

Benjamin van Giffen, University of St.Gallen, St. Gallen, Switzerland,  
benjamin.vangiffen@unisg.ch

## Abstract

*Algorithmic forecasts outperform human forecasts in many tasks. State-of-the-art machine learning (ML) algorithms have even widened that gap. Since sales forecasting plays a key role in business profitability, ML based sales forecasting can have significant advantages. However, individuals are resistant to use algorithmic forecasts. To overcome this algorithm aversion, explainable AI (XAI), where an explanation interface (XI) provides model predictions and explanations to the user, can help. However, current XAI techniques are incomprehensible for laymen. Despite the economic relevance of sales forecasting, there is no significant research effort towards aiding non-expert users make better decisions using ML forecasting systems by designing appropriate XI. We contribute to this research gap by designing a model-agnostic XI for laymen. We propose a design theory for XIs, instantiate our theory and report initial formative evaluation results. A real-world evaluation context is used: A medium-sized Swiss bakery chain provides past sales data and human forecasts.*

*Keywords: Explanation Interface, Sales Forecasting, Machine Learning, Design Science Research.*

## 1 Introduction

Algorithmic forecasts outperform human forecasts in many tasks (Grove et al., 2000). A meta-study found a 10% average performance gain of algorithmic forecasts compared to human forecasts (Grove et al., 2000). State-of-the-art machine learning (ML) algorithms have widened that gap (Blohm et al., 2020). Accurate sales forecasting plays a key role in businesses' profitability (Sun et al., 2008), because business decisions based on sales forecasting are considered particularly impactful in organizations, since numerous other functions depend on them (Mentzer and Bienstock, 1998). Hence, sales forecasting is recognized as one of the key value propositions of Artificial Intelligence (AI) (Cam et al., 2019). We delimit AI and ML by referring to AI as the superordinate phenomenon of building machines to mimic human behavior. The elaborations in this paper generally refer to ML, the most prevalent technical implementation of AI (Engel et al., 2022), although the general term AI is used in some contexts. Despite the significant advantages, individuals are resistant to use ML based forecasts. This algorithm aversion can have several causes.

**Decision makers do not understand algorithmic forecasts.** When selecting algorithms, companies have to make a trade-off between comprehensibility and performance. High-performing algorithms in the field of ML are less comprehensible and explainable than white box models such as linear regression models (Herm et al., 2021). On the other hand, Poursabzi-Sangdeh et al. (2021) show that participants of their study were less able to detect mistakes of white-box models compared to black-box models and attribute this phenomenon to information overload.

**Greater intolerance for algorithmic than for human errors.** People are particularly averse to the predictions of algorithms after seeing how they make predictions, even when those predictions outperform human predictions. If humans and algorithms make the same mistake, people lose confidence in the algorithm more quickly (Dietvorst et al., 2015).

**Lower willingness to trust and use algorithmic forecasts.** In practice, algorithmic advice is often ignored by humans in hybrid human-algorithmic systems (Prah and Van Swol, 2017), especially, when the task is perceived as subjective (Castelo et al., 2019). One way to mitigate this ignoring of advice is to increase trust in the algorithms giving the advice (Önköl et al., 2019).

**Low algorithmic literacy and experience of individuals.** When decision makers have low algorithmic literacy, they may have false expectations towards the algorithm, leading to algorithm aversion. Opposed to that, experience with algorithmic decision support correlates positively with the use of algorithmic decision support (Burton et al., 2020).

A potential solution to overcome algorithm aversion is explainable AI (XAI). In an XAI system, users interact with an explanation interface (XI) to request model predictions and explanations. This way, the behavior of ML forecasting systems becomes more comprehensible to humans (Gunning et al., 2019). However, XAI approaches for time series forecasting are scarce (Ates et al., 2020) and underexplored for users with little statistical education (Liao et al., 2020). Most research focuses on algorithm-centric explainability for ML experts. Since many decision-makers in the sales forecasting domain are no ML experts, XAI approaches for lay users are particularly important. In this paper, we focus on example-based explanations, which are highly promising for time series forecasts but lack evaluation (Rojat et al., 2021). The overarching goal of this research endeavor is to help increase decision quality of human decision makers who use ML forecasting systems. Decision quality is referred to as the accuracy with which decision makers forecast the next business day's sales (Geva and Saar-Tsechansky, 2021). Following the idea of hybrid intelligence, we aim to optimize the collaboration between human and AI in the team. For that, we design a XI tailored to the needs of lay users to drive the digitalization of work. We refer to lay users as individuals who are experts in their domain, but have little knowledge of ML or statistics. In this sense, lay users are the majority of users working in various business departments.

