University of St.Gallen
School of Business, Law, Social Sciences,
International Affairs and Computer Science (HSG)

# Explainable AI
# Development of a machine learning based sales forecast and implementation of an interactive web application

**Master Thesis**

Philipp Schmelzer
14-603-254
Philipp.schmelzer@student.unisg.ch
Master in Business Innovation
University of St. Gallen

**Supervisor**

Professor Dr. Ivo Blohm
Assistant Professor of Data Science and Management
IWI - Institute of Information Management

St. Gallen, 25th of May 2021

## Abstract

Our present and future lives are increasingly determined by algorithms based on artificial intelligence. However, these often represent a "black box" and it cannot be determined how exactly these algorithms achieve the outcome. This master thesis employs the principles of explainable artificial intelligence to develop an interactive web application that explains the results of a machine learning based sales forecast in a way that is understandable to non-specialist users. The explanation interface, which was developed in collaboration with business experts from the catering industry following an iterative prototyping cycle based on the principles of design science research, incorporates design principles that were identified through a literature review and adapted to the specific context of the work with the help of expert interviews. The design features in the explanation interface follow the design principles "Added Layer of Interpretability", "Ease of Use", "Interactivity", "Comparability" and "Factor / Feature Importance" and satisfy the defined design requirements "Improving Transparency", "Improving Trust" and "Improving User Understanding of Bias and Model Behaviour". As particularly relevant for the fulfilment of the objectives set in the thesis, the design features "Overall Overview", "Landing Page / Explanatory Tour", "User Input Functions", "Comparison Elements" and "Influencing Factors" were identified.

**Keywords:** Explainable Artificial Intelligence, Machine-Learning, Post-Hoc Explanation, Explanation Interface, Times Series Forecasting, Sales Prediction, Shiny R, Design Science Research

## Table of Content

## List of Figures

## List of Tables

## List of Acronyms and Abbreviations

| | |
|---|---|
| AI | Artificial Intelligence |
| BAFU | Federal Office for the Environment (Switzerland) |
| DARPA | Defence Advanced Research Projects Agency |
| DSR | Design Science Research |
| DL | Deep Learning |
| FAT | Conference for Fairness, Accountability, and Transparency |
| GUI | Graphical User Interface |
| ML | Machine Learning |
| UI | User Interface |
| RQ | Research Question |
| RMSE | Root Mean Squared Error |
| XAI | Explainable Artificial Intelligence |

# 1 Introduction and Formulation of Research Questions

The global market for artificial intelligence (AI) is growing strongly and it is increasingly finding its way into all areas of business and life (GrandViewResearch, 2020). This is exacerbated by the ever-increasing capacity of global computing power and the rapid increase in the amount of data available worldwide (Hagras, 2018; IDC & Seagate, 2018). Gartner emphasises the growing importance of AI for the future, listing the technology among the top 10 Strategic Technology Trends for 2018. Current development shows that AI with its various applications is an integral part of our society and economy; it will shape business models in the future and permanently change entire industries (Panetta, 2017).

Applications of AI and machine learning (ML) have many profitable use cases in both business and daily life. It is standard, for example, for credit granting mechanisms to be monitored and evaluated by AI without further oversight by human personnel (Bussmann, Giudici, Marinelli, & Papenbrock, 2020). In the United States, an AI-based system is used to assess the recidivism rate of individuals sentenced for a crime. This has generated considerable public attention because it was discovered that the AI-System systematically discriminates against black U.S. citizens (Holland, 2016; Kirkpatrick, 2017). The efficiency and objectivity of AI are also being relied upon in recruitment processes. For instance, companies such as Amazon rely on AI technology to process job applications. However this has also come under criticism, as it has been found that women, among others, are systematically disadvantaged by this system, (Dastin, 2018; Köchling & Wehner, 2020).

There is a growing desire for more accountability from the various AI stakeholders, not only because of past cases of discrimination found in AI technology but also due to the increasing number of important decisions made using AI technology. Due to the increasingly significant impact of AI on private and economic life, the results and recommendations of this technology must be made accessible in an explainable and interpretable manner to an ever-widening audience. But algorithms tend to be complex and opaque without organizations fully understanding exactly how they function (Adadi & Berrada, 2018). These algorithms are perceived by the social and economic community as a "black box" which are producing results for which it is unclear how they arise or how they are to be interpreted. This is further aggravated by the current lack of organisational knowledge and skills regarding AI and ML (Oxborough, Rao, Cameron, & Westermann, 2018). Because of these developments, there is a demand for transparent algorithms that incorporate an ethical component: The functionality and the algorithm itself should be constructed or explained in a way it can be better understood and therefore the results better interpreted. Furthermore, data sets should be free of bias against race, gender and other criteria. (Guidotti, Monreale, & Ruggieri, 2018). Even the European Union has taken up this issue and has passed laws to regulate algorithmic decision-making, to guarantee a "right to meaningful explanation" of the underlying logic of a model (Goodman & Flaxman, 2016).

In order to resolve this tension about the comprehensibility of an algorithm, Explainable Artificial Intelligence (XAI) offers an approach to achieve more transparency in the field of AI. Its objective is to develop a set of tools that generates explainable models while keeping a high level of performance (Das, Member, Rad, & Member, 2020). Still, this improved transparency comes with a trade-off between performance and the explainability of an algorithm (Oxborough et al., 2018, p. 10). For example, in DARPA's XAI research project, a wide variety of learning algorithms are compared in terms of their performance and accuracy. It was discovered that deep learning methods and neural networks exhibit the best performance but are also extremely difficult to comprehend or explain. Decision trees, on the other hand, have a relatively low performance but are easy to explain. (Gunning & Aha, 2019, p. 46).

The following figure illustrates the concept of XAI and the differences between a black box approach as it is used today and an explainable AI as it is urgently needed in the future.

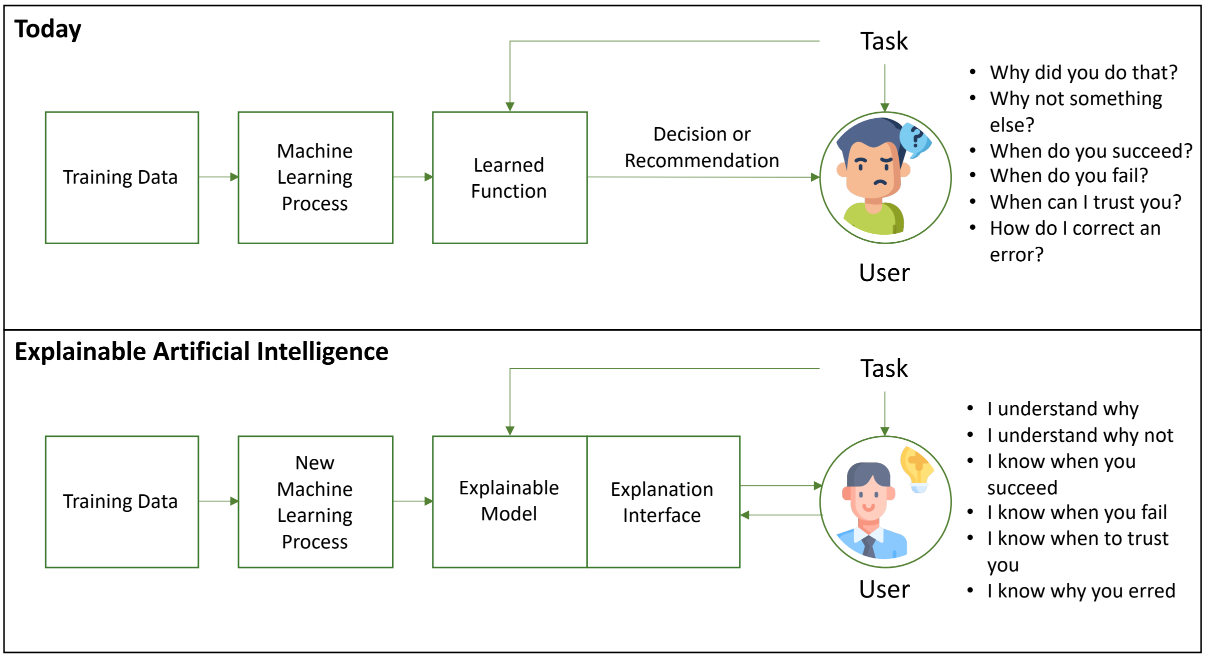



*Figure 1: Concepts of AI and XAI, according to Gunning & Aha, 2019, p. 48.*

The core difference between the two approaches is that a "normal" AI model may give a user a decision or recommendation in which the user may not fully understand the logic behind it. In such a case, the user does not know why exactly the algorithm came to this result, whether the calculation is flawed, or whether the user can trust the result. Although the user has received additional information for decision making, the validity of that decision cannot be verified or explained. In the case of an XAI application, the user is aware of what the algorithm's recommendation says and how to interpret it. Further, the user knows the errors and biases in the system and knows when to trust the recommendation. Because of this improved transparency and interpretability, the user has an enhanced capacity to make a well-reasoned decision. (Gunning & Aha, 2019, pp. 47–49).

An explainable algorithm can be realized in two different ways. On the one hand, transparency can be achieved by focusing on opening the black box of the algorithm and explaining its inner workings. This approach is about understanding how the algorithm itself works, i.e., it provides the user with an explainable model. On the other hand, the focus can be on the explainability itself, i.e., an attempt is made to explain the result of the algorithm to a user. This approach is mainly employed when companies want to understand the results of algorithms when they are used for decision-making support. (Oxborough et al., 2018, p. 5).

Thus, an intrinsic explanation occurs when the explanatory capacity is already built into the architecture of the algorithm. An explanation method developed in this way is not transferable to other algorithms and can only be used for this individual model. If the explanation is not adapted to the architecture of the individual algorithm and can be used for several already trained algorithms, one speaks of a post-hoc explanation method (Das et al., 2020, p. 2).

The motivation for this research work stems from the aspiration to bring AI-based applications closer to non-technical people, enabling them to make their daily operational tasks easier and more efficient. The complex and sometimes difficult to explain functionality of predictive models is a major barrier for the adaptation of such applications in many non-technical areas. However, it is precisely in these industries, which have a low level of technologization, where there is an enormous potential for optimization. Food waste, for example, is a widespread problem in the catering industry. According to a study commissioned by the BAFU around 14% of all food waste occurs in the catering industry (Beretta & Hellweg, 2019, p. 32). This waste in turn has a strong negative effect on the operating profit of companies, but it also places a heavy burden on the environment. Accurate and detailed forecasting of food demand can lead to data-driven menu planning with less waste and a better operating result. However, to realize this optimization, the prediction and its results must match the organizational knowledge level of the foodservice system. Only if the users of the application can understand, interpret and trust the results of the demand forecast, a sufficient level of adaptation can be guaranteed. This goal is to be achieved with the help of the principles of XAI.

In order to achieve the goals of this thesis, design principles are derived from the current scientific field of XAI and adapted for practice in order to apply them in a practical business case to support business users in their daily work in the catering industry. In a subsequent step, the design principles will be further developed through expert interviews and used for the development of a prototype of a Graphical User Interface (GUI) that business users can utilise in their daily work to better interpret and understand the results of the Random Forest-based sales forecast. This algorithm is used to predict the future sales of a given institution, which according to DARPA (2019) has relatively high performance but is difficult to comprehend and explain (p. 46). Since the results of the prediction are utilized to support decision-making processes, the focus lies on a post-hoc explanation of the results, which are implemented with an explanation interface designed in the form of an interactive and user-friendly website.

The developed interactive web application, based on the derived design principles, is used to transform the time series forecasting results into user-centric and understandable insights to improve the efficiency of daily decision making. The resulting interactive GUI is designed in a way to ensure that even non-specialist users can access the visualisation intuitively and quickly, that the training period can be kept as short as possible by making the application easy to use, and that frustrations in using the application can be avoided.

The structure of the thesis is based on the research questions below. For each of the following research questions, the corresponding procedure and the methods used to answer the question, as well as the desired outcome, are explained.

**Research question 1:**

RQ 1: *"What are the underlying design principles of explainable AI?"*

To be able to answer the first research question, Chapter 2.1 lays the foundation for the theoretical principles of XAI. Design requirements are derived based on the scientific literature, which the resulting web application must fulfil in order to meet the requirements of XAI. Based on a literature research design principles are collected, which can be employed to translate the design requirements into specific design features in the resulting web application. In the Chapter 2.2, these design principles are evaluated and checked for their applicability in practice. With the help of these identified design principles, the first research question is answered.

**Research question 2:**

RQ 2*: "What design principles must be incorporated in a visualization in order to illustrate the results of a time series forecasting model in a comprehensible, intuitive and user-friendly way?".*

To answer the second research question, seven expert interviews were conducted to gather a list of user needs. Based on the derived needs and requirements of the participants, the design principles are adapted to the specifications of the catering industry in order to ensure that they are suitable for use in the web application. In parallel, a time series prediction model suitable for the business case was developed, which will be implemented in a web application using Shiny R in a subsequent step.

**Research question 3:**

RQ 3*: "What are the relevant design features of an explanation interface that display the results of a sales forecast, based on the design principles of XAI and optimized for the needs and requirements of professional users in the catering industry?"*

To address the third research question, the design principles collected in the second research question are transformed into design features. This first prototype will be evaluated and further developed in a second iterative cycle by business experts in a design workshop. The results of this second iteration will be transformed into a web application using Shiny R and evaluated again by expert interviews in a third iterative cycle. The resulting feedback will again be used to further develop the prototype. The design workshop, the expert interviews, and the iterative prototyping will be used to ensure that the design features in the final prototype comply with the adapted design principles of XAI and that the prototype meets the wishes and requirements of the business users.

# 2 Related Work

This chapter serves as a theoretical basis for the topics referred to in the main body of the thesis. An overview of the current state of science in the field of XAI is given and the goals and research issues are explained. With the help of a literature review, various design principles are derived with which XAI can be implemented in theory. After prioritising the relevant principles and assessing whether they can be applied in practice, the first research question is answered. Finally, the subsequent chapter explains how time series prediction is implemented theoretically.

## 2.1 Explainable Artificial Intelligence – State of Art in Science

XAI is a research topic that has gained attention over the past decade. There are multiple international conferences and workshops (e.g. (Aha, Darell, Pazzani, et al., 2017; Diakopoulos, Friedler, Arenas, et al., 2018; Kim, Varshney, & Weller, 2018; Wilson, Kim, & Herlands, 2016) dedicated to further developing this field. It is important to note that XAI is not an independent research topic, but instead overlaps with many other branches of AI research. Adadi & Berrada (2018) propose a framework for the AI research environment where the XAI research branch can be seen as part of the AI Third Wave Movement, representing the basis and prerequisite for the future development of an artificial superintelligence.

The importance of XAI is further underlined by numerous intersections with other AI related research areas. According to Wierzynski (2018), XAI research is influenced by developments in the following research topics and seeks to contribute to solving the challenges and emerging issues henceforth:

**Bias:** It is of greatest interest to the developer of an AI system and the various stakeholders to ensure that the AI system does present bias or stereotype. These biases can be inherited from current datasets. Instead, it is expected that AI provides an objective and rational decision support for the user and does not influence his or her decision in an undesirable direction (Wierzynski, 2018, p. 28).

**Fairness:** If an AI system is used to support a decision, it must be possible for the user to determine whether the decision or recommendation is fair and based on objectivity. The result of the algorithm must be fair and equitable for minorities and all layers of society (Wierzynski, 2018, p .28).

**Safety:** The problems solved with the help of AI are so complex that not all possible solutions can always be tested and verified. Therefore, the results of an algorithm must be trustworthy and realistic without explaining exactly how the underlying system works (Doshi-Velez & Kim, 2017, p. 3.; Wierzynski, 2018, p. 28).

**Causality:** An underlying goal of any AI system should be not only to draw correct conclusions but also to explain the underlying model and the causality between input and the resulting outcome (Wierzynski, 2018, p .28).

**Transparency:** This research area explores how the workings of an algorithm can be explained in a simple and understandable way to a non-specialized audience, enabling such audience to independently evaluate how these systems affect their lives (Weller, 2017).

There exist multiple definitions of XAI within the research community, as well as its overarching goals. The study by Barredo Arrieta et al., (2019) provides a comprehensive overview of the different goals pursued with XAI for the individual stakeholders of the technology (p. 86). However, as this thesis focuses on the end-user of the model, the goals for the remaining stakeholders of XAI will not be discussed. The DARPA, which has a research project dedicated to the field describes the objective of XAI as follows: “produce more explainable models, while maintaining a high level of learning performance (prediction accuracy), as well as to enable human users to understand, appropriately, trust, and effectively manage the emerging generation of artificially intelligent partners” (Turek, n.d.). Another definition of the goals of XAI is given by the international conference FAT: “The aim of XAI is to ensure that algorithmic decisions as well as any data driving those decisions can be explained to end-users and other stakeholders in non-technical terms.” (Diakopoulos et al., 2018). In addition to the goals already mentioned, Das et al., (2020) mention another objective of XAI: “XAI promotes fairness and helps mitigate biases introduced to the AI decision either from input datasets or poor neural network architecture” (p. 5).

Based on these goals, three design requirements are derived that the proposed prototype of the thesis has to fulfil. First, by providing understandable explanations about the functionality, the explanation interface should improve the transparency of the algorithm, enabling users to understand the recommendations and results of the algorithm and thus make them accessible for their decision-making process (Das et al., 2020, p. 5). Secondly, the functions of the explanation interface are intended to increase the level of trust that users have in the algorithm in order to improve their certainty in the decision-making process (Das et al., 2020, p. 5). Third, the explanation interface should be designed in such a way that the user improves his understanding of possible errors and biases during the usage and can better assess the behaviour of the model and the results (Das et al., 2020, p. 5). If all three design requirements are satisfactorily implemented in the resulting prototype, the model will be able to optimally support the user's decision-making and the user will be able to make informed decisions, as shown in Figure 1.