To address the outlined research gap, we design a model agnostic XI for sales forecasting. A design science research (DSR) approach is followed to answer the question: How can XAI design facilitate the decision quality of ML sales forecasting systems? We propose an emerging design theory for XI that draws on XAI, the theory of human explanations and cognitive load theory. We instantiate our theory as XI and report initial formative evaluation results as proof-of-concept. We use a real world context for this project: A medium-sized Swiss bakery chain that provides past sales data and human forecasts serves as the evaluation context. The final design theory will contribute to extending XAI approaches for time series forecasting and constitute prescriptive knowledge in the form of empirically validated design theory (Gregor and Hevner, 2013). For that, we investigate (1) the effectiveness of two selected example-based explanation approaches (factual and counterfactual examples), (2) the cognitive mechanisms through which example-based explanations may foster decision quality (3) the boundary conditions in which XI may improve decision quality of lay-users.

## **2 Conceptual and Theoretical Background**

The DSR project is grounded in three streams of literature: (1) XAI for sales forecasting; (2) the theory of human explanations; (3) cognitive load theory.

### **2.1 Explainable AI for Sales Forecasting**

Since numerous organizational decisions rely on forecasts, forecasting is important for business strategy (Mentzer and Bienstock, 1998). ML enables forecasting future events more accurately than traditional approaches (Kelleher et al., 2015), but the accuracy improvement comes at the cost of comprehensibility of predictions. XAI makes ML systems' behavior more comprehensible to humans

by providing explanations (Gunning et al. 2019). XIs serve as user interfaces which allow interacting with the model to request predictions and explanations (Mohseni et al., 2020). Effective explanations help mitigate human biases to prevent erroneous decisions (Wang et al., 2019). Unsatisfied explanation needs are diametral to decision quality (Liao et al., 2020). However, research has not yet answered how lay users' explanation needs can be effectively satisfied by designing XIs.

One can distinguish between global (explaining the inner logic of a ML model, i.e. the general mapping of inputs to outputs) and local (explaining specific model outputs) explainability (Adadi and Berradda, 2018). Local approaches can further be example-, feature-, and rule-based. Example-based explanations provide examples of either similar inputs that led to similar forecasts (factual examples) or that led to highly diverging forecasts (counterfactual examples) (Dandl et al., 2020). Feature-based approaches specify how much each input feature contributes to the prediction of a model (Apley and Zhu, 2020), providing insight into the workings of the algorithm for experts. Examples include SHAP (SHapley Additive exPlanations) or LIME (Local Interpretable Model-agnostic Explanations). Rule-based explanations provide a decision rule that is derived from the logic of a locally interpretable model (Guidotti et al., 2018). Explanations for time series can further be divided by the way they are generated. Backpropagation-based approaches perform a single forward and backward pass in the network. This way, sub-series in the input data that are most influential on the predicted classification are found. Perturbation-based approaches remove or alter parts of the time series and calculate the probability shift for a given output. Symbolic Aggregate approXimation (SAX), a data mining based approach, transforms the time series into strings to find patterns (Lin, 2003). Example-based approaches are k-nearest-neighbor (KNN) to find the closest example in the dataset to show how similar samples would behave (Lee, 2012) and shapelets (Ye and Keogh, 2011) to find subsequences of time series which are most representative of a class.

The described approaches have three disadvantages we would like to address specifically. First, most of the approaches aim to explain time series classification (rather than forecasting) problems and focus on feature-based explanations. In contrast to this, example-based explanations are very promising for time series forecasts but lack evaluation (Rojat et al., 2021). The potential of example-based explanations stems from their alignment with human reasoning processes and the minimal cognitive load for the user. Second, the approaches often fail to help users with low technical literacy (Cheng et al., 2019), especially if the data is hardly interpretable and highly dimensional such as in ML sales forecasting. Here, many abstract features (such as lag variables) are used, which are by nature incomprehensible for lay users. Even white-box models (e.g. regression models), which are commonly referred to as inherently explainable, produce outputs that are difficult to analyze by lay users (Hagras, 2018). Third, existing research is rather algorithm-centric although these mostly feature-based approaches have limited value for increasing decision quality of lay users (Alufaisan et al., 2021) and are thus inappropriate.

Against this background, the review paper by Abdul et al. (2018) identifies trends in XAI and notes that there is a need for increased usability, practical interpretability, and effectiveness for real-world users. Most explanations approaches in XAI research are static, assuming that there is a single message to be delivered. Interactive approaches "allow users to explore the system's behavior freely" (Abdul et al., 2018). Cheng et al. (2019) examine a XI for non-expert stakeholders that provides counterfactual explanations but focusses on ML classification tasks. In this study, addressing the disadvantages named above, we add to existing XAI approaches by designing a model agnostic XI that provides interactive example-based explanations for time series.

## 2.2 Theory of Human Explanations

Analyzing how people reason and explain uncovers valuable preconceptions about how to construct effective artificial explanations (Liao et al., 2020). Explanations can be looked at on three different levels:

**Explanations as cognition.** Explanations are created through cognitive processes involving abductive reasoning, causal attribution and selection and evaluation of explanations. Rather than directly

explaining the cause of an event, people emphasize the contrast between an event that occurred and another event that did not occur. The quality of the explanation is assessed based on whether inconsistencies in the mental models were resolved (Miller, 2019). Hence, explanations refer to specific forecasts, not models as whole.

**Explanations as products.** Explanations are products of these cognitive processes and have a contrastive nature, i.e., differences are inferred from (counter-) factual examples (Lombrozo, 2012). A factual example is a close neighbour of a given data point with a similar prediction. A counterfactual example is a close neighbour of a given data point but with a sufficiently close but different prediction (Dandl et al., 2020).

**Explanations as interaction.** Explanations are answers to “why questions” and involve the transfer of knowledge (Lombrozo, 2006). They are therefore social processes where explainer and explainee interact to close knowledge and belief gaps. They are needed due to unexpected or inconsistent events that challenge the user’s mental model (Hilton and Slugoski, 1986). Thus, explanations facilitate learning and improve one’s mental model through generalizing from the causes accountable for the discrepant events (Lombrozo, 2006).

Some key criteria for XI design for lay users of ML systems can be derived from these insights. First, explanations must be local, not global. Explanations are created through abductive reasoning, i.e. by starting with a (set of) observation and then finding the most likely conclusion from the observation. Local explanations constitute a way to adapt to this approach. In sales forecasting, events that need explanation are single unusually high or low forecasts, hence decision makers need local explanations. The inner logic of the ML forecasting system is irrelevant for them, as it has no direct business consequence. Second, example-based explanations are preferred against feature-based explanations. As human explanations have a contrastive nature, i.e. people rather explain the difference between two specific events than the cause of an event directly, XI design should adopt this approach by providing factual and counterfactual examples. Third, explanation of performance should be part of the XI. People create various potential explanations in abductive reasoning. The subsequent selection process is influenced by prior beliefs and knowledge. If the performance of past forecasts is communicated, people might be more likely to accept the forecast and explanation bundle (Zhang et al., 2020). In summary, lay users have no option but to explain algorithmic forecasts in the way they explain phenomena in general. It is necessary to provide human-consumable explanations, as it is the most seamless way for lay users to make sense of algorithmic outputs (Liao et al., 2020).

## **2.3 Cognitive Load Theory**

Cognitive load is an additive construct whose assumption is that humans use a short-term working memory that processes information and a long-term memory for storing information. The working memory processes all conscious cognitive tasks but the number of processable information items is limited (Sweller, 1988). One can distinguish between intrinsic (inherent complexity of tasks) and extraneous (complexity of how tasks are presented) cognitive load (Blohm et al., 2016). Details on the historical accuracy of the algorithm and the visualization of forecasts pose extraneous cognitive load on the user. The number of forecasts and explaining examples displayed and the complexity of explanations pose intrinsic cognitive load on the user. Although we assume that XAI approaches will be beneficial for users’ decision quality, too much information on the XI could have a negative impact. Thus, XIs should minimize cognitive load as it is diametral to task performance. Considerable load is posed for users who actively contribute cognitive effort to process ML forecasts. Explanations are an additional source of cognitive load and may lead to cognitive overload and decreased decision quality. Consequently, the explanations’ goals may not be met. In fact, existing research shows that interfaces that pose too high cognitive load lead to inferior decision quality in forecasting (Blohm et al., 2016). This situation is even more problematic for lay users who are unfamiliar with ML forecasting systems. Lay users have less domain knowledge and experience and hence less extensive cognitive structures that allow them to process several elements invoking cognitive load simultaneously. Thus, cognitive load theory suggests that different XAI approaches are needed for lay users.

### 3 Methodology

DSR is concerned with creating innovative artifacts (e.g. design theories) seeking to extend the boundaries of human and organizational capabilities (Hevner et al., 2004). A design theory can be defined as an integrated prescription that consists of a particular class of requirements, effective development practices, and a system solution with distinctive features (Walls et al., 1992). We follow an established DSR process to develop and test a design theory for an XI and adopt the six phases of Peffers et al. (2008). (1) The problem of currently unused potential through ML sales forecasting is identified. (2) Theoretical insights are integrated with practical considerations to define the design objective, i.e., helping decision-makers make better decisions. (3) The design theory is instantiated as a XI. (4) The solution is demonstrated to domain experts to capture their feedback. (5) The artifact will be evaluated regarding its ability to satisfy the design objective. (6) The results are communicated in terms of a design theory. This process is performed in an iterative way, allowing for continuous refinement of the design results (Hevner et al., 2004). Following the schema provided by Gregor et al. (2020), each design principle (DP) contains the aim, user, context, mechanism and a rationale. The artifact is positioned as exaptation as we apply solutions derived existing research (see chapter 2) to a new problem, namely the explanation needs of lay users who use a new technology advance, i.e., ML decision support. The way ML models derive decisions is often too complex to grasp by lay users, leading to difficulties to trust and use such systems. This presents a new challenges in the field of ML decision support (Gregor and Hevner, 2013).

We address this DSR project with a medium-sized Swiss bakery chain that provides access to historic sales data and human sales predictions, thus serving as an evaluation context for our XI. The unique data access, the comparably low data literacy of decision makers, and the perishability of food - making forecasting particularly important - renders an ideal testbed for our project.