To achieve this goal of a transparent AI system, various possibilities have been described by the scientific community. In the work of Guidotti et al. (2018), a taxonomy is presented that helps to characterize the different approaches of how to explain a black box algorithm (p. 11), these categories are listed in the following figure.

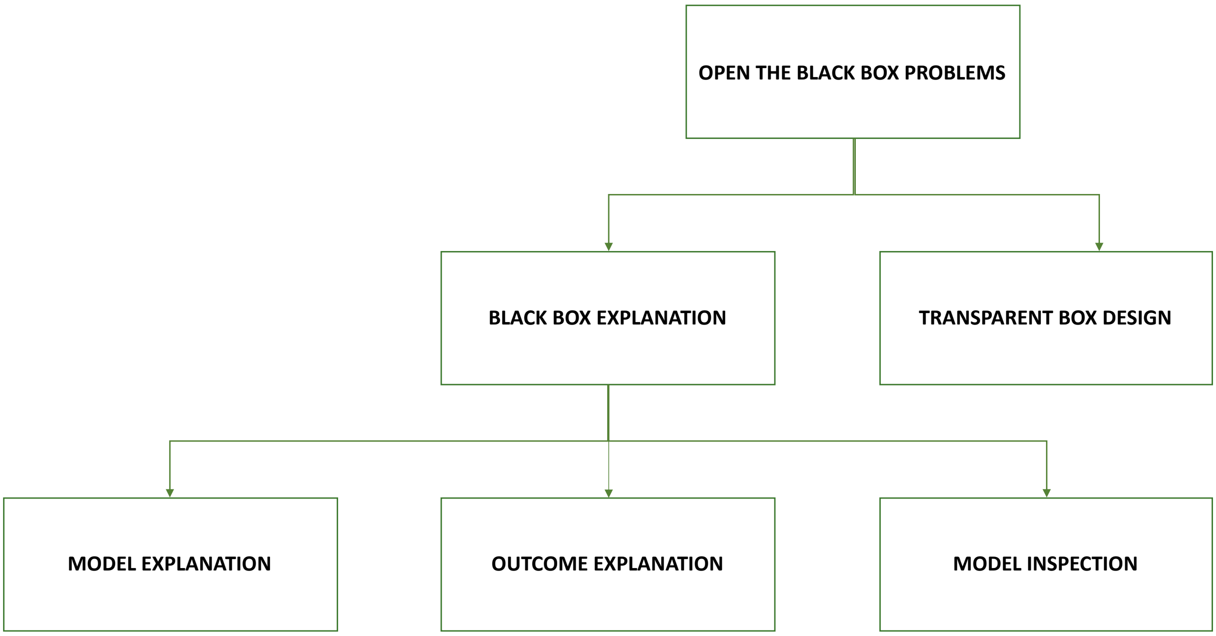



*Figure 2: Different approaches to explain a black box algorithm, according to Guidotti et. al, 2018, p. 11.*

There are two ways to solve the problem of black box algorithms: explaining the black box or developing a transparent algorithm. The first way centres on developing an interpretable model that derives insights from a data set with the help of already explainable mechanisms. In the second case, the focus turns to make the results of an inexplicable model understandable in retrospect. This can be done in three different ways: model explanation, outcome explanation or model inspection. Model Explanation aims to achieve an overall understanding of the way the model works in order to be able to better interpret the results. In outcome explanation, the black box problem is solved by explaining the results produced by the model. The focus of the model inspection is to better understand how the model works and how changing input factors affect the outcome. (Guidotti et al., 2018, p. 11).

There are several mathematical and statistical approaches that are used to implement a higher level of transparency in algorithms intrinsically or post-hoc. However, explaining them will not be the focus of this thesis. For more information, please refer to the works of Adadi & Berrada (2018), Barredo Arrieta et al. (2019), and Das et al. (2020), which deal extensively with this topic.

This thesis follows the approach of outcome explanation through an explanation interface in order to explain the results of the developed algorithm, i.e., the results of the prediction and the inner workings of the algorithm are explained post-hoc (see Figure 2). This decision reflects the results of the first round of interviews with business experts. These revealed that people working in the catering industry are not tech-savvy and that an explanation based on a technical model explanation, or a model inspection is inappropriate. All interviewees stressed that it is not crucial to them to understand the inner workings of the model, but that they need to understand the outputs of the model in order to use them to support their decision-making process. Therefore, this work does not develop a new inherently transparent model or take an intrinsic approach to explain a black box algorithm but creates an explanation interface in order to achieve the goals of the thesis (see Appendix B3 for documentation of the expert interviews).

In order to be able to pursue this approach of outcome explanation, an analysis of the current state of science was conducted. During a literature, review a number of design principles were compiled. In the analysis, a total of 22 studies were considered and 16 relevant design principles were identified. The results of the literature analysis are summarised in Appendix A.

In a subsequent step, the design principles were examined and ranked according to their relevance. The resulting design principles are described in more detail in the following table (see Table 1). These selected principles are reviewed for their applicability in practice in the following chapter and serve as a basis for the iterative development of the explanation interface later in the thesis.

| # | **Design Principle** | **Description** | **Mentions in Literature** (n = 22, multiple answers possible) |
|---|---|---|---|
| 1 | Ease of Use (incl. Consistency and Swift Execution times) | The user must not be confused by technical language or complicated expressions or visualisations. This creates resistance to integrating the tool into daily work and hinders the building of trust and credibility. The general handling must be intuitive and as simple as possible. | 35 Mentions |
| 2 | Interactivity (incl. User has control, What-if Scenarios) | Interactivity generates interest in the user to try out the application and experiment with it, increasing comprehensibility and acceptance. | 18 Mentions |
| 3 | Added Layer of Interpretability | Besides the actual visualisation of the results of the algorithms, the user needs an additional level for explanations in order to generate understanding and trust. | 17 Mentions |

| | | | |
|---|---|---|---|
| 4 | Backtracing / Reversing Steps | The user must be able to see what effects his input has. Furthermore, it must be easy for them to make adjustments to the interface, to give input and to undo their input. | 6 Mentions |
| 5 | Feature / Factor Importance (incl. Correlated Input Features) | The granularity of the information shown must be consistent with the needs of the users and it must be apparent which factors play a role in the prediction. | 5 Mentions |

*Table 1: Suggested design principles based on literature review.*

## 2.2 Explainable Artificial Intelligence – State of Art in Practice

This chapter explains the most important elements of implementing XAI in practice. Based on this information, the design principles derived in the previous chapter are reviewed for their applicability in practice and thus the first research question is answered.

The following graphic, taken from the work of Gunning and Aha (2019), presents the current framework used in practice to explain the implementation of XAI (p. 50). Comparing both the taxonomy of Guidotti et al. (2018) with the Gunning and Aha framework, it can be seen that the post-hoc explanations: "Model Explanation" and "Outcome Explanation" prevail in practice. Gunning and Aha (2019) examined eleven different American institutions with regard to their implementation of XAI and all of them focused on model explanation and outcome explanation through an explanation interface further underlining their importance (p. 55).

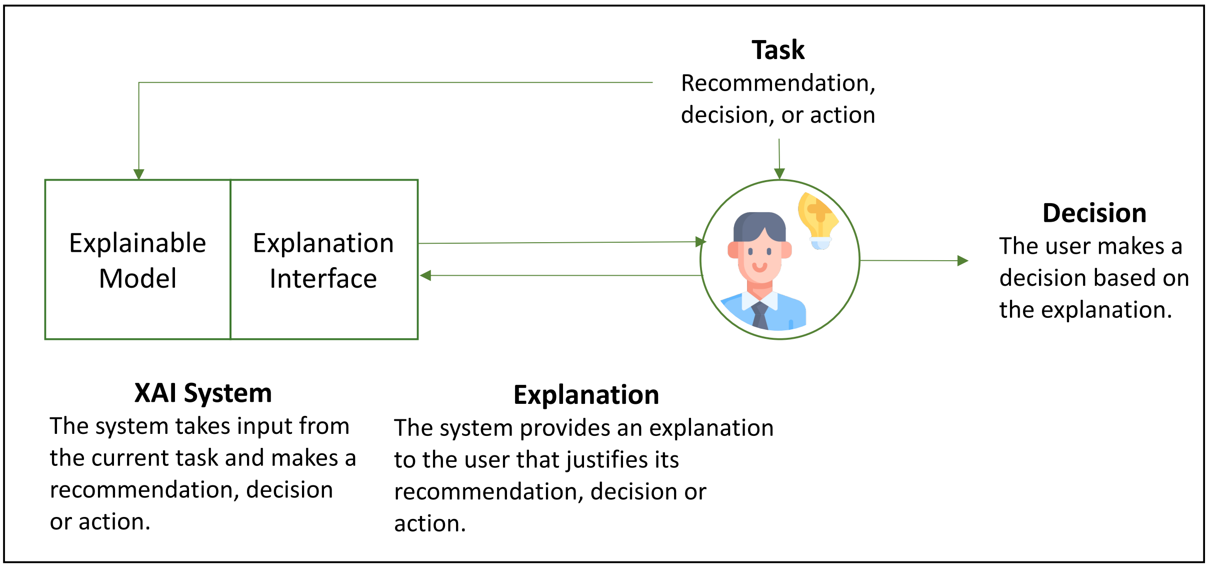



*Figure 3 :Explanation framework according to Gunning & Aha, 2019, p. 50.*

The different elements in an explanation interface and the purpose of it are described in the works of Kulesza, Burnett, Wong and Stumpf (2015), Lim, Dey and Avrahami (2009) and Smith-Renner, Rua and Colony (2019), among others. In these projects, an interactive interface is employed with the aim of explaining complex processes in algorithms to the users of the system in order to generate a high degree of transparency and user feedback (Smith-Renner et al., 2019, p. 4). The most important elements in an interactive explanation interface include the classification itself, the confidence interval, the features of the classifier that are understandable to humans and other possible classification variants (Smith-Renner et al., 2019, p. 4). However, the content of the interface must not be overloaded or too complicated, because the primary goal is to guarantee transparency and not to further overwhelm the user (Kulesza, Stumpf, Burnett, et al., 2013, pp. 8–10). Furthermore, the trust of the user tends to be negatively influenced if the confidence interval is too low (Lim & Dey, 2011b, pp. 422–424).

The structure and composition of the explanations themselves is another important factor that needs to be considered in order to develop an explanation interface that is optimally suited to the user. There are already many different research papers on this topic (e.g. Grice & Paul, 1975; Kulesza et al., 2015; Sajja, Aggarwal, Mukherjee, et al., 2020), which are influenced and shaped by other research areas such as social sciences (Miller, 2019).

There are many different guidelines on how the explanations should be designed, but in his work Kulesza et al. (2015) sets out a concise set of eight principles for explanations that are an inspiration for the present work: Explanations should be iterative, sound, complete, not overwhelming, actionable, reversible, integrate user feedback and respect incremental change. The first four principles are linked to the overarching category of explainability (pp. 127-128). In order to generate a satisfying user experience, it is essential to make predictive models interpretable. If a recommendation of a model contradicts the intuition or gut feeling of the user, the model quickly loses its reputation and is no longer used. (Sajja et al., 2020, p. 3). As the complexity and importance of a task increases, this problem is exacerbated and adequate explanations become even more important (Bunt, Lount, & Lauzon, 2012, p. 6).

The first principle states that explanations should facilitate an iterative learning process characterised by easily understood and sequentially arranged information. Second, the explanations of a model should not be over-simplistic. Optimally, they should represent the model in its complexity and not reduce it in the explanations. Furthermore, the explanations should also be "complete", i.e., the model should be depicted in its entirety in the explanations so that the user understands how different functions of the model interact with each other. The fourth principle, however, calls for a balance between too detailed and too simple explanations. The focus should always be on comprehensibility for the user; the information should never overwhelm the user. (Kulesza et al., 2015, pp. 127–128).

The last four principles from Kulesza et al. (2015) relate to the overarching category of "Correctability". Correctability describes the interaction between the user and the model, in which the user can adjust the model with the help of inputs (p. 128).

The principle of "be actionable" assumes that users will ignore explanations if they do not understand them. Therefore, the explanations must be designed in such a way that the user can benefit from them. Reversibility is another principle that encourages the user to provide more feedback and input to the model. If the user thinks that his changes could harm the prediction or the model, he will play around and try out the model less and thus slow down his learning and adaptation process (Kulesza et al., 2015, p. 128). The seventh principle is based on the research of Yang & Newman (2013), who found that users' willingness to give feedback decreases if the feedback is not integrated into the model (p. 101). The last principle is closely related to the iterative learning approach of the first principle. Users of a system are more willing to read explanations if they can improve their understanding of the prediction model (Kulesza et al., 2013, pp. 9–10).

In addition to the guidelines for explanations in the interface, there are also many scientific studies on the specifics of UI design. In addition to the "Eight Golden Rules of Interface Design" of Shneiderman, Plaisant, Cohen, Jacobs, & Elmqvist (2016) there are a large number of authors who have devoted themselves to this problem (Babich, 2019; Dudley & Kristensson, 2018; Pu, Chen, & Hu, 2012). In their work on the "Dark Patterns of UX Design", Gray, Kou, Battles, Hoggatt, & Toombs (2018) go one step further and highlight the "worst practices" in the design of user interfaces. However, due to the breadth of the research area, only the "6 Guidelines of User Interfaces" by Dudley and Kristensson (2018) are presented in this thesis, as they are most relevant to the development of an explanation interface and the general research purpose of this thesis.

| # | **Guideline** | | **Description** |
|---|---|---|---|
| 1 | Make task goals and constraints explicit | | Due to the user-driven nature of the operation of the UI, it is necessary for the user to understand their tasks, capabilities and constraints in order to utilise an effective problem-solving strategy to accomplish their task. |
| 2 | Support user understanding of model uncertainty and confidence | | It is important that the user is able to interpret the uncertainty and significance of the model so that it is possible for the user to establish which expectations the user can have of the model during operation and of the prediction result. |
| 3 | Capture intent rather than input | | Since with user input there is always a risk of uncertainty about the meaning of the input, reducing this uncertainty is important for training the model. The design of an interface can help to determine the real intent of a user. |
| 4 | Provide effective data representations | | The visualisation of data in the interface should always be oriented towards the possibilities of human perception so that the user can quickly derive information from them. |
| 5 | Exploit interactivity and promote | Interaction for Understanding | The interactivity of an interface improves the understanding of the underlying model, the functions of the interface and the used data sets. |

| | | | |
|---|---|---|---|
| | rich interactions | Make the most of the user | In order to develop a model further, users need a way to trace past inputs and see the impact of them. Furthermore, users want to have the feeling that they are contributing to the model, so it is important to include their feedback in the development. |
| 6 | Engage the user | | A sound model manages to motivate the users to complete the task and at the same time not to overwhelm them. |

*Table 2: 6 Guidelines of user interfaces according to Dudley & Kristensson (2018), pp. 22-29.*

Based on this information, the design principles that were collected in Chapter 2.1 will now be examined for their applicability in practice. For descriptions of the individual principles, please refer to Table 1.

The first design principle to be analysed, "Ease of Use", revolves around the simplicity of the provided explanations and the operation of the explanation interface, which should be characterised by a clear and concise structure. Based on the analysis of the current state of practice, this principle is applicable for the further course of the thesis. The principle is used in several research papers, for example, it is listed as one of eight principles for explanations in the work of Kulesza et al. (2015, p. 128) and is used in the work of Lim and Dey (2011, pp. 159–160) which confirms the applicability of the principle.

The second design principle, "Interactivity", which emphasises human-machine interaction, is also well documented in practice by many works. In the work of Gunning and Aha (2019), when characterising the explanation interface in research projects from different institutions, interactivity or a human-machine interaction is mentioned in five out of eleven cases (pp. 55-56). The wide use of this principle is also evident in the projects of Sajja et al. (2020), Dudley and Kristensson (2018) and Smith-Renner et al. (2019). Because of this broad application of this principle, it is applicable for further use in this thesis.

The third design principle "Added Layer of Explainability", which describes the need for visualisation of the results and an additional layer of explanation to generate understanding and trust, is a combination of the two design principles "Visualisation" and "Explanation". The need for explanation of the results of a black box algorithm is widely covered in practice and is the reason for the emergence of the research area of XAI. The principle of visualisation is also widely used in practice and is mentioned in the work of Gunning and Aha (2019) in six out of eleven analysed research projects. Furthermore, it is applied, among others, in the work of Lipton (2018), Schlegel, Arnout, El-Assady, Oelke, & Keim (2019), Sundararajan, Xu, Taly, Sayres, & Najmi (2019) and van der Maaten & Hinton (2008). Due to this broad usage in the current state of practice, the resulting design principle "Added Layer of Explainability" is appropriate for the further course of the work.

The fourth design principle, "Backtracing", which proposes the traceability of user inputs and their corresponding effects on the model, is also used extensively in practice. For example, it is part of the eight principles of explanation by Kulesza et al. (2015) and part of the six guidelines for UI interfaces by Dudley and Kristensson (2018), among others, and has also been studied in (Kapoor, Lee, Tan, & Horvitz, 2010) project, therefore the principle is applicable through the further course of the thesis.

The fifth design principle "Factor Importance", which emphasises the importance of understanding the impact of individual input factors for prediction, is not as widely supported in practice as the other four design principles, but is used in various research projects, such as the "Intelligibility Toolkit" provided by Lim and Dey (2010) and the work of Sajja et al. (2020). Since the importance of the individual input factors is crucial for explaining sales predictions, this design principle will also be considered in the further course of the work.