### 4 Artefact Description and Preliminary Results

We present the preliminary results from the first iteration that intends to conceptualize an initial design theory for XIs, including design requirements, principles, and features. Design requirements are derived from a comprehensive literature review and describe general goals for the design of the theory and its prototypal implementation as XI (Walls et al., 1992). Design principles serve as prescriptive guidelines for instantiating our design theory (Meth et al., 2015) and are implemented as design features to address design requirements. To ensure grounding and viability of our initial design theory, we use formative evaluation (Venable et al., 2016) and report preliminary findings.

#### 4.1 Objectives, Design and Development of the Design Theory

Based on existing research in the fields of XAI, sales forecasting, human explanations, and cognitive load theory, we derive three **design requirements** (DR) for XIs to reap the benefits of ML sales forecasting.

**DR1.** Help people make better decisions. By focusing on XAI approaches that are comprehensible for lay users, XIs should facilitate users to learn why single predictions were made. The XI should allow users to learn domain specific insights, so that they can improve the quality of their decisions. By allowing users to appropriately assess the ML system's capabilities, users' perception of the ML system's accuracy is calibrated which leads to higher decision quality (Mohseni et al., 2020).

**DR2.** Support human reasoning processes. Aligning XI design with the characteristics of human explanations increases the decision quality of humans. If in harmony with human explanations, artificial explanations are more effective and efficient which leads to higher decision quality.

**DR3.** Minimize cognitive load. XIs need to increase the efficiency of the explanation process, by minimizing the cognitive load that is posed on the user.

To conceptualize the design theory, the design requirements are translated into actionable **design principles** (DP) and context-specific design features (Meth et al. 2015), as illustrated in Figure 1.

**DP1.** XIs should focus on local explanations to support human reasoning (DR2) and allow users to request evidence for the algorithm’s output. Explanations of specific model predictions assist human reasoning processes (DR2) and foster the decision quality of users (DR1) (Ye and Johnson, 1995).

**DP2.** XIs should show users the benefit of algorithmic predictions. By summarizing relevant outcomes appropriately, users can efficiently gain understanding of the ML system’s performance capabilities and thereby increase decision quality (DR1) (Few, 2006; Mohseni et al., 2020).

**DP3.** XIs should support example-based explanations. As human explanations have a contrastive nature, XIs should adapt by providing example-based explanations (DR2) (Mohseni et al., 2020). Example-based explanations mimic human reasoning better than feature-based explanations, so that there is no need to build new cognitive structures for processing. Consequently, they can be better processed with existing cognitive structures and thus impose lower cognitive load (DR3).

**DP4.** XIs should leverage visual perception. Visualizations of information pose low cognitive load on the user (DR3) and comply with the nature of how humans process explanations (Miller, 2019) (DR2).

**DP5.** XIs should be designed for interaction. Effective and efficient explanations can only be provided through meaningful interaction which represents the explanatory conversation in the digital sense (Miller, 2019). Unexpected predictions trigger ‘why’ questions which are answered by explanations provided on request of the user (Miller, 2019). Such interactions support human reasoning (DR2) and allows for discovering erroneous predictions, thus facilitating decision quality (DR1).

The following describes **design features** (DF) that map onto our design principles to address design requirements. The control panel (DF1) enables users to interactively generate, modify and submit forecasts (DP5). The visualization panel (DF2) displays historic time series and forecast (DP4). Explanations are generated upon the user’s request (DP5), thereby minimizing cognitive load. By default, the prototype only shows forecasts of different products per day. Users may drill down to request an explanation by selecting a day on the visualization panel. The performance explanation panel (DF3) displays the algorithm’s improvement over a naive forecast (DP2). Factual explanations (DF4) enable the user to identify most similar historic days supporting the forecast using a KNN algorithm (MacQueen, 1967) that calculates the k nearest data points to the selected to be predicted day. In our implementation, KNN uses the squared euclidean distance to identify the seven most similar days in the past. Individuals can thus construct a mental model of the forecast by creating analogies from representative cases in which the algorithm’s outcome was accurate. The XI provides factual example-based explanations (DP3) encompassing two data tables. The selected data point is described in the stop table. The bottom table presents the most similar historic data points in ranked order. These tables use lay-user-friendly, context-specific feature descriptions and support interactivity through customizable column visibility and sorting (DP5). Comparing instances allows attributing potential causes as forecast explanations. Counterfactual examples (DF5) create contrastive explanations. Users can make specific predictions by altering feature values. This way, they can create counterfactual explanations (DP3) by examining how different feature values affect the forecast. By enabling meaningful comparisons that are consistent with human reasoning, users can locally explain predictions (DP1).