Based on the results of the preceding analysis, research question 1 will be answered:

RQ 1: *"What are the underlying design principles of explainable AI?"*

A variety of design principles exist with which XAI can be implemented (see Appendix A). From this multitude, however, only five design principles were identified, which are used further in this work. The underlying design principles of XAI employed consist of the following design principles: 1) Ease of Use, 2) Interactivity, 3) Added Layer of Explainability, 3) Backtracing / Reversing Steps and 5) Factor / Feature Importance, since the applicability of all five proposed design principles has been confirmed in practice by various studies.

## 2.3 Explainable Artificial Intelligence – Time Series Forecasting

The term forecasting describes a prediction of a certain event in the future. It represents a task that is used every day in many areas of activity in our lives. In this context, many of the forecasts utilise the so-called time series as the basis for the predictions. A time series is a set of data points that have been gathered in relation to a variable in a time sequence (Montgomery, Jennings, & Kulahci, 2015, pp. 1–2). A common example of a time series, which is also used in this thesis, is a transactional dataset, which is created for example by a cash register system in a restaurant or supermarket. For each transaction, the system records the time stamp, which products were purchased, and the quantity of products purchased. Based on this dataset, it is possible to determine the demand for individual products at a specific point in time, i.e., at a specific time or a specific day, or an entire time series, i.e., a week, a month or a year.

Predictions can be conducted in two possible ways. The qualitative forecast, which tries to predict the future in a subjective way without historical data, and the quantitative forecast, which uses historical data in a forecasting model to predict a specific future event. To make a prediction, the model searches the data set for certain patterns and describes the statistical relationships between the current and past values. This means that the model infers future values from past and current values (Montgomery et al., 2015, pp. 4-5).

There are many different types of predictive models, but regression models, smoothing models and general time series models are currently the most commonly used prediction techniques. Regression models use the relationships and interactions between the variable to be predicted and several related variables to make a prediction (Montgomery et al., 2015, p. 5). For example, Ashenfelter's prediction model uses several variables, such as temperature in the growing season or the amount of rainfall in winter, to predict the quality of future wine (Corsi & Ashenfelter, 2019). The smoothing models use a simple statistical function in which they determine the future value of a particular variable based on a past observation. The time series models use the statistical properties of the underlying historical data to create a predictive model, which in turn determines the future value using the statistical method of least squares (Charnes et al., 1974; Montgomery et al., 2015, p. 5).

# 3 Method & Data

In this chapter, the theoretical foundations of the methods and techniques employed in the thesis are explained and their application justified. First, the structure of the research design is explained, as well as the research gap and the design science approach used. In the following chapters it is explained why the programming language R was used in this context and an analysis of the employed data set is conducted. The subsequent chapter explains the framework that was used to structure the time series prediction. In the last chapter, the practical implementation of this process is explained, and the results of the prediction are presented.

## 3.1 Research Methodology

As shown in Chapter 2, XAI is a topic that is already being researched extensively. For example, a review of the current state of research is given by Adadi & Berrada (2018) while Barredo Arrieta et al. (2020) and Gunning and Aha (2019), among others, provide an introduction into the field. Furthermore, there are research papers that examine XAI regarding a specific industry, such as Holzinger et al. (2017) for the medical sector, Sajja et al. (2020) for the fashion retail sector, or more specifically Schlegel et al. (2019) for time series forecasting. However, the current design principles of XAI focus on an overarching explanation of the black box of an algorithm and not on how the results of a prediction can be specifically explained in a way that is understandable to people outside the field to enable them to trust the predictions and integrate them into their daily work. In the current literature, there is no research on the extent to which the requirements of XAI can be transformed into application-related design principles and how these principles can be used to provide added value to end users in the catering industry.

This research gap is addressed in this paper in the following way. First, the existing design principles on XAI in current research are compiled and sorted according to their relevance, thus answering the first research question. In a second step, these design principles are verified and adapted with the help of expert interviews, thus answering the second research question. In the last step, an explanation interface for end users in the catering industry is developed and evaluated in an iterative process on the basis of the adapted design principles. With its help, users should be able to interpret the results of a time series prediction in a comfortable and intuitive way. The resulting design features are compiled and sorted according to their relevance in order to be able to answer the last research question.

To aid in filling these research gaps, a Design Science Research (DSR) Project was initiated according to the work of Hevner, March, Park, & Ram (2004) and March & Smith (1995). According to the theory in Design Science, practical problems can be solved by software prototypes i.e., with artifacts that possess theoretical and structural knowledge. A main feature of DSR, the combination of theoretical and practical knowledge, is also employed in this thesis

(Hevner et al., 2004). General research on DSR concludes that such artifacts should be developed in a process that follows a continuous process of incremental improvement and evaluation (Hevner et al., 2004; Takeda, Veerkamp, & Yoshikawa, 1990). This design science process has been further developed through several papers (Kuechler & Vaishnavi, 2008; Peffers, Tuunanen, Rothenberger, & Chatterjee, 2007).

The research design of this paper is also based on Kuechler and Vaishnavi (2008) suggested framework. The framework fits well to the context of the present thesis due to three reasons. The first phase in Kuechler and Vaishnavi (2008) process, "problem awareness," is characterized by the linkage of theoretical knowledge from the literature and practical knowledge gathered through expert interviews. Following this logic, the design principles used in this work are first derived from theory and then further developed with the help of the experts. A second emphasis is placed on the joint prototyping and evaluation of the artifact with future users. The resulting prototype of this work was developed together with the experts and improved by their feedback. Last but not least, this work also relies on the iterative approach proposed by Kuechler and Vaishnavi (2008) and (Peffers et al., 2007) as the artifact is improved through three rounds of evaluation.

***Employed Design Science Process***

After establishing the theoretical foundations of DSR and why it is used in this thesis, the employed design process is explained. It consists out of one overarching design cycle with six different phases and three iterations of prototyping. The following illustration shows the different phases of the design cycle and the corresponding outcomes.

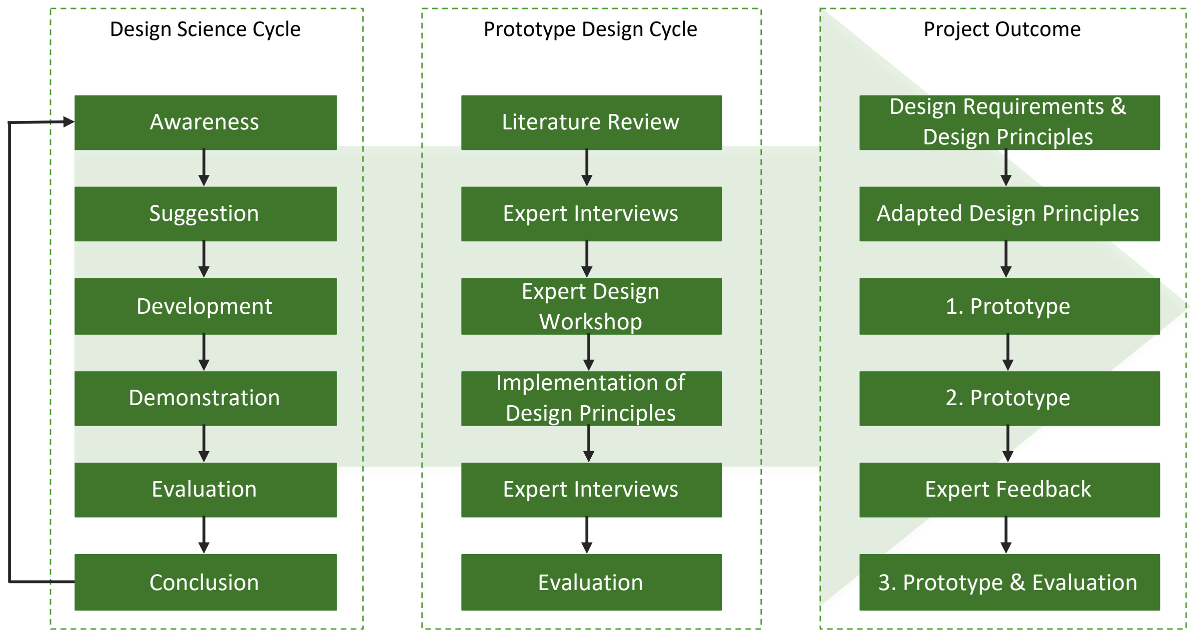



*Figure 4: Employed design science approach based on Kuechler and Vaishnavi (2008) and Meth, Mueller and Maedche (2015).*

First, the design requirements are defined based on the goals of XAI. Additionally, Design principles are derived and reviewed for applicability. In the second phase, the needs and requirements of business users are collected in expert interviews. The findings are analysed, compared and combined with the results from the literature review in order to create a set of adapted design principles that are suitable for the specific needs and requirements of the catering industry. The adapted design principles are utilized to develop a prototype that fulfils the underlying criteria of XAI. In order to develop this artifact, a time series-based prediction model is programmed. The results of this model serve as a numerical backbone in the artifact and are utilized for testing and illustration purposes in order to make it easier for the users to experiment with and comprehend the artifact's functions.

The design and structure of the prototype is developed in an iterative learning and development process during steps three and four. The first cycle of prototyping consists of creating several graphical elements, based on inputs from the expert interviews and various best practices. In a subsequent design workshop, which followed the principles of design thinking, the structure, functions and design of the visualisation are further developed and evaluated with the help of experts. Based on the results of the design workshop and incorporating the adapted design principles, the results from the time series prediction are employed in a web application utilizing the Shiny package within the R programming language. In the evaluation step, the resulting prototype is reviewed and further developed by conducting five evaluation interviews with business experts. This approach ensures that all the incorporated functions, the design and the handling are adapted to the needs of the business users and that the web application with its graphic elements meet the criteria of XAI. Eventually, the application is constructed in such a way that the comprehensibility for persons outside the field is ensured. Furthermore, the design of the application is focused on presenting the results of the prediction to the user in such a way that they are considered realistic, trustworthy and credible as well as intuitively understandable.

In this thesis, the terms design requirements, design principles, and design features are defined based on the research of (Meth et al., 2015). The design requirements correspond to general conditions that should be achieved by each artifact, in accordance with the previously defined objective of the artifact (Meth et al., 2015, p. 807). The gathered design principles refer to the general operational abilities of the developed artifact based on the previously established design requirements (Meth et al., 2015, p. 807). During the prototyping these design principles were matched to their corresponding design features in the implementation of the artifact, i.e., the web application. Those represent a possible way of translating the abstract design principles into tangible elements of the artifact fulfilling the predefined objective of the artifact (Meth et al., 2015, p. 807).

## 3.2 R Language Introduction

In this thesis the programming language R, along with its extensive range of libraries, was utilized to create the time series forecasting model and the webpage, in which the results of the prediction and the adapted design principles are incorporated. R is a free and open-source programming language for statistical calculations and graphics. A special feature of R is the numerous packages or libraries available online, containing additional functions to fulfil various tasks from different areas (Ihaka, 1998; Ihaka & Gentleman, 1996). This programming language was employed due to the fact, that it has extensive infrastructure for a time series analysis in the base version of R (Hyndman, 2021) as well as useful open source libraries like the stats package (R Core Team, 2021) and the forecast package (Hyndman & Khandakar, 2008).

***Employed Methods and packages***

Considering that the performance of the forecasting model has no real influence on the extent to which the results of the model are perceived as explainable and interpretable, a relatively straightforward configuration of the model was employed without further adapting and fine-tuning the performance of the algorithm with respect to the present task. The predictive model, which is based on the lecture of Blohm (2020), essentially consists of two parts. A Random Forest algorithm employed with the "randomForest" package (Liaw & Wiener, 2018) and the "stats" package (R Core Team, 2021) that performs the prediction.

According to the work of Ho (1995) and Breiman (2001), the Random Forest is a machine learning algorithm that is based on the ensemble learning technique, i.e., multiple decision trees are used to derive the output. It is employed in order to solve problems with the help of regressions and classifications. Due to its versatility and simplicity of operation, the algorithm is popular in the data science community and is extensively employed to resolve a variety of different data science problems. Other advantages include: it performs well with a small training set, it can handle many different input variables and complex data structures, and its performance can be adapted through changing a small number of parameters (Biau & Scornet, 2016; Scornet, Biau, & Vert, 2015). The challenge of using a Random Forest is that it provides high accuracy but low interpretability of the prediction. As the forecast is calculated via complex, multilinear relationships, it can only deliver a statement about the importance of the individual input factors. This classifies the prediction technique as a black box model and is another reason why the Random Forest algorithm was chosen in this thesis. (Vidal & Schiffer, 2020). For more extensive details on the method and operation of Random Forest algorithms, reference the work of Ho (1995) and Breiman (2001).

The stats package serves as a set of tools "for statistical calculations and random number generation" (R Core Team, 2021, p. 1). More specifically, the predict function is used to derive predictions from the results of previously trained models (Chambers & Hastie, 1992).

To obtain the forecast for the forecast horizon, the Random Forest was employed for regression. For this purpose, the R package "randomForest" was used analogously to the instructions in Liaw & Wiener (2002). The Random Forest uses a regression function to train a model from the different input variables using the training set (Tyralis & Papacharalampous, 2017). This model is then used with the help of the predict function from the stats package to generate the forecasted value based on the test set. In order to programme the whole algorithm besides the base R packages, several open-source packages were utilized to simplify the data handling and modelling. All the employed packages can be found in the following table.

| **Package name** | **Description** | **Source** |
|---|---|---|
| randomForest | This package was used to create a regression based on a Random Forest algorithm to train and fit the forecasting model. | Liaw & Wiener, 2018 |
| forecast | This package was employed to decompose the time series into the seasonal components. | Hyndman, Athanasopoulos, Bergmeir, et al., 2021 |
| data.table | This package was used to simplify the handling and manipulation of large data sets. | Dowle, Srinivasan, Gorecki, et al., 2021 |
| ggplot2 | This package was used to simplify the configuration and generation of plotting data. | W. Wickham, Henry, Pedersen, et al., 2020 |
| dplyr | This package was used to simplify the handling and manipulation of data frames. | H. Wickham, Francois, Henry, & Mueller, 2021 |
| stringr | This package was used to simplify the handling and manipulation of strings in R. | H. Wickham, 2019 |

*Table 3: Employed R packages.*

### 3.3 Employed Dataset

In this chapter, the data set used for time series forecasting is explained in more detail. One dataset with transactional data from a small bakery was used to perform the prediction. It can be accessed in the dedicated GitHub repository following the link provided in Appendix E. In Figure 5 the original data set is shown without any pre-processing or data modelling. The data set is comprised of four variables regarding the transactions made in the bakery. The “Date” variable shows on which date the transaction was done, it ranges from the 30$^{th}$ of October 2018 to the 9$^{th}$ of

| | Date | Time | Transaction | Item |
|---|---|---|---|---|
| 1 | 2018-10-30 | 09:58:11 | 1 | Bread |
| 2 | 2018-10-30 | 10:05:34 | 2 | Scandinavian |
| 3 | 2018-10-30 | 10:05:34 | 2 | Scandinavian |
| 4 | 2018-10-30 | 10:07:57 | 3 | Hot chocolate |
| 5 | 2018-10-30 | 10:07:57 | 3 | Jam |
| 6 | 2018-10-30 | 10:07:57 | 3 | Cookies |
| 7 | 2018-10-30 | 10:08:41 | 4 | Muffin |
| 8 | 2018-10-30 | 10:13:03 | 5 | Coffee |
| 9 | 2018-10-30 | 10:13:03 | 5 | Pastry |
| 10 | 2018-10-30 | 10:13:03 | 5 | Bread |

*Figure 5: Original configuration of the data set.*

April 2019. These dates indicate that the bakery was open every day except the 25$^{th}$ and 26$^{th}$ of December and the 1$^{st}$ and 2$^{nd}$ of January. In this time frame there were made a total of 21’293 sales. The “time” variable shows the exact time when the sale was made. Due to our interest was only to predict daily sales; this variable was excluded in the further work. The “transaction” variable indicates the number of transactions made. Given that one transaction could include multiple items this variable cannot serve as a unique identifier for the data set. The variable "item" indicates the type of product that was sold in the different transactions. In total, there were sales of 95 different products, but due to an item category called "None", it can be inferred that all transactions with this product must be some kind of input error and therefore were excluded from the data set. Therefore, 94 relevant product categories with a total sale of 20’507 were analysed. The data set can be accessed on the GitHub repository dedicated to this thesis.

In the following figure, the characteristics of the given data set are explored in further detail. According to graph A in Figure 6, the total sales are relatively stable in the given time frame. Of the top 10 products, it can be deduced that sales are largely made up of the same products, with the five best-selling products accounting for a total of almost 60% of all sales. From the distribution of sales, it can be derived that most sales are made on Monday, Wednesday is the lowest weekday in terms of sales. Based on the monthly distribution, November was the best sales month. After a slump in December, the sales figures increased until March, but remain clearly below the November level. The months of October and April are not represented in the data set, which complicates a proper interpretation.

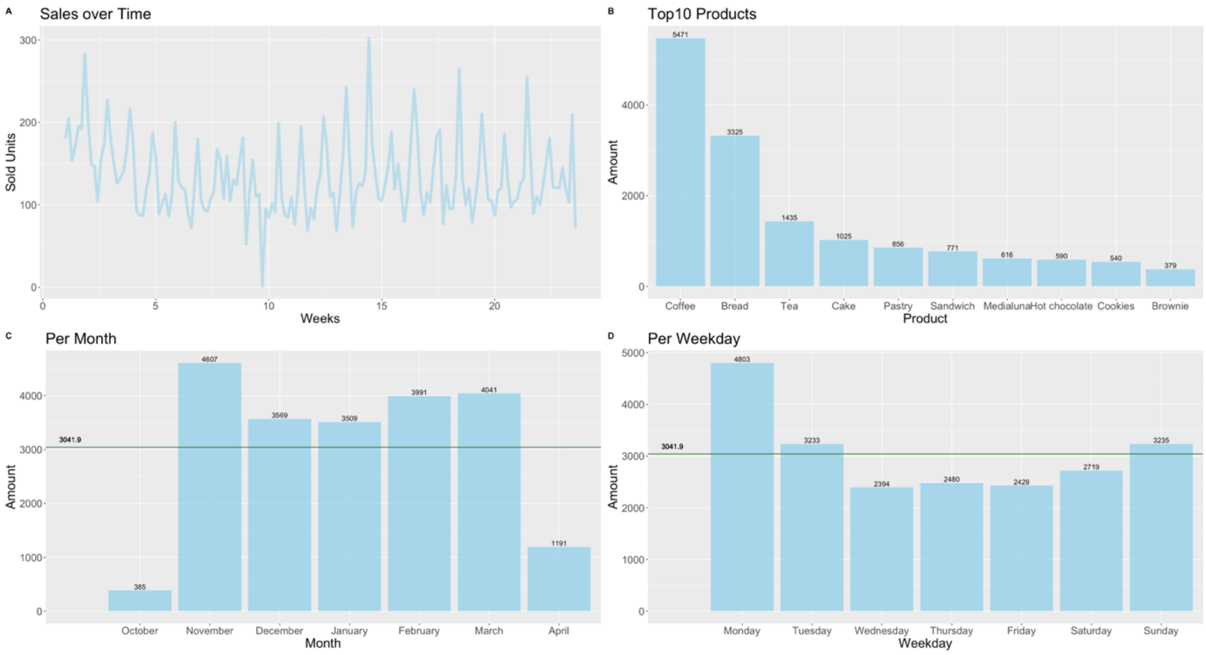



*Figure 6: Summary of the given data set.*

Due to the limited data set, it was not possible to make statements about all transactions in the period under consideration with the decompose function of the forecast package. Therefore, only the time series of coffee sales is analysed in the following figure. This product is the bakery's most important product and represents 27% of total sales.

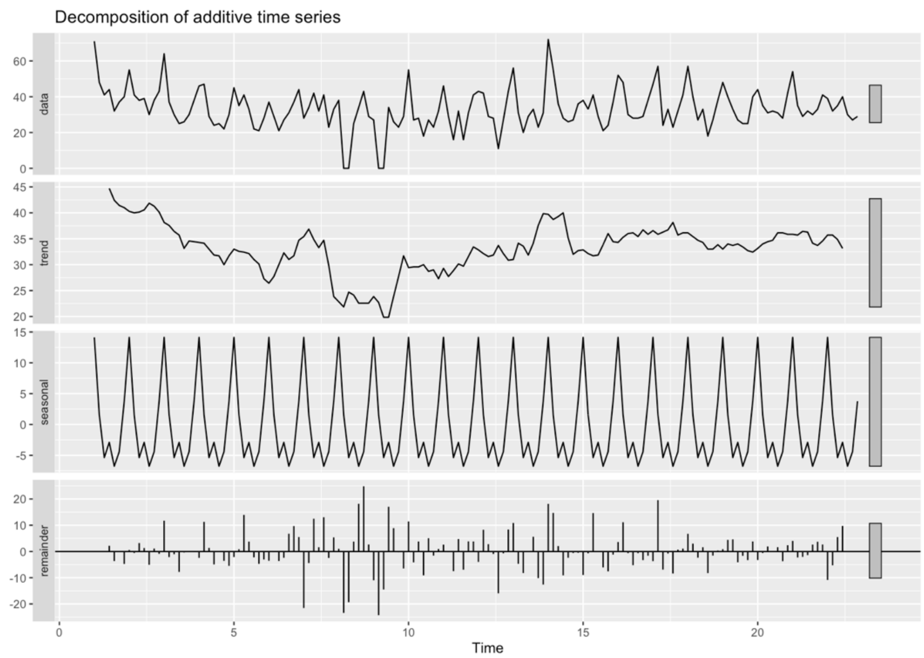



*Figure 7: Decomposition of the coffee time series.*

Analysing the decomposed time series of coffee show that sales figures drop directly at the beginning and reach their lowest level in week 14, after which a recovery of the sales figures takes place, although the figures do not reach the original volume. The coffee data set exhibits a seasonal component. This is due to the high sales figures on Monday and Sunday, which recur weekly. The residual, or noise, is high and characterised by large fluctuations, especially in weeks 5-10, which can be an indicator of many incorrect entries. This amount of noise can negatively affect the forecast.

In the analysis of the data set, it becomes evident that it is suitable for time series forecasting in principle, but that it contains some limitations that can negatively influence the result of the forecast. Firstly, the data set only covers just under seven months; for a meaningful forecast, a data set of up to two years would be optimal. Secondly, the sample is not well balanced, i.e., a large proportion of sales are accounted for by a small proportion of products and there are many products that are sold infrequently. In addition, there are many misclassified entries in the "None" category, which reduces the significance even further. Finally, the data set consists of individual products and not complete menus as they are usually sold in the system gastronomy. This could lead to irritations of the business users during the implementation and evaluation of the web application.

### 3.4 Underlying Process of the Time Series Forecast

After getting to know the employed data set and the principles of time series forecasting it is important to understand the theoretical foundations of the forecasting process utilized in this work. In this thesis, the forecasting process according to (Shmueli, Kenneth, & Lichtendahl, 2016, p. 17) is employed to ensure a structured and goal-oriented forecasting model, which is illustrated in Figure 8.

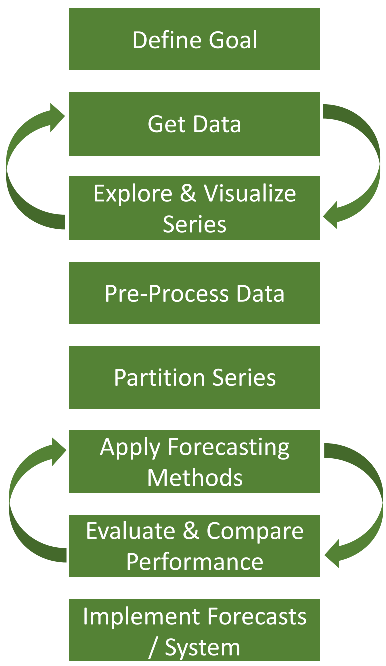


*Figure 8: Forecasting process, based on Shmueli et. al., 2016*

As in other data analysis tasks, the process is initiated by defining the goal the forecast should achieve. This stage in the process is of great importance, as different forecasting techniques are used depending on the goal and how the forecast will be utilised. In this context, it is distinguished between descriptive and predictive goals. In descriptive or time series analysis, the data of a time series is modelled according to different properties such as trends or seasonal characteristics. These can then be used for strategic planning or decision-making. If a predictive goal is pursued, i.e., in time series forecasting, then the predictive model uses the inherent information in the time series to determine a future value or event (Shmueli et al., 2016, pp. 18–19). Furthermore, it is central to determine the forecast horizon and the forecast interval. The first parameter indicates the time span of the value to be predicted, i.e., how far into the future the forecast model is required to predict. The forecast interval determines the frequency with which new forecasts are made. This is particularly important when it is essential to integrate new information and figures into a forecast in order to achieve optimal results (Montgomery et al., 2015, p. 6).

The second phase "Data Collection" is about guaranteeing the quality of the compiled data. Missing values, corrupted data or entry errors can have a strong negative impact on the prediction accuracy (Shmueli et al., 2016, p. 25). The importance of the first phase is also evident here. Depending on the goal of the forecast, the requirements for granularity, i.e., the extent to which the data set covers temporal, geographic or demographic aspects, to name but a few, may change. If the goal is to predict the daily sales figures of a restaurant, the forecasting model needs a granularity on a daily or hourly basis to be able to deliver adequate results. If data with too low a granularity is utilised instead, the results will not be optimal either. A data set for example on a per-second basis may contain too many irrelevant events having a negative impact on the prediction result (Shmueli et al., 2016, pp. 26-27).

In order to be able to select the optimal prediction model and the optimal data set, it is advisable to split a time series into its systematic and non-systematic parts. The systematic elements are distinguished between level, trend and seasonality, the non-systematic elements are referred to as noise, i.e., random variances that come from measurement errors or other sources. Level refers to the average value of a time series, trend refers to the difference in the time series from one season to the next and seasonality describes the short-term behaviour of the time series, which can be observed several times in the time series (Shmueli et al., 2016, p. 28). It is important to note that all time series have the component level, but do not necessarily have a value for trend and seasonality. It should also be noted that the systematic elements can never be observed separately from noise. This is where the various forecasting techniques come in and try to measure the systemic part of the time series separately and determine the noise component. The systematic part is then used to determine the future value and the non-systematic part is needed to quantify the uncertainty of the prediction (Shmueli et al., 2016, p. 28).

The goal of the next step "Data Exploration and Visualisation" is to use visual analytics to derive problems or initial insights from the data set. In this way, approximations can be made to the values of level, trend and seasonality (Shmueli et al., 2016, pp. 30-39).

In the following step "Data Pre-processing", the data set is prepared for the analysis and optimised for it. The individual tasks in this step include the elimination of missing values, the adjustment of intervals in the time series and the correction of extreme outliers. Missing values can generate a hole in the time series. Such holes pose a problem for many forecasting techniques and negatively affect the validity and accuracy of the results. If irregular intervals are present in a time series, there might be a change in the observation period, i.e., the observation intervals are not the same; this is the case if, for example, not only daily values but also annual values are present in a time series. Extreme outliers in the data set can weaken the predictive power of a model. However, prior to removing such extreme values, it is necessary to assess whether the value was caused by an entry error, or by a rare event or other reasons. This is required to assess whether it is justified to delete this entry (Shmueli et al., 2016, pp. 39-41).

An important subsequent step to enable the evaluation of the forecast is the partitioning of the data set into a test and training set. The algorithm is trained on the training set and its performance is scrutinized on the test set. This measure was introduced to prevent the algorithm from optimising and adapting itself too much on the training set and overfitting on this specific data set. An algorithm suffering from overfitting will perform well on the dataset it is trained on but will not perform well on new datasets (Shmueli et al., 2016, p. 45).

In a time series forecast, the test set always consists of as many values that correspond to the forecast horizon. This ensures that the performance of the actual forecast can be checked. If the test set is reduced, it is impossible to evaluate the exact performance of the model regarding the forecast horizon. However, if the test set is unnecessarily increased, there is a risk that the training set will be too small and contain too little information for training the model, which in turn will have a negative impact on performance as well (Shmueli et al., 2016, p. 48).

In the last phase, the finalised forecast model is implemented in the target application in order to fulfil the goal set in step one and to achieve the desired benefit.

## 3.5 Results of the Time Series Forecasting Model

Based on the underlying process explained in Chapter 3.4 and the data set analysed in Chapter 3.3, the following chapter explains the theoretical structure, practical sequence and results of the time series prediction model. The entire code of the time series prediction and the implementation in Shiny can be found in the GitHub repository dedicated to this master thesis (Appendix E). The following figure illustrates all the core elements of the time series prediction model.

| | | |
|---|---|---|
| Goal | • Support decision-making in catering industry<br>• Prediction of food sales based on historical data<br>• Forecast Horizon: 28 days | |
| Data | • Transactional data set from a small bakery<br>• 20'507 total observations, 95 different products<br>• Time frame: 30th of October to 9th of April | |
| Time series | • Creation of even weeks, starting with Monday ending with Sunday<br>• Resulting in 23 weeks usable for the prediction | |
| Additional features | • Is closed: indicates if the bakery is open or closed on the given date<br>• Weekday and number of week: indicates the specific weekday and number of the week on the given date<br>• Seasonality: indicates the seasonal implications on the given product<br>• Time lags 7-14: indicate the amount of sold products 7-14 days earlier | |
| Data Partition | • Split into train set and test set: forecast horizon of 28 days resulting in 4 weeks for the test set | |
| Prediction | **Individual Products**<br>1. Creation of data.table with all features and products<br>2. Run a for-loop to create the data table to predict the individual items with a random forest<br>3. Within the for loop, employment of a second for loop to predict the values for weeks 11-14 with stats package<br>4. Collection of the actual and predicted values in data table | **Total Sales**<br>1. Creation of data table with all features and total sales<br>2. Training of the random forest<br>3. Run a for-loop to predict the values for weeks 11-14 of total sales with stats package<br>4. Collection of the actual and predicted values in data table |
| Preparation for Shiny Implementation | • Creation of different data tables to show actual and predicted values for individual and total sales on weekly and monthly basis<br>• Plotting of graphs to visualize actual and predicted values for individual and total sales on weekly and monthly basis | |

*Figure 9: Summary of key elements of time series prediction.*

The first phase of the process focuses on the goal definition. The overall objective of this forecast is to predict food sales based on historical transactional data for the catering industry. The purpose is to support the business users in this field in their daily decision-making processes and to make the ordering process of this industry more efficient, more cost-saving and less wasteful. In order to generate the greatest possible benefit for the business users, the forecast horizon was set at 28 days. This serves to enable the most forward-looking planning possible and to create a week-by-week overview of future menu plans. Therefore, the forecast must be easy to handle, as it is recalculated week by week. Furthermore, the data integration must also be quick and easy, as the most recent historical data should be integrated swiftly. In total, two forecasts were programmed, one for the individual products and one for the total sales in the specified period. Steps two and three, i.e., data collection and data analysis, will not be dealt with in detail, as this was done in Chapter 3.3.

In the pre-processing phase, the data set was changed so that the mapped weeks always start with Monday and end with Sunday. This results in 23 weeks that can be utilized for the forecast. In addition, several new features were created to optimise the quality of the prediction. These were applied to both the individual forecast and the total sales forecast. Firstly, an "is closed" feature was added to indicate whether the bakery was open or closed on a given date. To develop this feature, the data set was searched by day and the days on which sales were zero were marked as "closed". Furthermore, weekdays and calendar weeks were specified for the given time series and integrated into the data set. With the help of the forecast package, the time series was analysed with the decompose function and broken down into its individual elements. Using this function, a frequency of seven was employed, i.e., it was searched for a week-based seasonality element. It should be noted that a component of seasonality could be found for each individual product, but not for the whole dataset. However, in order to integrate a seasonal component into the data set for total sales, the component from each individual product was utilized and integrated cumulatively into the data set for total sales. Finally, so-called "time lags" were integrated into the data set; they look back from a certain date to a specified number of days in the past and indicate the sales figures on this date. In this work, the seven time lags spanning from lag7 to lag14 were used, which means that at any point in time in the time series, there is a 7-14 day look back in time to compare the current sales value to past values. The following figure shows the final configuration of the data set used for the prediction of individual sales, in this example for the product "Scandinavian".

| Id | 19 sales | is_closed | day_of_week | week_of_year | quantity_lag7 | quantity_lag8 | quantity_lag9 | quantity_lag10 | quantity_lag11 | quantity_lag12 | quantity_lag13 | quantity_lag14 | seasonal |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 4 | 0 | 5 | 47 | 8 | 9 | 2 | 1 | 4 | 10 | 0 | 14 | 3.363011871 |
| 2 | 3 | 0 | 1 | 47 | 1 | 8 | 9 | 2 | 1 | 4 | 10 | 0 | -0.922702414 |
| 3 | 1 | 0 | 4 | 47 | 1 | 1 | 8 | 9 | 2 | 1 | 4 | 10 | -1.149593171 |
| 4 | 3 | 0 | 2 | 47 | 5 | 1 | 1 | 8 | 9 | 2 | 1 | 4 | -0.052020808 |
| 5 | 0 | 0 | 3 | 47 | 5 | 5 | 1 | 1 | 8 | 9 | 2 | 1 | -0.813458717 |
| 6 | 0 | 0 | 6 | 47 | 5 | 5 | 5 | 1 | 1 | 8 | 9 | 2 | -0.872282246 |
| 7 | 0 | 0 | 7 | 48 | 2 | 5 | 5 | 5 | 1 | 1 | 8 | 9 | 0.447045485 |
| 8 | 9 | 0 | 5 | 48 | 4 | 2 | 5 | 5 | 5 | 1 | 1 | 8 | 3.363011871 |
| 9 | 0 | 0 | 1 | 48 | 3 | 4 | 2 | 5 | 5 | 5 | 1 | 1 | -0.922702414 |
| 10 | 0 | 0 | 4 | 48 | 1 | 3 | 4 | 2 | 5 | 5 | 5 | 1 | -1.149593171 |
| 11 | 2 | 0 | 2 | 48 | 3 | 1 | 3 | 4 | 2 | 5 | 5 | 5 | -0.052020808 |
| 12 | 0 | 0 | 3 | 48 | 0 | 3 | 1 | 3 | 4 | 2 | 5 | 5 | -0.813458717 |
| 13 | 1 | 0 | 6 | 48 | 0 | 0 | 3 | 1 | 3 | 4 | 2 | 5 | -0.872282246 |
| 14 | 1 | 0 | 7 | 49 | 0 | 0 | 0 | 3 | 1 | 3 | 4 | 2 | 0.447045485 |
| 15 | 4 | 0 | 5 | 49 | 9 | 0 | 0 | 0 | 3 | 1 | 3 | 4 | 3.363011871 |
| 16 | 0 | 0 | 1 | 49 | 0 | 9 | 0 | 0 | 0 | 3 | 1 | 3 | -0.922702414 |
| 17 | 0 | 0 | 4 | 49 | 0 | 0 | 9 | 0 | 0 | 0 | 3 | 1 | -1.149593171 |
| 18 | 0 | 0 | 2 | 49 | 2 | 0 | 0 | 9 | 0 | 0 | 0 | 3 | -0.052020808 |
| 19 | 0 | 0 | 3 | 49 | 0 | 2 | 0 | 0 | 9 | 0 | 0 | 0 | -0.813458717 |
| 20 | 0 | 0 | 6 | 49 | 1 | 0 | 2 | 0 | 0 | 9 | 0 | 0 | -0.872282246 |

*Figure 10: Final configuration of the dataset prior forecasting.*

The prediction for the individual products was done using two nested for-loops. The first was designed to perform five tasks. First, a data set was created for each of the 95 products, containing only the relevant data for the specific product. This was then split into a test set and a training set, with the test set corresponding to the size of the forecast horizon, i.e., exactly four weeks. In a next step, the corresponding features for seasonality and time lags were integrated and then the Random Forest algorithm was trained on the training set and the auxiliary vehicles in order to collect information during the second for-loop were set up. The result of the first for-loop is a Random Forest model trained on the training set, which is then used in the second for-loop to predict the values within the forecast horizon of the individual product with the help of the predict function of the stats package. This was done on a week-by-week basis, i.e., one week into the future was predicted at a time. After the forecast, the data of this already predicted week was integrated into the data set and based on this, the next week was predicted. This second for-loop was carried out a total of four times for each product. With the auxiliary vehicles from the first for-loop, information on the predicted values, the actual values and the performance of the prediction are collected for each product so that it can be analysed later.