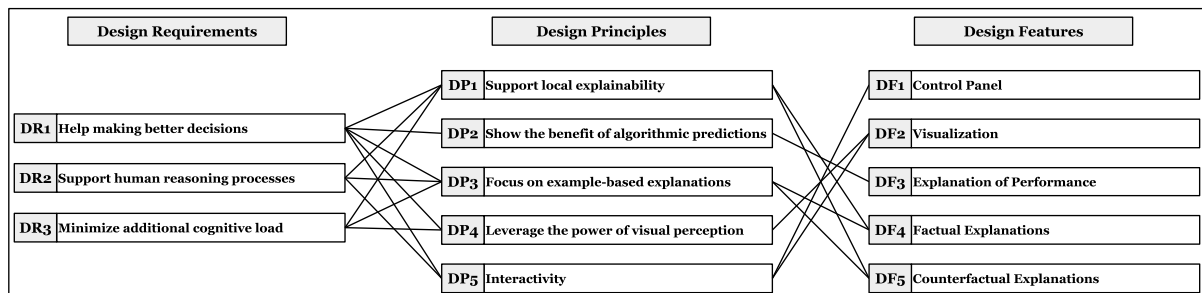


Figure 1. Preliminary conceptualization of design requirements, principles and features.

## 4.2 Initial Implementation as Explanation Interface

The ML forecasting model uses transactional point-of-sale bakery data. The time series covers 5 years of sales and contains a timestamp and the features type of product, amount, and price. External data was used to add weather indicators (radiation, rainfall, temperature), national holidays, and school vacations. Performing extensive autocorrelation analysis, we also included lagged features (lag 1, 2, and 7) as well as a rolling average over 7 days. Based on these features, an XGBoost model (Chen and Guestrin, 2016) makes one-step ahead forecasts, that is, the sales of the next day. We opted for the XGBoost model on the basis of a performance comparison (mean absolute percentage error, MAPE) of various ML algorithms. Out of six tested models (linear regression, decision tree, random forest, XGBoost, ARIMA, Neural Network), XGBoost performed best. However, our generic prescriptions are not bound to the peculiarities of a specific algorithm. We first evaluate the accuracy of our model to establish a baseline regarding business value (cycle 0). The ML model outperforms the forecast of the bakery's branch managers by 6.5% (MAPE) on average over all products. The branch managers' forecasts are provided to the production department and are referred to as human forecast in this paper. Considering a margin of 70% (SBC, 2019), this increase in forecast accuracy equals yearly savings potential of approximately CHF 13000 due to reduced waste in the branch with the highest unit sales. This equals 3.1% of the yearly revenue in that branch. Thus, following the ML predictions without any change would already lead to significant savings. This confirmation is important to justify any effort to design an XI for ML forecasting. The prototypical XI is illustrated in figure 2.

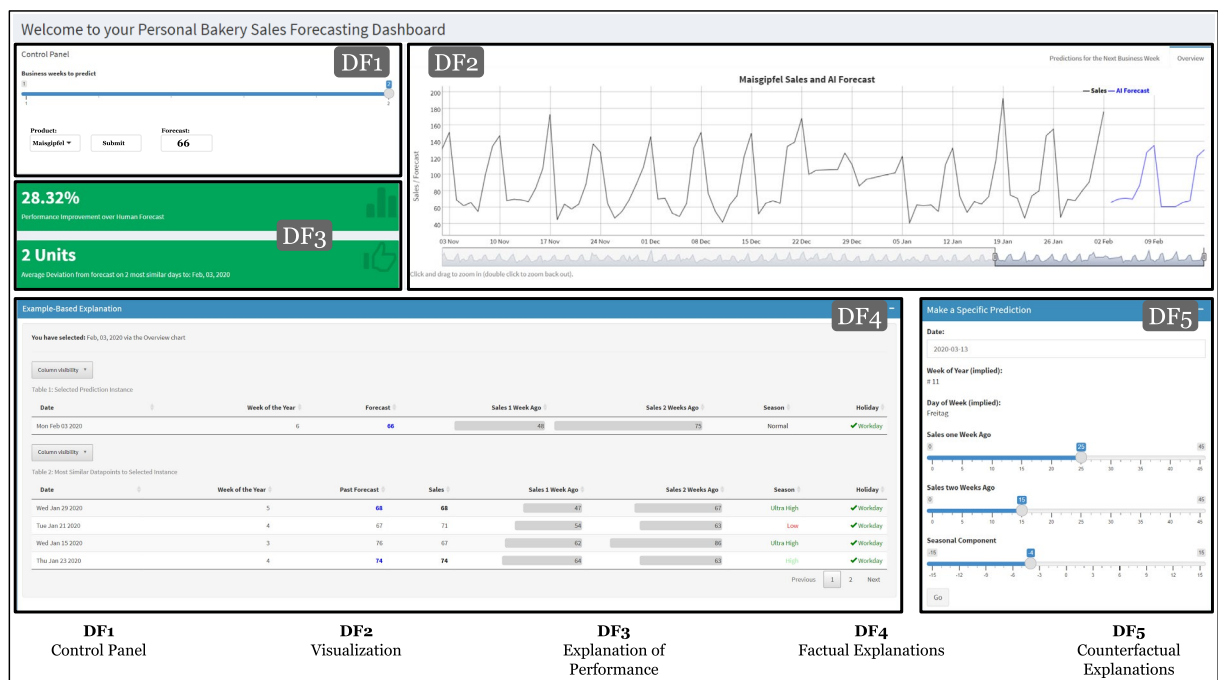


Figure 2. Prototypical explanation interface for sales forecasting.