In total, these two for-loops are run through once for each individual product, so a total of 95 Random Forest models are calculated and predictions made. The following figure shows the result of the prediction. For each individual product, the four predicted weeks with the corresponding actual values are shown. The section only shows the results of the predictions for the products "Bread" and "Scandinavian".

| | Mo | Tue | Wed | Thur | Fri | Sat | Sun | Goods |
|---|---|---|---|---|---|---|---|---|
| bread_predicted_KW11 | 29.2816533333334 | 17.7888633333334 | 17.793999047619 | 18.3134983333333 | 17.37516 | 17.2078583333334 | 23.6079816666666 | Bread |
| bread_predicted_KW12 | 29.65143 | 17.0117283333334 | 18.163815 | 16.1418933333333 | 18.2368266666666 | 18.4748816666667 | 22.2471216666666 | Bread |
| bread_predicted_KW13 | 30.6728933333334 | 20.4689916666666 | 17.5304749999999 | 16.2561233333333 | 18.346515 | 19.47585 | 20.9990383333333 | Bread |
| bread_predicted_KW14 | 30.0702883333333 | 19.7478216666667 | 18.0663873809524 | 16.0054366666667 | 17.4843216666666 | 19.1355116666666 | 22.716715 | Bread |
| bread_actual_KW11 | 30 | 18 | 10 | 12 | 13 | 22 | 13 | Bread |
| bread_actual_KW12 | 28 | 18 | 13 | 17 | 17 | 23 | 20 | Bread |
| bread_actual_KW13 | 40 | 26 | 7 | 8 | 15 | 13 | 27 | Bread |
| bread_actual_KW14 | 36 | 17 | 15 | 18 | 20 | 16 | 15 | Bread |
| Scandinavian_predicted_KW11 | 3.51137739177488 | 0.648326904761904 | 0.340577380952381 | 1.19547205627706 | 0.324031587301587 | 0.295610360750361 | 1.2511458968809 | Scandinavian |
| Scandinavian_predicted_KW12 | 5.78136178571429 | 0.580305238095238 | 0.142480216450217 | 0.817241818181815 | 0.512068690476189 | 0.234691551226552 | 1.02137708735709 | Scandinavian |
| Scandinavian_predicted_KW13 | 4.2101494047619 | 0.611046904761904 | 0.254070216450217 | 0.633374437229433 | 0.289702142857142 | 0.401013019480519 | 0.910633634976136 | Scandinavian |
| Scandinavian_predicted_KW14 | 3.25326155844155 | 0.245032142857143 | 0.399565 | 1.07374027056277 | 0.27222992063492 | 0.306959765512266 | 1.13764728576978 | Scandinavian |
| Scandinavian_actual_KW11 | 3 | 0 | 0 | 0 | 0 | 1 | 1 | Scandinavian |
| Scandinavian_actual_KW12 | 2 | 0 | 0 | 1 | 1 | 0 | 2 | Scandinavian |
| Scandinavian_actual_KW13 | 7 | 0 | 0 | 0 | 0 | 0 | 0 | Scandinavian |
| Scandinavian_actual_KW14 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | Scandinavian |

*Figure 11: Excerpt from the forecast results for individual products.*

The prediction for the entire sales is based on the prediction for the individual products. Since this problem does not require a large number of individual predictions to be made, but only one prediction regarding the cumulative data set, this prediction is less complex and only requires one for-loop. In a preliminary step, a data set is created with all relevant features and data points for the prediction. The "is closed" feature and the days of the week or weeks of the year could be taken from the previous forecasting model. The time lags and the seasonal component were cumulated and calculated based on the data set for the individual products. Based on this data set, the split into a training set and a test set was carried out and the Random Forest was trained.

In the next step, this trained model was put into a for-loop to make the prediction for the forecast horizon. This for-loop and the collection of results and performance works analogously to the second for-loop in the forecast for the individual products. The following figure shows the result of the forecast for the total sales.

| | Mo | Tue | Wed | Thur | Fri | Sat | Sun |
|---|---|---|---|---|---|---|---|
| predicted_Kw11 | 200.1944 | 158.3385 | 114.5295 | 106.7540 | 100.26940 | 110.8382 | 146.1549 |
| predicted_KW12 | 204.9773 | 139.9484 | 104.9104 | 108.3524 | 100.42086 | 110.6100 | 138.6213 |
| predicted_KW13 | 202.0639 | 122.7348 | 106.4687 | 106.8690 | 101.44244 | 114.7388 | 145.2802 |
| predicted_KW14 | 195.4020 | 134.1619 | 104.4162 | 105.5984 | 95.28421 | 109.0557 | 146.8136 |
| actual_Kw11 | 210.0000 | 151.0000 | 106.0000 | 105.0000 | 87.00000 | 117.0000 | 120.0000 |
| actual_KW12 | 186.0000 | 131.0000 | 97.0000 | 104.0000 | 107.00000 | 126.0000 | 132.0000 |
| actual_KW13 | 254.0000 | 163.0000 | 89.0000 | 110.0000 | 100.00000 | 124.0000 | 154.0000 |
| actual_KW14 | 181.0000 | 121.0000 | 121.0000 | 120.0000 | 145.00000 | 119.0000 | 103.0000 |

*Figure 12: Forecast results for total sales.*

The performance of the prediction is evaluated in this thesis with the RMSE indicator, i.e., the root squared mean error. It is used to compare the values of a prediction with the actual values and makes a statement about the accuracy of the prediction (Chai & Draxler, 2014).

In this thesis, it was calculated with a function that follows this formula:

$$RMSE = \sqrt{\sum_{i=1}^{n} (x_i - y_i) \frac{(x_i - y_i)^2}{n}}$$

$x_i$ *represents the predicted values*

$y_i$ *represents the observed values*

$n$ *represents the number of observations* (Moody, 2019).

The following figure shows an excerpt of the RMSE values for the five most sold products. The RMSE values are difficult to compare with each other because the sales figures for coffee and pastry, for example, differ significantly. This has a negative impact on the significance of a comparison.

| | Item | RSME |
|---|---|---|
| 7 | Coffee | 4.7825 |
| 1 | Bread | 5.1725 |
| 10 | Tea | 3.14 |
| 25 | Cake | 3.655 |
| 8 | Pastry | 3.0975 |

*Figure 13: Excerpt of RMSE performance for individual products.*

The RMSE for total sales is 19.599 when the seasonal component is included. If this variable is not included in the data set, the RMSE improves slightly to 19.29. This could be due to the cumulative calculation of the seasonal component, in which a mean was taken over all seasonal components of the 95 products. In order to understand these RMSE figures, the average sales volume for bread can be seen in Figure 10 and the total sales volume can be seen in Figure 11.

As explained earlier, the Random Forest algorithm is a black box, which means that the prediction is the result of multilinear relationships, which makes the interpretation of the results difficult. However, the importance of the individual variables can be measured and evaluated. The following figures show the importance of the variables for the prediction of the product "Coffee", representing the best-selling product, and the prediction for total sales.

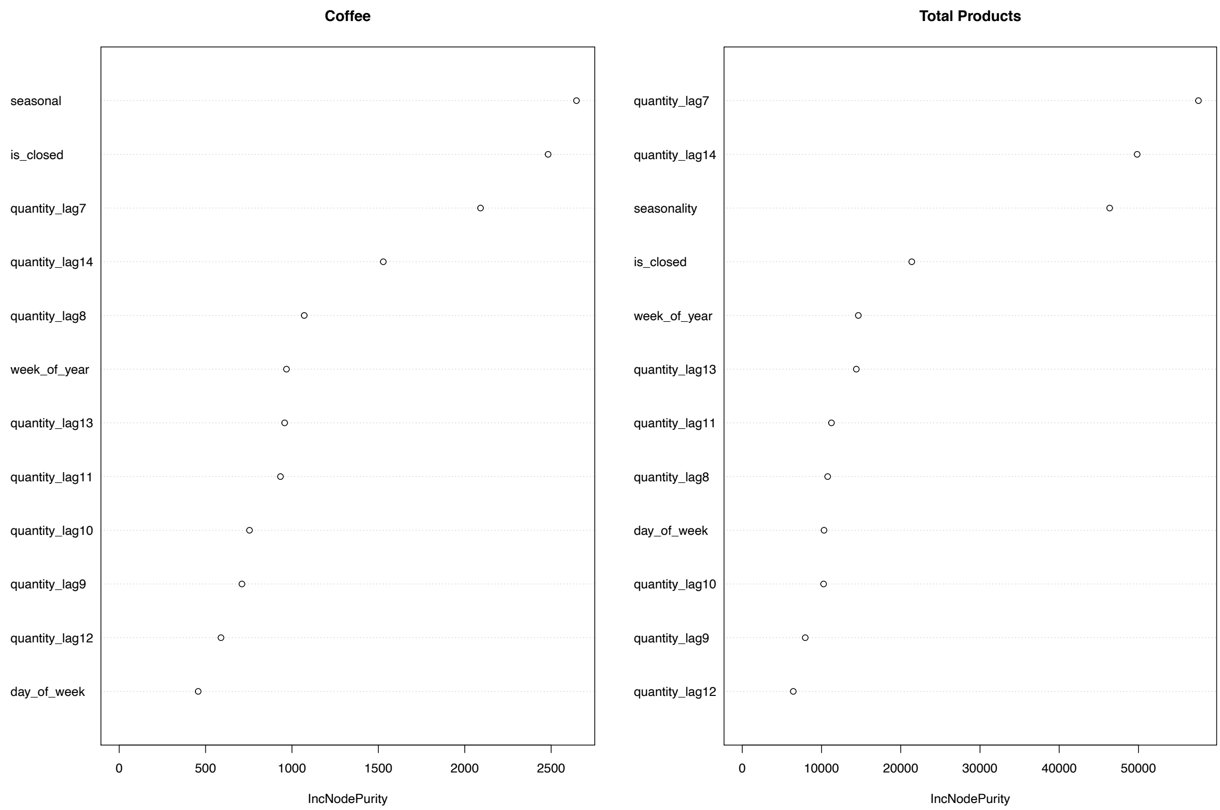



*Figure 14: Importance of different variables for both of the forecast models.*

The IncNodePurity represents the Mean Decrease Gini and is used to show the importance of a variable (Louppe, Wehenkel, Sutera, & Geurts, 2013). In the figure, the variables are ordered in descending order of importance.

In the coffee prediction, the seasonal variable appears to be the most important for the prediction. This could be related to the fact that coffee sales follow a clear seasonal pattern, as shown in Figure 7. The "is closed" variable is in second place and thus also has a large impact on the result. This could be related to the fact that the bakery was only closed on four days in the given period, on 25$^{th}$ and 26$^{th}$ of December and 1$^{st}$ and 2$^{nd}$ of January.

The variables lag7 and lag14 achieve the greatest influence of all lag variables. The variables "week of year" and "day of week" do not have much influence on the prediction result, which could be caused by the small size of the data set; this could not sufficiently incorporate special features on certain days or weeks throughout the year. Alternatively, the seasonal component could act as a proxy for these two factors and thus weaken the significance.

An analysis of the importance of the variables for the total products shows that variables lag7 and lag14 have the greatest importance for the prediction. The reason why the seasonal component dropped to third place could be that this variable was calculated over the average of all seasonal components of all 95 products. This could have introduced bias or errors as sales figures and seasonal components vary greatly between products. Compared to the individual prediction, where the importance of the individual variables is decreasing evenly, the other variables drop off sharply here, which means that the prediction of total sales is mainly based on the variables lag7, lag 14 and the seasonal component.

The last step was to prepare and visualise the various results of the forecasts so that they could be implemented in Shiny and the web application. Based on the requirements and needs identified in the expert interviews, the business user should have both figures and visualisations of the different predictions to support their decision-making process. For this purpose, several csv-files were first modelled in order to be able to efficiently transfer the numerical results into Shiny. Furthermore, visualisations were created that show the weekly and monthly development and compare the predicted and actual values in order to give the business user a simple and understandable platform for his decision-making process.

# 4 Results

In this chapter, the results of the master thesis are presented. First, the adapted design principles derived from the expert interviews are explained and research question 2 is answered. Subsequently, the development process from which the explanation interface emerged is illustrated, whereby a particular emphasis is placed on explaining the reasons for the individual changes made to the design features and where they originated.

## 4.1 Design Principles derived from Interviews

In this chapter, the results of the expert interviews are presented and transformed into design principles in order to make the results of the time series prediction understandable and interpretable for non-specialist business users. The design principles developed in this way provide the answer to the second research question and form the foundation for answering the third research question, as well as the development of the web application. The development process is documented in Chapter 4.2.

The seven interview participants consist of people who are either currently working or have previous experience in a catering environment or large restaurant. One interviewee, Mrs. Bold who is head of HSG's G division and responsible for the operation of three bar and event venues, has no previous experience in the catering industry.

All seven interviewees wanted an explanation of the results of the predictive model. There is a consensus across all interviewees that a system with unexplainable or uninterpretable results would not be useful for the daily decision-making process. No participant would blindly trust an application that delivers results that are neither realistic, understandable nor explainable. It can therefore be deduced that the general trend towards XAI, which is present in other industries, also finds application in the system gastronomy. Conversely, this means that all applications based on AI or ML that are to be accepted by the people in this industry require an additional layer of interpretability that can be understood by the end user. An interview with Mr. Betzl discusses how the catering industry displays a certain antipathy towards technology. Mr. Betzl made it clear that in the catering industry, the gut feeling and experience of the chef is what counts most and that this is how the needs for the coming days and weeks are elicited. Therefore, it can be concluded that a product based on a black box algorithm only has a chance to be accepted in this market if the results of a prediction of such an application are perceived as trustworthy and realistic. Thus, the design principle "Added Layer of Interpretability", which is often found in research, is also applied in the practical implementation in this specific context.

A further analysis of the interviews further proves that a clear, concise and easy-to-use application is preferred to an application that is overloaded with functions and based on technical explanations. The post-hoc explanation and interpretability of the results of the prediction model were consistently rated as more important than an explanation of the intrinsic functioning and the technical specifications of the algorithm. For this reason, no in-depth explanations of the employed algorithm or technical specifications were provided in the implementation. In all interviews, a simply structured and clear implementation was preferred, which makes use of intuitively understandable functions and is explained in a simple language that can be understood by people with little technical knowledge. The majority of interviews emphasised that the application must not overwhelm the users with functionality and technical terminology to prevent a low adaptation of the application. For this reason, “Ease of Use” is derived as a second design principle and is applied in the implementation.

An additional functionality, which was mentioned in almost all interviews to increase the trust and acceptance of the prediction results, was the interactivity of the application. It was requested that the user can try out different elements in order to get an impression of the functionality and their interrelationships. Furthermore, some input options should be available to the user in order to guarantee an interactive menu planning with a variable time horizon. For an individual optimisation of the prediction, it is necessary that the predicted demand can be adjusted, or the prediction period can be changed. In the development of the application, it is central that the user has the feeling that they are in the centre and have control over the application. In a total of six out of seven interviews, the experts described interactivity as a key functionality that supports the credibility of an application. Because of this, interactivity is integrated as a third design principle for the implementation.

In the interviews of Mrs. Bold, Mr. Remus and Mrs. Vögeli, additional emphasis was placed on the comparability of the prediction results. Without a certain comparability of the results, the plausibility and confidence in the results are negatively influenced. Therefore, for each predicted value, the interview participants requested a comparative value from the previous period based on the historical data. In addition, a weekly view is required in order to be able to analyse the development of demand week by week. Furthermore, there should be the possibility of a scenario-like planning in order to test different menu combinations and to create the menu plan for the following weeks. For this purpose, the forecast horizon was set to 28 days. Due to these needs, the design principle "Comparability" was adopted and integrated into the implementation.

The last design principle derived from the expert interviews is "Factor Importance". According to the interviews, it is necessary to see which factors influence the prediction in order to achieve a better interpretability of the results. Since the explanations must be intuitive and easy to understand, a visualisation with icons was chosen. The icons are specifically adapted to each catering facility so that, for example, special events or seasonal features are integrated into the forecast. This need was evident in the interviews with Mrs. Bold, Mr. Remus and Mrs. Vögeli who explained the dependence of demand on various events, such as the exam period at universities or special recurring events. The influencing factors, the icons and their explanations are explained in more detail in Chapter 4.2. The following illustration summarises and describes the factors identified in the expert interviews and answers the second research question:

RQ 2: *"What design principles must be incorporated in a visualization in order to illustrate the results of a time series forecasting model in a comprehensible, intuitive and user-friendly way?"*.

| # | **Principle** | **Explanation** |
|---|---|---|
| 1 | Added Layer of Interpretability | The interview partners all asked for an interpretation aid of the results to ensure trust and comprehensibility. |
| 2 | Ease of Use | To ensure the ease of use of the programme, a simple, clear and concise presentation of information was key for the interview partners. Explanations in text form should be kept short and without the use of technical language. In daily use, a quick learning curve and a fast-responding programme are essential. |
| 3 | Interactivity (incl. User has control, What-if Scenarios) | According to the interview partners, the comprehensibility of the results and the trust in them can be improved by an interactivity. In the catering industry, it is paramount to see what the total demand for menus is and how this is divided among the three or four menus offered. Being able to adjust these numbers in the tool seems crucial |
| 4 | Comparability | Some interviewees felt it was important to be able to compare the results with different values based on the historical data in order to better assess whether the algorithm predicts realistic numbers and to achieve a better plannability of the menus. |
| 5 | Factor Importance | For most interview partners, it is important to know the factors that go into the prediction. This should be displayed in conjunction with small visualisations (icons) whose message is quick and easy to understand. In this way, specific factors (weather, season, special events) should be visualised. |

*Table 4: Derived design principles from expert interviews.*

## 4.2 Prototyping of the Explanation Interface

The next section describes the individual development steps of the GUI and the webpage in detail. The description follows the chronological order of the prototyping. In addition to describing the functionality of the elements of the prototype, the following tables explain why these elements are needed in the user interface, thus explaining the specific benefit for the business user and where the idea for the specific elements originated.

### 4.2.1 1. Cycle of Prototyping

The first iterative cycle of prototyping is initiated with the originally proposed graphical elements for the user interface. These were further developed based on the collected expert feedback in two prototyping rounds during the design workshop. The following table shows all the elements of the GUI that were developed based on the expert interviews or best practice seen on the web. These elements serve as a basis for the design workshop with the business experts and should enable a swift transition to prototyping during the workshop. The table lists the respective element of the GUI, the corresponding description, the goal and purpose of the implementation, and the source from where the idea for the element originated.

| **Proposed graphical elements for the user interface** | | | | |
|---|---|---|---|---|
| **#** | **Element of GUI** | **Description** | **Importance** | **Source** |
| 1 | Navigation bar | The navigation bar contains the tabs for the demand forecast and the detailed explanations of the functionality, is located on the left side and stands out in colour from the rest of the application. | It serves as a control and overview element for the user, which makes operation with the application simple and intuitive. | Best practice on Fiddler.ai, 2020; RStudio, 2020b |
| 2 | Demand forecast | In this window, the user sees the input screen, the demand history, the menu prediction with the different recommended menus, the selected date, the factors influencing the prediction and the top 3 alternatives for the menu recommendation. | The central element of the application. It offers the user a clear structure that presents the results from the predictions in a comprehensible and trustworthy way, so that they can be used in- | Demand forecasting and menu planning as central elements of the application. Interviews: Eric Meier, Michaela Frank, Thomas Betzl, Jessica Svahn Bold, Michael Remus |

| | | | | |
|---|---|---|---|---|
| | | | tuitively in everyday operation without a great introduction. | & Maik G., Anonymous, Doris Vögeli |
| 3 | Explanations | In this window, the user can view detailed explanations of all elements of the visualisation and how the forecast works. | The easy-to-understand but detailed descriptions make it easier for the user to understand and trust the functioning of the application and the forecast. Furthermore, the operation and training time are shortened even further. | A separate tab dedicated to general explanations was classified as crucial in each interview. |
| 4 | Input mask | In the input mask, the user can provide various inputs for the algorithm in order to overwrite the prediction and adjust it to his or her liking. For example, one can adjust the number of menus or the factors that influence the prediction. | Several studies have shown that the interactivity of an application enhances user access and understanding, thus increasing the credibility of and trust in the application. Furthermore, by playing around and trying things out, the user can acquire a better understanding of the functionalities. | Interactivity was rated as important in the following interviews.<br>Eric Meier, Michaela Frank, Jessica Svahn Bold, Michael Remus & Maik G., Anonymous, Doris Vögeli |
| 5 | Course of demand | The demand curve shows the demand for the different menus over the course of the year and the week. | Through the easy-to-understand graphs, users can get a better understanding of the demand for their menus over time and thus identify possible trends and seasonalities and compare different menus. | Interview Michael Remus & Maik G., Anonymous, Doris Vögeli |

| | | | | |
|---|---|---|---|---|
| 6 | Cumulative menu number | This field shows the predicted cumulative demand for menus on a selected day. | This field allows the user to assess whether the predictions are realistic and allows a comparison with the user's "gut feeling" and implicit knowledge. | Interview Michael Remus & Maik G., Anonymous |
| 7 | Previous year's figure | This field shows the respective previous year's value for the total demand or the demand for menus 1-4. | This field allows the user an intuitively understandable and quick comparison with a secured, historical value and thus helps to understand and assess the value of the forecast. | Interview<br>Eric Meier, Michael Remus & Maik G. and Anonymous |
| 8 | Menu Display 1-4 | These fields show the menu recommendations for the user with the resulting quantities of the individual menus. | The simple and clear presentation helps the user to quickly understand the results from the algorithm. | The element that displays the different menus was rated as important in every interview. |
| 9 | Top 3 Alternatives | These fields show the user alternative options for the recommended menu suggestions. | This field helps with menu planning by providing the user with a manageable variety of menus from which to choose. | Interview<br>Eric Meier, Michaela Frank, Michael Remus & Maik G. |
| 10 | Icons | When designing the application, care was taken to convey as much information as possible with icons. | Icons convey a large amount of information in a space-saving, clear and intuitive way and thus increase the comprehensibility and clarity of the application. Furthermore, the recognition value of individual functions or elements is increased. | Interview<br>Michael Remus & Maik G, Anonymous, |

| | | | | |
|---|---|---|---|---|
| 11 | Factors influencing the prediction | This field displays the factors that influence the prediction in the algorithm. Possible icon categories and their sub-elements of icons are: Guest volume:<br>Low - Medium - High<br>Weather:<br>Sunny - Cloudy - Rainy<br>Season: Warm - Cold<br>Terrace operation: Yes - No<br>Special events: Yes - No | Based on these icons, the user is shown in a simple and understandable way how the algorithm weights different factors on the selected day. As these can be tailored to the maximum to the individual specifics of the user's operation, they increase the credibility and confidence in the prediction. | Interview<br>Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 12 | Guest volume | The icons for this category describe the guest volume and thus the demand for menus on the selected day. | If necessary, this factor is precisely adapted to the specifics of the institution in question, thus increasing the quality of the predictions and their credibility. | Interview<br>Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 13 | Weather | The icons for this category describe the weather on the selected day and thus the effects of the different weather conditions. | If necessary, this factor is precisely adapted to the specifics of the institution in question, thus increasing the quality of the predictions and their credibility. | Interview<br>Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 14 | Season | The icons for this category describe the season on the selected day and thus the impact on the demand for the various menus. | If necessary, this factor is precisely adapted to the specifics of the institution in question, thus increasing the quality of the predictions and their credibility. | Interview<br>Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 15 | Terrace operation | The icons for this category describe whether it is possible to cater for the outdoor | If necessary, this factor is precisely adapted to the specifics of the institution in | Interview |

| | | | | |
|---|---|---|---|---|
| | | area on the selected day, and thus the impact on the demand for the various menus. | question, thus increasing the quality of the predictions and their credibility. | Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 16 | Special events | The icons for this category visualise whether a special event is taking place on the selected day, which changes the demand for menus. | If necessary, this factor is precisely adapted to the specifics of the institution in question, thus increasing the quality of the predictions and their credibility. | Interview<br>Jessica Svahn Bold, Michael Remus & Maik G, Anonymous, Doris Vögeli |
| 17 | Un-Do Button | A button is to be installed in the input mask with which users can undo their actions step by step. | According to several studies, the introduction of such a button increases the ease of use of an application and user acceptance. | Backtracing / Reversing Steps Babich, 2019; Dudley et al., 2018; Kulesza et al., 2015; Lim & Dey, 2011; Nielsen, 1994; Shneiderman et al., 2016. |

*Table 5: Proposed graphical elements for the first prototype.*

***Feedback: Design Workshop with Business Experts***

The design workshop aimed to gather ideas for the design of the user interface for the visualization of demand forecasting in the system gastronomy and to prototype it. In order to accelerate the creativity and the start of prototyping, individual graphical elements and functions of the web application were already created before the design workshop.

To achieve the goal of the workshop, two experts were invited. Before the actual prototyping, it was ensured that the participants understood the basic design principles of developing a user-centred UI and that they understood the prepared elements and their underlying functions. In the subsequent brainstorming session, ideas were exchanged, which were then used to create initial prototypes in two rounds of development. The objectives and detailed process of the Design Workshop and a visualization of the design elements can be found in Appendix C.

The design elements were well received by the two experts because they were visually designed and clearly structured. The selection of icons for the influencing factors of the prediction was found to be good. According to the experts, the level of interactivity was adequate and covers the most important functions. However, it was commented that a lot of information is displayed in one window. To overcome this obstacle, it was suggested that the demand history and reporting be moved to a new window. The experts expressed the need for a weekly view, which could be used to plan the menus for entire weeks. In addition, a drop-down menu was developed to enable an easy selection of menus. The whole documentation of the design workshop and the resulting prototypes from the two different prototyping rounds can be found in Appendix C.

***Integration of the Expert Feedback into the Prototype***

In the second development round, the prototype from the first round was further developed. For this purpose, the elements and windows in the prototype were rearranged to create more space and overview on the surface. Furthermore, the various functions were divided into categories and corresponding tabs were developed to streamline the interface and make it clearer for the business user. The resulting prototype from this development round is documented in Appendix C. In the following tables, only elements that have either been changed or redesigned are listed. The changes made in this development phase all originate from the design workshop and can be found in the documentation.

| **Changed elements** | | | | |
|---|---|---|---|---|
| **#** | **Element of GUI** | **Change** | **Importance** | **Source** |
| 1 | Navigation bar | The navigation bar has been extended to include the elements of the weekly view and the demand history. | The new Demand History window was created to streamline and simplify the Demand Forecast window. | Design Workshop<br>Michaela Frank, Eric Meier |
| 2 | Number of menus | In the design workshop, the experts expressed that three predicted menu recommendations would be sufficient for their operation. | By reducing the number of menus, the application as a whole becomes clearer. | Design Workshop<br>Michaela Frank, Eric Meier |
| 3 | Icons | In the workshop it was noted that the icons for closed terrace operation and no special events were not necessary. | By reducing the number of icons, the application can be made simpler and clearer. | Design Workshop<br>Michaela Frank, Eric Meier |

| | | | | |
|---|---|---|---|---|
| 4 | Course of demand | A new window has been created in the navigation bar for the graphics to illustrate the demand for the individual menus. | Removing the graphs creates more space in the demand forecast window and thus enables a clearer and more structured presentation of the forecast. Furthermore, the graphs can take up more space in a separate window to be displayed larger and more detailed. In addition, this further improves the thematic congruence of the individual windows. | Design Workshop<br>Michaela Frank, Eric Meier |

*Table 6: Changed elements implemented in the second prototype.*

| **New elements** | | | | |
|---|---|---|---|---|
| **#** | **Element of GUI** | **Description** | **Importance** | **Source** |
| 1 | Weekly view | The weekly view follows the same system as the demand forecast for one day. In addition, the selected day and the following four days are displayed to provide an estimate of the weekly requirement of the three recommended menus.<br>It should be noted that the input mask only changes the forecast in the demand forecast window. The icons of the influencing factors are displayed for the entire week. As with the demand forecast, the demand for the day and the comparative value from the previous year are visualised. | The weekly view simplifies menu planning for the user, as they can see at a glance when which menu is most in demand. The added value the user gets from using the is improved. | Design Workshop<br>Michaela Frank, Eric Meier |
| 2 | Dropdown for menu customisation | The drop-down menu allows the user to quickly and intuitively customise the recommended menu according to their preferences. The list from which the user | The dropdown helps the user to quickly compare individual menus. He thus receives valid decision support for his menu | Design Workshop<br>Michaela Frank, Eric Meier |

| | | | | |
|---|---|---|---|---|
| | | can choose is based on the individual institution's data set and can be adapted to the specifics of each institution. | planning. The added value for the user, which he receives through the use of the menu, is improved. | |

*Table 7: New elements implemented in the second prototype.*

### 4.2.2 2. Cycle of Prototyping

The second iterative cycle of prototyping begins with the initial implementation of the explanation interface in a Shiny Web App. The Code for the Shiny App can be accessed in the GitHub repository dedicated to this thesis (see Appendix E). The basic functions of Shiny are explained before the initial configuration in Shiny is presented. In a next step, the feedback from the various evaluation interviews is analysed and illustrated.

***Employment of Shiny R package***

The Shiny R package was created by RStudio (RStudio, 2020a) and provides a free and open-source toolbox to enable developers to create user-friendly and interactive web applications based on the R platform without needing in-depth knowledge of HTML or JavaScript. Shiny is an extension of the R platform and therefore provides the possibility to visualise any output generated in R. This makes it possible to present the results of complex statistical models in an easily explainable and interactive way in a web browser (Resnizky, 2015, p. 8).

To program a web application in Shiny, the normal R Console is employed. The development of the application is done in two different R files, the configuration of the user interface in the ui.R file and the server function in the server.R file. In the ui.R file, the design, the elements and the structure of the app are determined, i.e., everything that can be seen later in the browser. The instructions for operating the app are written in the server.R file. The shinyApp function merges the information from the two files and generates the GUI based on the related ui.R and server.R files (RStudio, 2020c).

Similar to R, Shiny also has countless packages that can be used to create complex functions and interactions. The following table lists all the packages used to develop the web application.

| Package name | Description | Source |
|---|---|---|
| shiny | Like the R base package, the shiny package serves as a platform for the most basic functions | RStudio, 2020a |
| shinydashboard | This package introduces new features and functions designed to simplify the creation of interactive dashboards. | Chang & Borges-Ribeiro, 2018 |
| shinyWidgets | This package introduces a new set of input controls and UI elements and gives the developer the ability to customise them. | Perrier, Meyer, Granjon, & Fellow, 2021 |
| shinyjs | This package allows extensive customisation of design elements on the user interface through JavaScript operations. | Attali, 2020 |
| rintrojs | This package allows easy integration of step-by-step instructions, interactive notes or landing pages for the web application. | Ganz & Mehrabani, 2019 |

*Table 8: Employed Shiny packages*

***Implementation in Shiny***

The following figure schematically illustrates the theoretical structure and the relationships of the different files of the web application developed in this thesis based on Shiny R. The data and plot input originates from the time series prediction explained in Chapter 3.5.

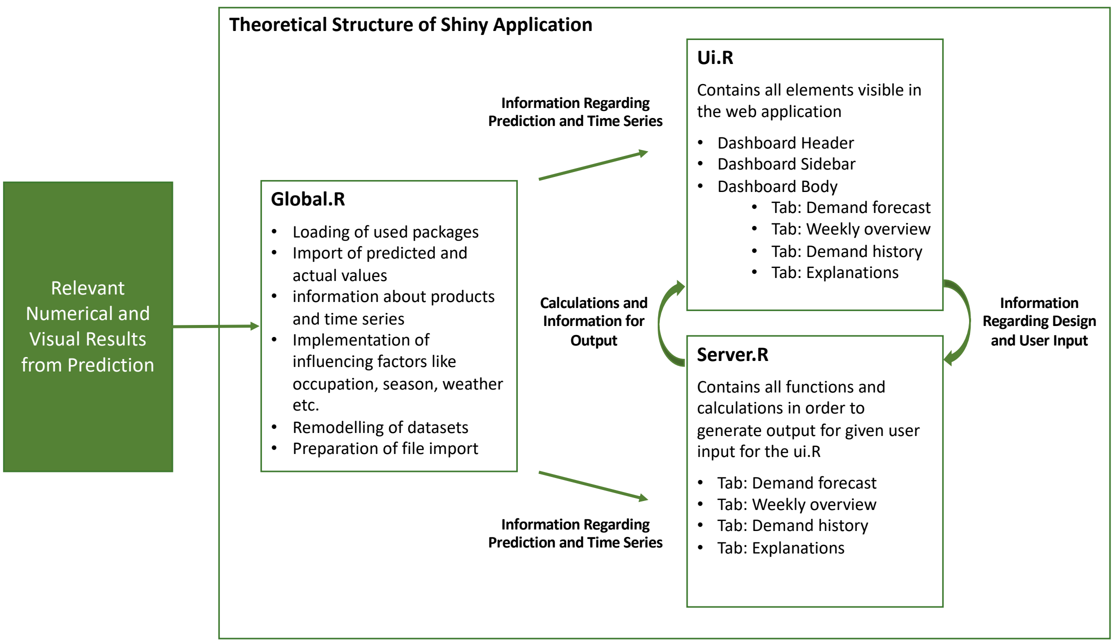



*Figure 15: Structure of the Shiny implementation.*

Based on an analysis of the average sales values, it could be determined that only 21 relevant product categories exist in the present data set. In this work, a product is considered relevant if it was sold on average at least once a week. Products that were only offered and sold over a certain timeframe are not further considered by this general definition. However, since the complete data set only covers 23 weeks, a product that was only offered in a limited timeframe would only have to have been sold 23 times per week to be classified as relevant. Since this threshold is low, it can be assumed that the relevant products are included in the data set of relevant products, even if they were not offered for the entire period. All 21 relevant products are listed in the following figure. These 21 products and their corresponding results from the prediction serve as the basis for the Shiny implementation and all functions within the Shiny system are aligned on the corresponding data set.

```
 [1] Bread           Scandinavian   Hot chocolate  Cookies        Muffin
 [6] Coffee          Pastry         Medialuna      Tea            Farm House
[11] Juice           Soup           Cake           Coke           Sandwich
[16] Alfajores       Brownie        Truffles       Toast          Scone
[21] Spanish Brunch
```

*Figure 16: Relevant product categories.*

In addition to server.R and ui.R which were explained in the preceding chapter, a file called global.R was created due to the complexity in this implementation. As can be seen in Figure 15 above, this file represents a kind of cornerstone and information pool for the functions and processes of the other files. In this file, all packages required for the operation of the web application are provided, all relevant information is loaded, modelled and prepared for further use. The influencing factors of the forecast are implemented in this file, which should make it easier for the business users to understand and interpret the results of the forecast. In total, five such factors have been developed via corresponding icons in the UI of the web application. The purpose and design of these are described in detail in the subsequent chapter.