## 4.3 Preliminary Evaluation Results

To evaluate our design theory, we follow the framework proposed by Venable et al. (2016) and engage in a process of formative evaluation. We validated our design principles and design features through five expert interviews with decision-makers of five different Swiss bakeries (two production managers and three CEOs). In summary, the interviews confirmed that the components of our conceptualization improve decision makers' decision quality. Hence, the central design requirements can be considered as being met. However, in terms of completeness, it seems that one design principle was missing. Although our XI featured a control panel (DF3), the notion of providing "on-demand explanations"

was not fully addressed. The interviewees called for displaying predictions for all products at once to be more efficient when making decisions (DR1). However, if the user's mental model is challenged, the interviewees appreciate the possibility to drill down to the details of the prediction's explanation.

Following a central mantra of information dashboarding - "overview first, details on demand" (Shneiderman, 2003) - we thus want to add new design principles and rearrange our design theory and instantiation accordingly. In terms of design features, factual and counterfactual explanations have been considered as most telling such that we want to strengthen their role in the next design iteration.

## **5 Next Steps and Expected Contributions**

Further design-and-evaluate cycles are planned to complete our DSR project. We assessed our preliminary design theory and the XI prototype by a formative evaluation. We will finish cycle 1 by integrating the results of the evaluation in our design theory. That is, we will provide an overview of the products' forecasts when opening the XI. In cycle 2, we will conduct a series of lab experiments. Given our initial evaluation results, we will focus on the effect of example-based explanations on the participant's decision quality. In the first experiment, we will employ a 2x2 between-subjects design with factual and counterfactual examples being the experimental factors. The control group will use an XI that only consists of the control panel (DF1) and the visualization (DF2). The other groups will additionally have access to factual explanations only (DF4), counter-factual explanations only (DF5), or both. In line with existing research, participants will be recruited from online platforms such as Prolific (50 per condition, 200 in total) (Palan and Schitter, 2018). The task will be to forecast the next business day's sales of 20 products using the XI. We will employ a post-experimental survey to evaluate the degree to which our design requirements are met as well as the relevance of our design principles. In additional experiments, we will vary this baseline scenario to increase cognitive load. In experiment 2, we will add performance explanations (DF3) that may help individuals to select effective explanations but also may reflect an additional source of (extraneous) cognitive load. In experiment 3, we are increasing the intrinsic cognitive load by integrating more complex factual and counterfactual explanations. For instance, we might use influential examples that have particularly influenced a specific forecast (Koh and Liang, 2017) for factual explanations and counter-factual explanations according to Dandl et al. (2020). This experimental setup allows not only to investigate the effect of our design features and the overall design theory, but also to better understand the boundary conditions under which XAI may provide value in sales forecasting. In the final third cycle, we will perform a field test in a bakery to evaluate the effectiveness of the XI in the field, i.e. in a real-world context. We will track the decision quality by measuring the prediction accuracy of the decision-makers when given the XI. A control group will be given the same algorithmic predictions but no XI.

The final design theory will contribute to extending XAI approaches for time series forecasting and constitute prescriptive knowledge in the form of empirically validated design theory (Gregor and Hevner, 2013). We contribute to theory by investigating (1) the effectiveness of two selected example-based explanation approaches (factual and counterfactual examples) in time series forecasting (e.g. Ates et al., 2021; Dandl et al., 2020; Ye and Keogh, 2011; Koh and Liang, 2017), (2) the cognitive mechanisms through which example-based explanations may foster decision quality by posing the proper amount of cognitive load on the user (e.g., Miller, 2019; Lombrozo, 2006; Lombrozo, 2012), (3) the boundary conditions in which XI may improve decision quality of lay-users (Liao, 2020; Cheng et al., 2019; Wang et al., 2019). For practice, we hope to provide guidance for building XIs for lay users in sales forecasting by providing generalizable design guidelines and to enable decision makers to make better sales forecast decisions.

## References

- Abdul, A., Vermeulen, J., Wang, D., Lim, B. Y. and Kankanhalli, M. (2018). "Trends and Trajectories for Explainable, Accountable and Intelligible Systems: An HCI Research Agenda." *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, 1–18. New York, NY, USA: Association for Computing Machinery.
- Adadi, A. and Berrada, M. (2018). "Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)", *IEEE Access*, 6, pp. 52138–52160.
- Alufaisan, Y., Marusich, L. R., Bakdash, J. Z., Zhou, Y. and Kantarcioglu M. (2021). "Does Explainable Artificial Intelligence Improve Human Decision-Making?." *Proceedings of the AAAI Conference on Artificial Intelligence* 35 (8), 6618–6626.
- Apley, D. W. and Zhu, J. (2020). "Visualizing the effects of predictor variables in black box supervised learning models." *Journal of the Royal Statistical Society. Series B: Statistical methodology* 82 (4), 1059–1086.
- Ates, E., Aksar, B., Leung, V. J. and Coskun, A. K. (2021). "Counterfactual explanations for multivariate time series". In: *2021 International Conference on Applied Artificial Intelligence (ICAPAI)*. IEEE.
- Blohm, I., Antretter, T., Sirén, C., Grichnik, D. and Wincent, J. (2020). "It's a Peoples Game, Isn't It?! A Comparison Between the Investment Returns of Business Angels and Machine Learning Algorithms", *Entrepreneurship Theory and Practice*.
- Blohm I., Riedl, C., Füller, J. and Leimeister, J. M. (2016). "Rate or Trade? Identifying Winning Ideas in Open Idea Sourcing", *Information Systems Research* 27 (1), 27–48.
- Burton, J. W., Stein, M. K. and Jensen, T. B. (2020). "A systematic review of algorithm aversion in augmented decision making", *Journal of behavioral decision making* 33 (2), 220–239.
- Cam, A., Chui, M. and Hall, B. (2019). *Global AI Survey: AI proves its worth, but few scale impact*. URL: <http://dl.n.jaipuria.ac.in:8080/jspui/bitstream/123456789/1323/1/Global-AI-Survey-AI-proves-its-worth-but-few-scale-impact.pdf> (visited on 16. November 2021).
- Castelo, N., Bos, M. W. and Lehmann, D. R. (2019). "Task-Dependent Algorithm Aversion", *Journal of marketing research* 56 (5), 809–825.
- Chen, T. and Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. In: *arXiv [cs.LG]*. Available at: <http://arxiv.org/abs/1603.02754>.
- Cheng, H. F., Wang, R., Zhang, Z., O'Connell, F., Gray, T., Harper, F. M. and Zhu, H. (2019). Explaining decision-making algorithms through UI: Strategies to help non-expert stakeholders. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland, UK.
- Dandl, S., Molnar, C., Binder, M., Bischl, B. (2020). Multi-Objective Counterfactual Explanations. In: *Parallel Problem Solving from Nature – PPSN XVI*. Springer International Publishing, 448–469.
- Dietvorst, B. J., Simmons, J. P., Massey, C. (2015). "Algorithm aversion: people erroneously avoid algorithms after seeing them err." *Journal of Experimental Psychology: General* 144 (1), 114–126.
- Engel, C., Ebel, P. and Leimeister, J. M. (2022). "Cognitive Automation." *Electronic Markets*.
- Few, S. (2006) *Information Dashboard Design: The Effective Visual Communication of Data*. O'Reilly.
- Geva, T. and Saar-Tsechansky, M. (2021) "Who is a better decision maker? Data-driven expert ranking under unobserved quality." *Production and operations management* 30 (1), 127–144.
- Gregor, S., Chandra Kruse, L., Seidel, S., Kruse, L., and Seidel, S. (2020). "Research perspectives: The anatomy of a design principle." *Journal of the Association for Information Systems* 21(6), 1622–1652.
- Gregor, S. and Hevner, A. R. (2013) "Positioning and presenting design science research for maximum impact." *Management Information Systems Quarterly* 37 (2), 337–355.
- Grove, W. M. et al. (2000) "Clinical versus mechanical prediction: a meta-analysis." *Psychological assessment* 12 (1), 19–30.