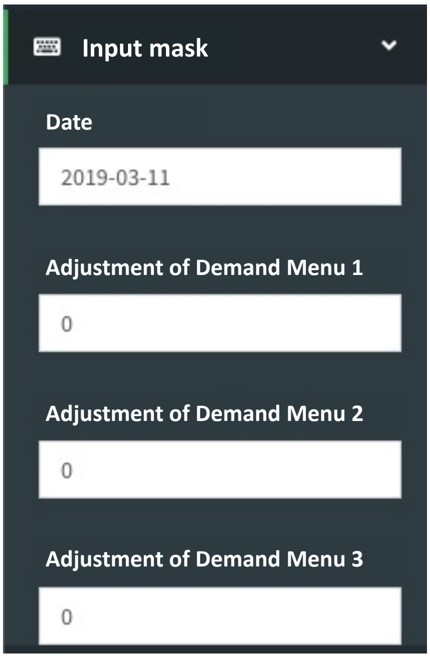



*Figure 17: Design of input mask.*

The second important element of the Shiny structure is the ui.R file. All design elements that are visible on the web application are developed here. Besides the configuration of the general elements like the header and the sidebar of the dashboard, the design, the structure and the elements of the different windows are engineered in this file. In total, four tabs will be implemented: the demand forecast, the weekly overview, the demand history and the explanations for the whole web application. The purpose and the exact functions are explained in detail in the subsequent chapter. It is important that each element in the ui.R file has a corresponding function in the server.R file that allows the element to be interactive. The ui.R and server.R are in constant correspondence with each other, with the ui.R file sending user input information to the server.R file. The sent information is converted there via functions into output information, which is sent back to the ui.R file and displayed to the user.

The dynamics and the interactivity of the web application will be briefly explained using the input mask as an example. The structure of this element is illustrated in Figure 17. In ui.R, the design of the input mask is implemented within the sidebars configuration. For this purpose, an item is defined under which an input field for a date and three fields for numerical input are created. These four fields are designed interactively for the user, i.e., in the date field the user can select between the dates of the forecast horizon. The information is retrieved by a function that accesses information stored in the global.R file. The other three fields are pure input fields, meaning the user can freely enter any whole numbers to change the displayed number of required menus for the next time period. If the entry is successful, the information will be forwarded to the corresponding function in the server.R file. In the case of an entry in the date field, the corresponding information is sought out in the stored data set and displayed at the specified position in the demand forecast tab. For example, if the user wants to replace the demand of menu 1 with the number 50, the function in the server.R file overwrites the original value from the stored data set and displays the new value of 50 in the predefined graphic element. The change in menu 1 also changes the total demand for menus. However, since this value consists of the sum of three numbers, the server.R function must first check which of the three numerical input fields in the input mask have been changed. These changes are checked with the help of an If function and multiple else-if conditions. In this example, the If function detects that input field 1 has been adjusted, but input fields 2 and 3 have not. Based on this information, the function now adjusts the three values which comprise the total demand accordingly and displays the new value again in the predefined graphical element. The relationship between the different elements is shown in the following graphic.

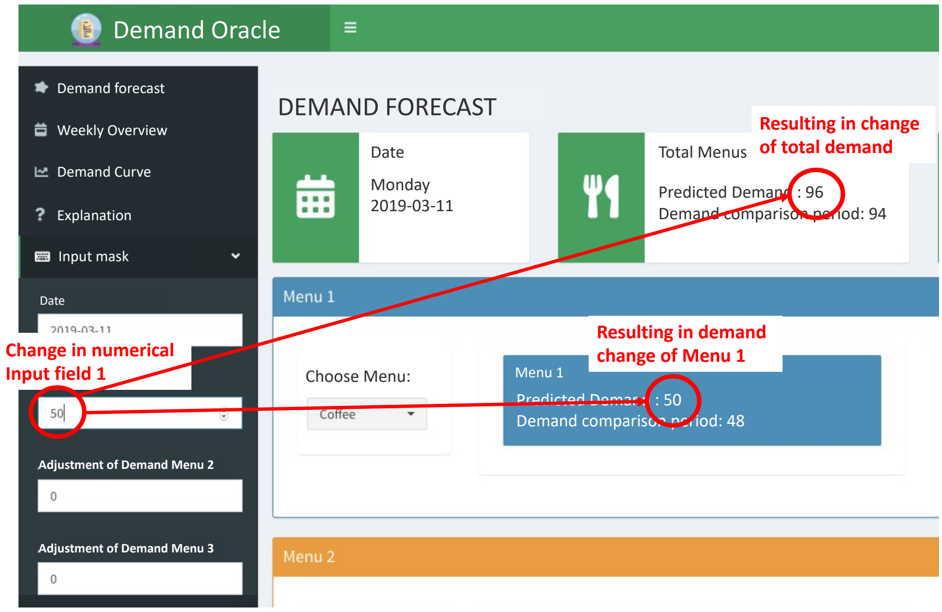



*Figure 18: Visualisation of adjustment dynamic in the UI related to user input in the input mask.*

***Initial Configuration of Shiny Web Application***

In the next development step, the prototype from the design workshop was transformed into a web application using the R package Shiny and other supplementary packages (see previous chapter). Due to the way Shiny works and design considerations, there are small differences between the prototype resulting from the design workshop and the web application programmed with the help of R and Shiny. These deviations mainly concern the arrangement and structure of the application.

If individual elements or functionalities have been implemented only partially, in a modified form or not at all, the corresponding explanations, in addition to the description of the implementation of all visual elements and functionalities, can be found in the following tables.

| Implemented elements | | |
|---|---|---|
| # | **Elements of GUI** | **Description** |
| 1 | Navigation bar | The functionality of the elements was integrated as discussed in the design workshop. |
| 2 | Demand forecast | |
| 3 | Weekly overview | |
| 4 | Course of demand | |
| 5 | Explanations | |
| 6 | Input mask incl. functions | |
| 7 | Cumulative menu number | |
| 8 | Factors influencing the prediction | |
| 9 | Icons | |
| 10 | Menu recommendation 1-3 | |
| 11 | Alternative options | |
| 12 | Dropdown menu selection | |
| 13 | Integration of the previous year's values | |

*Table 9: Elements implemented in the web application.*

| Changed elements | | | |
|---|---|---|---|
| # | **Elements of GUI** | **What was changed?** | **Why was it adapted?** |
| 1 | Overarching design and structure of the demand forecast | The functionalities were implemented in the Shiny app, except for the input mask, as discussed in the design workshop. The design, the colours and the arrangement of the elements were adapted based on the functionality of the Shiny implementation. | The input mask was moved to the navigation bar to generate more space for the remaining elements. Furthermore, an attempt was made to ensure a uniform, clear and structured overview in the Shiny app in order to develop a customised app for the user that is precisely tailored to his or her requirements and needs. |
| 2 | Overall design and structure of the weekly overview | The functionalities were implemented in the Shiny app, except for the input mask and the influencing factors as discussed in the design workshop. The design, the colours and the arrangement of the elements were adapted based on the functionality of the Shiny implementation. | The input mask was moved to the navigation bar to generate more space for the remaining elements on the interface and to provide the user with a simpler and clearer user interface.<br>The icons that influence the forecast have been determined not only for the whole week |

| | | | |
|---|---|---|---|
| | | | but for each individual day of the selected period, in order to provide the user with increased clarity and improved credibility of the application. |
| 3 | Graphs in the Demand History window | This window now displays the top 10 menus over the entire time course, as well as the weekly and monthly listing of the demand for the cumulative and individual menus. A drop-down menu allows you to switch between the weekly and monthly view and between the individual menus. | Based on the available data set, these graphs were selected. |
| 4 | Design and text in the Explanations window | The explanations for the various functions and icons can be viewed via a drop-down menu and are not all displayed at once. | By displaying only the selected functionality and icon, the user is not overwhelmed by an abundance of information at once and, therefore can better absorb the explanations piece by piece. |
| 5 | Icons for categories terrace operation and special events | Unlike in the Design Workshop, there are now icons again in the Shiny App for "no special event" and "terrace operation not open". | This change was implemented because now a constant number of icons is displayed, and the user thus realises more easily if no event is pending, or the terrace is not open. |
| 6 | Factors influencing the prediction | In the Weekly Overview window, the icons for the factors influencing the forecast are displayed for each day of the selected period rather than by week. | The icons displayed offer the user little explanatory potential if they are only determined for the whole week, as the individual factors such as the number of guests or the weather are subject to daily fluctuations. |

*Table 10: Changed elements implemented in the web application.*

| **Not implemented Elements** | | |
|---|---|---|
| **#** | **Elements of GUI** | **Why was it not implemented?** |
| 1 | Manual adjustment of the influencing | The question arises at which level the influencing factors should be adaptable. At the current state of the application, it is planned to integrate the individual factors already via the data set. This means that the number of guests is derived from the actual number of visitors, the weather situation, the seasonality, the frequency of special events and the |

| | | |
|---|---|---|
| | factors in the input mask | opening of the terrace as well as the effects of the variables mentioned are already integrated in advance and can be adjusted in this way. |
| 2 | Un-Do button in the input mask | No easy-to-program function could be found in the Shiny environment that comes close to the functionality of an Un-Do button. For this reason, an implementation was not carried out. |
| 3 | Icons for individual menus | The recognition value of icons is high for individual foods or products and offers added value for the user. However, this is not the case for individual dishes that represent a combination of different products. In order not to confuse the user, this functionality was omitted. |

*Table 11: Not implemented elements in the web application.*

***Feedback: Evaluation of the Prototype***

A total of five evaluation interviews were conducted to test the design and functionality of the prototype. The documentation of the evaluation interviews is provided in Appendix D. A summary of the results is illustrated in Figure 19. Overall, the experts were satisfied with the functionality of the prototype.

| Question | Eric Meier | Jessica Svahn Bold | Michaela Frank | Michael Remus | Doris Vögeli |
|---|---|---|---|---|---|
| Do you trust the predictions of the application? | Yes | Yes | Yes | Yes | Yes |
| Are you able to understand the predictions? | Yes | Yes | Yes | Yes | Yes |
| Do the predictions seem realistic to you? | Yes | Yes | Yes | Yes | Yes |
| Do the predictions seem credible to you? | Yes | Yes | Yes | Yes | Yes |
| Are the explanations in the app understandable for you? | Yes | Yes | Yes | Yes | Yes |
| Is the operation of the app intuitive for you? | Yes | Yes | Yes | Yes | Yes |
| Could you imagine using such an app for menu planning? | Yes | No | Yes | Yes | Yes |

*Figure 19: Results of the qualitative evaluation.*

According to the statements in the evaluation interviews, all five experts trust the application and are capable of comprehending the prediction. All participants consider the predictions to be credible and realistic. The explanations in the application were rated as understandable by all participants and the operation was described as simple and intuitive. All participants working in either a restaurant or canteen would consider using the application. However, one participant who manages three different bars and event locations, would require adaptation of individual functions and visualizations in order to ensure that the application optimally supports her daily work.

| Evaluated Element | Eric Meier | Jessica Svahn Bold | Michaela Frank | Michael Remus | Doris Vögeli | Total |
|---|---|---|---|---|---|---|
| Clarity of the application | 3.5 | 5 | 4 | 5 | - | 4.4 |
| Simplicity of the structure | 4.5 | 4 | 5 | 5 | - | 4.6 |
| Simplicity of operation | 4.5 | 5 | 5 | 5 | - | 4.9 |
| Comprehensibility of the explanation | 4 | 5 | 5 | 5 | - | 4.8 |
| Credibility of the predictions | 4.5 | 4 | 5 | 5 | - | 4.6 |
| Transparency of the predictions | 4.5 | 3 | 5 | 5 | - | 4.4 |
| Comprehensibility of the predictions | 4.5 | 5 | 4 | 5 | - | 4.6 |
| Usefulness in daily work | 3.5 | 4 | 5 | 5 | - | 4.4 |
| Usability in daily work | 4 | 5 | 5 | 5 | - | 4.8 |

*Figure 20: Result of the numerical evaluation.*

In Figure 20, the answers of the five experts regarding the numerical evaluation were analysed. In the survey, participants were asked to rate the reviewed element on a scale of 1-5. The results are consistent with the findings from the qualitative questions. The ease of use of the application, the comprehensibility of the explanations and the usability in everyday work were rated excellent with 4.9 and 4.8 respectively. The structure and clarity as well as the usefulness of the application in everyday work were rated lowest with 4.4 yet remained good.

The analysis of the feedback from the individual evaluation interviews resulted in the fourth and final prototype of the web application. The documentation of the results of the evaluation interviews can be found in Appendix D. Since the functionalities of the application seemed sufficient for daily use to all interview partners and sufficient confidence was placed in the predictions, most of the changes relate to an improved design, which fosters clarity, and the structure of the application, improved transparency of the explanations as well an enhanced explanation of how to use the application. An exception to this is the evaluation interview with Mrs. Bold, who is responsible for the G area at the University of St. Gallen, i.e., for the operation of the bars and the event venues. In order to be able to use the application in such an environment, the functions would have to be optimised for bar operation.

### 4.2.3 3. Cycle of Prototyping

In the third iterative development step, the feedback from the evaluation interviews was used to further develop the prototype and adapt it to better meet the needs of the experts. No functionalities were removed from the application. Any new functions, changes to the function or proposed but not implemented elements are described in detail in the following tables.

| **New elements** | | | | |
|---|---|---|---|---|
| **#** | **Elements of GUI** | **Description** | **Importance** | **Source** |
| 1 | Landing page | The landing page is the first page a user sees when starting the application. It contains an introduction to the functionalities of the application. A returning user should not see this introduction again. | During the evaluation interviews it became clear that although the app is understandable and clearly structured, the app is overwhelming with all the functions without an initial introduction. The landing page was designed with the aim of keeping the familiarisation time low and increasing the understanding of the functions. | Evaluation Interview Eric Meier, Jessica Svahn Bold |
| 2 | Explanatory tour | The explanatory tour provides a step-by-step introduction for the user to explain all the core functions on the first visit. | In the evaluation interviews, the need for a step-by-step introduction became apparent in order to reduce the training time and to avoid resulting frustrations. | Evaluation Interview Eric Meier, Jessica Svahn Bold, Doris Vögeli |
| 3 | Demand trend graphs regarding forecasts | Where deemed appropriate, values for the comparison period and the forecast were integrated into the visualisations. | It is now possible for users to compare the predicted values with the values of the comparison period on a weekly and monthly basis to check the validity of the predictions. | Evaluation interview Eric Meier, Michael Remus |

| | | | | |
|---|---|---|---|---|
| 4 | Cumulative demand in the weekly overview | Similar to the "Demand forecast" window, a value has been integrated in the "Weekly overview" window that shows the weekly demand of the individual menus. | It is now quick and easy to see how the total demand of a menu changes in a week's time. | Evaluation Interview Eric Meier |

*Table 12: New elements implemented in the final prototype.*

| **Changed elements** | | | | |
|---|---|---|---|---|
| **#** | **Elements of GUI** | **Description** | **Importance** | **Source** |
| 1 | Extension and renaming of the comparison period | The value used to back up the predicted value was renamed "previous year period". Furthermore, in a final version of the application, it should be composed of several factors (e.g. demand from several previous years, months, similar seasons, etc.) and not only based on the previous year. | The current Corona situation shows that values based only on the previous year are not good comparative indicators. | Evaluation interview<br>Jessica Svahn Bold, Michaela Frank |
| 2 | Adapt and clarify explanations | An explanation of the variable "comparison period" was integrated. Furthermore, the explanations for the icons were expanded. | It was not clear to the interviewees in the evaluation interviews what exactly the variable "comparison period" comprises and how it is composed. | Evaluation interview<br>Michaela Frank, Michael Remus<br>Doris Vögeli |
| 3 | Renaming the variable "Current demand” | The variable "current demand" was renamed "estimated demand". | The phrase "current demand " was not clear to some interviewees and caused uncertainty. | Evaluation interview<br>Michaela Frank |
| 4 | Other | Typos, small design changes. | | |

*Table 13: Changes implemented in the final prototype.*

| Not implemented elements | | | | |
|---|---|---|---|---|
| **#** | **Elements of GUI** | **Description** | **Importance** | **Source** |
| 1 | Special version for bar operation | In order to be able to use the application in bar businesses as well, a special focus is required on the specifics in such an environment. | In this work, the focus is on system catering, so the functions and the design are optimised for use in this environment. However, it is conceivable that the final version of the application can be adapted to the specifics of a bar business without much effort. | Evaluation interview Jessica Svahn Bold, |
| 2 | Integration Instagram interface | Marketing on social media channels is becoming increasingly important. A good social media presence is particularly relevant for event catering and can have a decisive impact on sales figures. | In a next version of the application, a connection to social media channels is quite conceivable. However, the goal, functionality and content of such an interface must be worked out with all relevant stakeholders in order to develop a solution that is tailored to the needs of the users. | Evaluation interview Jessica Svahn Bold, |
| 3 | Rename menu selection | The data set on which the application is based contains the sales data of a bakery and includes mostly individual products and no menus. Due to this fact, the term "menu selection" in the windows of the application was described as partly misleading. | This element was not renamed because the user interface is geared to the needs of system catering. The underlying data set with the individual products was deposited to illustrate the functionality and thus only serves as a dummy. | Evaluation interview Michaela Frank |
| 4 | Coupling with other systems | The integration of further systems, such as the storage system or the ordering system, offers increased added value for the user. | In this work, the focus is on demand forecasting as well as a detailed explanation of how it works, so at this stage there is no integration of other systems. In a next version of the application, such an expansion of the functionality is well conceivable. | Evaluation interview Michaela Frank |

*Table 14: Not implemented elements in the final prototype.*

# 5 Discussion

In the following chapter, the results of this work will be discussed in order to determine how the most relevant design features of the explanation interface emerged. Additionally, an assessment will be made of whether the final prototype of the interface complies with the principles of XAI, i.e., whether it fulfils the three design requirements established at the beginning of the thesis. Based on these findings, the third research question will be answered at the end of this chapter. In order to assess whether the developed explanation interface follows the principles of XAI, a review of the development of the results of the first and second research questions will be conducted. The entire development process of the design features is shown in Figure 21.