- Guidotti, R., Monreale, A., Ruggieri, S., Pedreschi, D., Turini, F., Giannotti, F. (2018). Local Rule-Based Explanations of Black Box Decision Systems. In: arXiv [cs.AI]. Available at: <http://arxiv.org/abs/1805.10820>.
- Gunning, D., Stefik, M., Choi, J., Miller, T., Stumpf, S., Yang, G.-Z. (2019). “XAI-Explainable artificial intelligence.” *Science robotics* 4 (37).
- Hagras, H. (2018) “Toward Human-Understandable, Explainable AI.” *Computer* 51 (9), 28–36.
- Herm, L. V., Wanner, J., Seubert, F. and Janiesch, C. (2021). I Don’t Get It, but It Seems Valid! The Connection Between Explainability and Comprehensibility in (X)AI Research. In: *Proceedings of the 2021 European Conference on Information Systems*.
- Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). “Design science in information systems research.” *Management Information Systems Quarterly* 28 (1), 75-105.
- Hilton, D. J. and Slugoski, B. R. (1986). “Knowledge-Based Causal Attribution: The Abnormal Conditions Focus Model.” *Psychological Review* 93 (1), 75–88.
- Kelleher, J. D., Namee, B. M. and D’Arcy, A. (2015) *Fundamentals of Machine Learning for Predictive Data Analytics: Algorithms, Worked Examples, and Case Studies*. The MIT Press.
- Koh, P. W. and Liang, P. (2017). Understanding Black-box Predictions via Influence Functions. In: *arXiv [stat.ML]*. Available at: <http://arxiv.org/abs/1703.04730>.
- Lee, Y.-H., Wie, C.-P., Cheng, T.-H., Yang, C.-T. (2012). “Nearest-neighbor-based approach to time-series classification.” *Decision support systems* 53 (1), 207–217.
- Liao, Q. V., Gruen, D. and Miller, S. (2020). Questioning the AI: Informing Design Practices for Explainable AI User Experiences. In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*.
- Lin, J., Keogh, E., Lonardi, S. and Chiu, B. (2003). A symbolic representation of time series, with implications for streaming algorithms. In: *Proceedings of the 8th ACM SIGMOD workshop on Research issues in data mining and knowledge discovery*. New York, NY, USA.
- Lombrozo, T. (2006). “The structure and function of explanations.” *Trends in Cognitive Sciences* 10 (10), 464-470.
- Lombrozo, T. (2012). *Explanation and Abductive Inference*. In: Holyoak, K. J. and Morrison, R. G. (eds.) *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press.
- MacQueen, J. (1967). “Some methods for classification and analysis of multivariate observations.” *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability* 1 (14), 281–297.
- Mentzer, J. T. and Bienstock, C. C. (1998). *Sales Forecasting Management: Understanding the Techniques, Systems and Management of the Sales Forecasting Process*. SAGE Publications.
- Meth, H., Mueller, B. and Maedche, A. (2015). “Designing a Requirement Mining System.” *Journal of the Association for Information Systems* 16 (9), 799-837.
- Miller, T. (2019). “Explanation in artificial intelligence: Insights from the social sciences.” *Artificial intelligence* 267, 1–38.
- Mohseni, S., Zarei, N. and Ragan, E. D. (2020). A Multidisciplinary Survey and Framework for Design and Evaluation of Explainable AI Systems. In: *arXiv*. Available at: <http://arxiv.org/abs/1811.11839>.
- Önköl, D., Gönöl, M. S. and De Baets, S. (2019). “Trusting Forecasts.” *Futures & Foresight Science* 1 (3-4), 1-26.
- Palan, S. and Schitter, C. (2018). “Prolific.ac—A subject pool for online experiments.” *Journal of Behavioral and Experimental Finance* 17, 22–27.
- Peffer, K. et al. (2008). “A Design Science Research Methodology for Information Systems Research.” *Journal of Management Information Systems* 24 (3), 45–77.
- Poursabzi-Sangdeh, F., Goldstein, D. G., Hofman, J. M., Wortman Vaughan, J. W., Wallach, H. (2021). Manipulating and measuring model interpretability. In: *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. New York, NY, USA.
- Prahl, A. and Van Swol, L. (2017). “Understanding Algorithm Aversion: When Is Advice from Automation Discounted?” *Journal of Forecasting* 36 (6), 691–702.

- Rojat, T., Puget, R., Filliat, D., Del Ser, J., Gelin, R. and Díaz-Rodríguez, N. (2021). Explainable Artificial Intelligence (XAI) on TimeSeries Data: A Survey. In: *arXiv [cs.LG]*. Available at: <http://arxiv.org/abs/2104.00950>.
- SBC (2019). *Branchenspiegel für das Bäckerei-, Konditorei- und Confiserie-Gewerbe*. Schweizerischer Bäcker-Confiseurmeister-Verband (SBC). Available at: [https://www.swissbaker.ch/inhalte/03\\_dokumente/verband/de/sbc\\_branchenspiegel\\_2019\\_d.pdf](https://www.swissbaker.ch/inhalte/03_dokumente/verband/de/sbc_branchenspiegel_2019_d.pdf).
- Shneiderman, B. (2003). *The eyes have it: A task by data type taxonomy for information visualizations*. The Craft of Information Visualization. Elsevier, 364–371.
- Sun, ZL., Choi, T. M., Au, K. F. and Yu, Y. (2008). “Sales forecasting using extreme learning machine with applications in fashion retailing.” *Decision support systems* 46 (1), 411–419.
- Sweller, J. (1988). “Cognitive load during problem solving: Effects on learning.” *Cognitive science* 12 (2), 257–285.
- Venable, J., Pries-Heje, J. and Baskerville, R. (2016). “FEDS: A Framework for Evaluation in Design Science Research.” *European Journal of Information Systems* 25 (1), 77–89.
- Walls, J. G., Widmeyer, G. R. and El Sawy, O. A. (1992). “Building an Information System Design Theory for Vigilant EIS.” *Information Systems Research* 3 (1), 36–59.
- Wang, D., Yang, Q., Abdul, A. and Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. Glasgow, Scotland, UK.
- Ye, L. R. and Keogh, E. (2011). “Time series shapelets: a novel technique that allows accurate, interpretable and fast classification.” *Data mining and knowledge discovery* 22 (1), 149–182.
- Ye, L. R. and Johnson, P. E. (1995). “The impact of explanation facilities on user acceptance of expert systems advice” *Management Information Systems Quarterly* 19 (2), 157 – 172.
- Zhang, Y., Liao, Q. V. and Bellamy, R. K. E. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In: *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. New York, NY, USA.