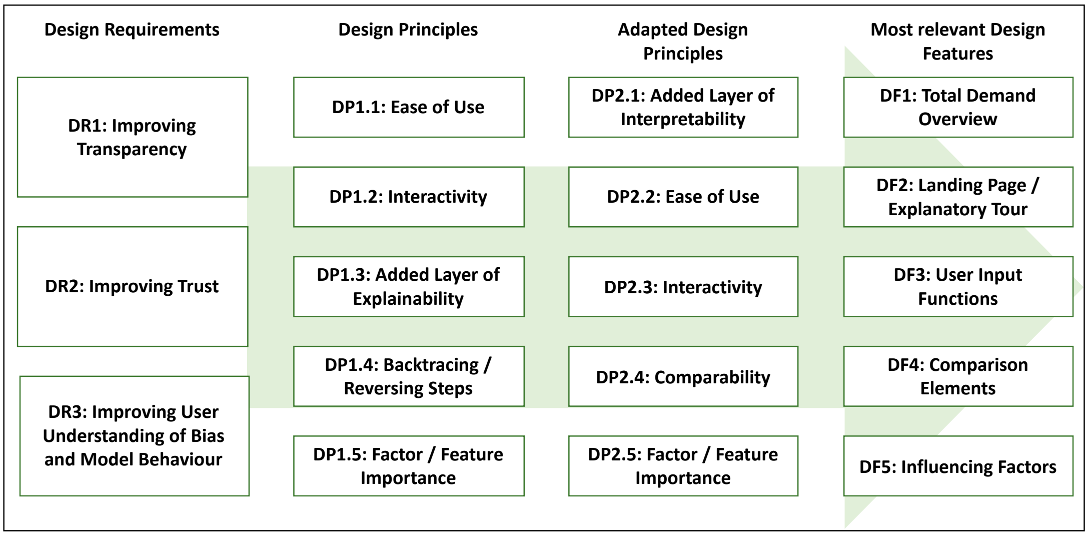



*Figure 21: Evolution of the key design features of the explanation interface.*

As a basis for the prototyping, the goals of XAI defined by science were considered, establishing three design requirements for this work. The resulting prototype is required to open the black box of the prediction model, increase user trust and improve the user's understanding of the prediction model and possible errors and biases (see Chapter 2.1). In a subsequent step, design principles were derived based on a literature review and evaluated whether they are applicable in practice in this form (see Chapter 2.2). The resulting design principles provided the basis for eliciting the requirements and preferences of the experts and were further developed through the feedback collected in the interviews (see Chapter 4.1). Building on these adapted design principles, an iterative learning and development process was initiated in order to implement the design features in the explanation interface in collaboration with the experts (see Chapter 4.2).

The following figure shows the individual relationships between the design requirements, the design principles and design features. Based on the results of the expert interviews and as seen in Figure 22, it can be stated that the design requirements are deeply related to the design principles.

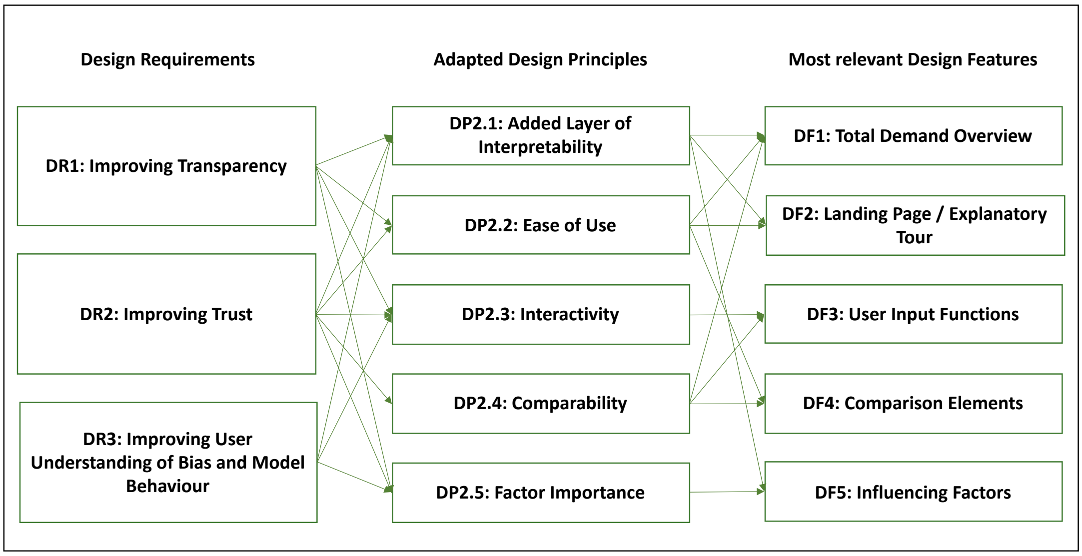



*Figure 22: Relationship between design requirements, design principles and design features.*

The design requirement "Improving Transparency" is supported by the application of Design Principles 1, 2, 3 and 5. They all pursue the goal of either making the functionality of the algorithm more tangible to the user or presenting the results of the prediction in a comprehensible way and making them accessible. Only the design principle "Comparability" does not contribute to an improved transparency of the algorithm This principle does not focus on the functionality of the model but on making the results of the prediction in the user's mental model tangible and comparable and thus comprehensible.

The second design requirement, "Improving Trust", is supported by all the employed design principles, which is not surprising since user acceptance and rapid adoption of a technology depends largely on whether the user has trust in it. The integration of an additional explanation interface leads to a better understanding of the predictive model and the results, which was confirmed in all expert interviews. Furthermore, in six out of seven interviews, simple explanations, easy-to-use functions and a certain degree of interactivity were named as prerequisites for building trust. In three out of seven interviews, the need for comparability between the results of the forecast and the historical data was mentioned, as well as a more detailed description of the factors that have an impact on the forecast in order to be able to trust the forecast model.

The third design requirement "Improving User Understanding of Bias and Model Behaviour" aims to increase the user's level of understanding of the model and its recommendations, enabling the user to recognise and identify possible errors in the predictions in order to ensure that the user can comprehend and assess the behaviour of the model. Thus, design principles 1, 3 and 5 were derived and implemented in the prototype. The additional layer of explanation ensures that the user develops a basic and theoretical understanding of all functions and elements. Furthermore, the interactivity allows the user to experiment with the different functions in order to generate a deeper and practical understanding of the forecasting model. By analysing the different input factors and their impact on the model, the user has the opportunity to familiarise himself with the functioning of the model and thus draw conclusions about the plausibility of the prediction results.

The five design principles, explained in detail in Chapter 4.2.3, were further developed and verified through the expert interviews and were implemented in the prototype with the help of several design features. From this multitude of design features of the explanation interface prototype, a total of five were selected that appeared to be particularly relevant in order to fulfil the requirements of an XAI-based application and ultimately to answer the third research question:

RQ 3*: "What are the relevant design features of an explanation interface that displays the results of a sales forecast, based on the design principles of XAI and optimized for the needs and requirements of professional users in the catering industry?"*

The first relevant design feature is the display of the total demand figure in the daily and weekly overview. The visualisation of total sales across all menus on a daily or weekly basis is crucial for the user. With this figure, it can be derived immediately whether the prediction has a realistic character and agrees with the chef's gut feeling, or whether it is marked by errors and not trustworthy. This functionality was described as essential in the interviews with Mr. Meier, Mr. Remus, Mrs. Vögeli and the anonymous person. This design feature improves the interpretability of the explanation interface, helps to make the app simple and clear and allows a comparison with the user's subjective perception of the current demand in order to better assess the quality of the result of the prediction. Therefore, the design feature was derived from design principles 2.1, 2.2 and 2.4 and helps to comply with design requirements 2 and 3.

The second design feature identified as relevant are the introductory features which include the landing page and the step-by-step explanatory tour. These two elements intend to educate the user on how to use the web application in a user-friendly and interactive way, ensuring a fast and straightforward learning process avoiding frustration for the user. Overall, such introductory features were requested in three out of five evaluation interviews (Mr. Meier, Mrs. Vögeli, Mrs. Bold). Due to the improved comprehensibility of the operation of the web application and the accelerated learning process, this design feature is based on design principles 2.1 and 2.2

and helps to meet the second and third design requirements, as the functionality improves the trust in the system and increases the level of user understanding of the algorithm.

The third relevant design feature revolves around the functions that enable the user to influence the web application through their input. The user can modify implausible values or change the suggested menus, for example, and thus customise the recommendation according to their preferences. These interactive features improve the user's understanding of possible prediction errors and how the prediction model works, and help the user become a trained operator of the forecasting model. In addition, the user can compare different menu predictions and analyse which menu is the best choice for a given day or week with certain given conditions. These kinds of interactive features were requested in six out of seven expert interviews. The implementation of the functions integrates the design principles of "Interactivity" and "Comparability" and helps to fulfil design requirements 1, 2 and 3.

The fourth relevant design feature addresses the comparison values provided to the user in the different screens to compare the plausibility of the forecast with historical data from a comparison period. For example, the daily and weekly overviews display values from the comparison period and the visualisations in the demand history window provide the option to compare the predicted values with the historical values. The necessity of incorporating such a feature in the web application became evident in the interviews with Mr. Meier, Mr. Remus, Mrs. Vögeli and the participant who wished to remain anonymous. These functions should give the user confidence in the operation of the web application and help to assess the quality of the forecast results. Therefore, the implementations of these design feature originate from the design principles of "Ease of Use" and "Comparability" and help to comply with design requirement 2 and 3.

The last relevant design feature that is identified is the display of factors that influence the prediction. This functionality allows the user to better understand the individual influencing factors and how the forecast model works.  This provides the basis for the user to better assess how changing input factors, such as the weather, affect the results of the forecast. In this way, the user develops a feeling for the forecast model and can intuitively assess whether the forecast seems realistic or when the system makes a mistake, which in turn improves trust in the system. The need for such a function was established in the interviews with Mrs. Bold, Mr. Remus and Mrs. Vögeli. The implementation of this feature follows the design principles 2.1 and 2.5 and addresses design requirement 1, 2 and 3.

After presenting the relevant implemented design features and comparing them with the corresponding design principles and design requirements, it can be stated that the design features used address all the design requirements that were derived from the objectives of XAI in Chapter 2. This assumption is further strengthened when the results of the evaluation interviews are integrated into this analysis. In the evaluation interviews, the experts provided positive feedback to the application and considered the results of the prediction to be trustworthy, comprehensible and understandable. The structure and operation of the application as well as the explanations are also described as simple and adequate (see Chapter 4.2.2). Consequently, the prototype of the web application fulfils all three design requirements and therefore the resulting explanation interface complies with the standards of XAI.

# 6 Limitations and Conclusion

The present prototype of an explanation interface based on the principles of XAI was developed in an iterative process in collaboration with business experts from the catering industry. The resulting prototype was designed in three cycles to ensure the accordance with the demands of the users and the fulfilment with the design requirements defined in the thesis. However, any design process is a continuous improvement process and therefore the prototype still has weaknesses and limitations that could be addressed through further iterations or research work.

The dataset used has too few data points to develop a genuinely good prediction for the multiple products. A dataset that would contain two years of transactional data would positively influence the predictions outcome. Furthermore, the different context in which the data was collected caused confusion during the evaluation interviews, as the transactional data from a bakery has different characteristics and specifics than those from a large catering institution.

Furthermore, the focus of the work was on the elicitation of the design principles and the derivation and further development of the design features in the resulting explanation interface; the rudimentary predictive model was not further developed. Thus, the features employed by the model are not particularly sophisticated and the quality of the prediction results could be positively influenced by a more intensive feature design and by fine-tuning the model parameters.

Moreover, the number of expert interviews only represents a small sample of the catering industry as a whole and thus the results cannot be decontextualised for the entirety of the industry. These results could be further validated through further interviews and evaluation rounds.

Another limitation of the work are the functions of the web application, which are based on dummy variables. Assumptions were made for the alternative menus and the influencing factors of the prediction that do not correspond to the historical data set. These features would have to be tested with a realistic dataset in a specific context and adapted to the respective institution. Furthermore, the interactivity of the web application could be further enhanced by using user feedback to improve the prediction model or the explanation interface. In principle, the work focused on the second design requirement "Improve Trust". The other two design requirements were addressed, but there is further potential for optimisation in the implementation of these requirements. In addition to the already mentioned opportunities for improvement, the evaluation interviews provided a variety of possibilities to enhance the prototype in further research work. On the one hand, the focus could be on developing the present system into a monolithic booking and ordering system for the catering industry. This could lead to further efficiency gains by integrating a storage or ordering system. On the other hand, the system could also be extended to other industries such as the event or bar business, as mentioned in the interview with Mrs. Bold.

## Bibliography


Adadi, A. & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, *6*, 52138–52160. https://doi.org/10.1109/ACCESS.2018.2870052

Attali, D. (2020). *shinyjs: Easily Improve the User Experience of Your Shiny Apps in Seconds*. Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/shinyjs/index.html

Babich, N. (2019). *The 4 Golden Rules of UI Design*. Date Accessed: 12.12.2020. https://xd.adobe.com/ideas/process/ui-design/4-golden-rules-ui-design/

Barredo Arrieta, A., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2019). Explainable Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, *58*(October 2019), 82–115. https://doi.org/10.1016/j.inffus.2019.12.012

Beretta, C. & Hellweg, S. (2019). Lebensmittelverluste in der Schweiz: Umweltbelastung und Vermeidungspotenzial. *ETH Zurich, Institut Für Umweltingenieurwissenschaften*.

Biau, G. & Scornet, E. (2016). A random forest guided tour. *Test*, *25*(2), 197–227.

Blohm, I. (2020). *Fighting Food Waste: Introduction to Time Series Forecasting.*

Breiman, L. (2001). Random Forests. *Machine Learning*, *45*, 5–43. https://doi.org/https://doi.org/10.1023/A:1010933404324

Bunt, A., Lount, M. & Lauzon, C. (2012). *Are explanations always important?* 169. https://doi.org/10.1145/2166966.2166996

Bussmann, N., Giudici, P., Marinelli, D. & Papenbrock, J. (2020). Explainable Machine Learning in Credit Risk Management. *Computational Economics*, *57*, 203–216. https://doi.org/https://doi.org/10.1007/s10614-020-10042-0

Chai, T. & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/https://doi.org/10.5194/gmd-7-1247-2014

Chambers, J. M. & Hastie, T. (1992). *Statistical Models in S*. Wadsworth & Brooks/Cole.

Chang, W. & Borges-Ribeiro, B. (2018). *Shinydashboard: Create Dashboards with "Shiny."* Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/shinydashboard/index.html

Charnes, A., Frome, E. L. & Yu, P. L. (1974). The Equivalence of Generalized Least Squares and Maximum Likelihood Estimates in the Exponential Family. *Journal of the American Statistical Association*, *71*(353), 169–171. https://doi.org/DOI: 10.1080/01621459.1976.10481508

Chromik, M., Völkel, S. T., Eiband, M. & Buschek, D. (2019). Dark patterns of explainability, transparency, and user control for intelligent systems. *CEUR Workshop Proceedings*, *2327*.

Corsi, A. & Ashenfelter, O. (2019). Predicting Italian Wine Quality from Weather Data and Expert Ratings. *Journal of Wine Economics*, *14*(3), 234–251. https://doi.org/https://doi.org/10.1017/jwe.2019.41

Das, A., Member, G. S., Rad, P. & Member, S. (2020). Opportunities and Challenges in Explainable Artificial Intelligence ( XAI ): A Survey. *IEEE*, 1–24.

Dastin, J. (2018). *Amazon scraps secret AI recruiting tool that showed bias against women*. Date Accessed: 11.11.2020.

Diakopoulos, N., Friedler, S., Arenas, M., Barocas, S., Hay, M., Howe, B., Jagadish, H. V., Unsworth, K., Sahuguet, A., Venkatasubramanian, S., Wilson, C., Yu, C. & Zevenbergen, B. (2018). *Principles for Accountable Algorithms and a Social Impact Statement for Algorithms*.

Doshi-Velez, F. & Kim, B. (2017). *Towards A Rigorous Science of Interpretable Machine Learning*. *Ml*, 1–13. http://arxiv.org/abs/1702.08608

Dowle, M., Srinivasan, A., Gorecki, J., Chirico, M., Stetsenko, P., Lianoglou, S., Antonyan, E., Bonsch, M., Parsonage, H., Ritchie, S., Ren, K. & Tan, X. (2021). *Package 'data.table.'* Date Accessed: 15.03.2021. https://cran.r-project.org/package=data.table

Dudley, J. J. & Kristensson, P. O. (2018). A Review of User Interface Design for. *ACM Transactions on Interactive Intelligent Systems*, *8*(2).

Fiddler.ai. (2020). *Explainable AI*. Date Accessed: 11.11.2020. https://www.fiddler.ai/explainable-ai

Ganz, C. & Mehrabani, A. (2019). *rintrojs: Wrapper for the "Intro.js" Library*. Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/rintrojs/index.html

Goodman, B. & Flaxman, S. (2016). European Union regulations on algorithmic decision-making and a "right to explanation." *ICML Workshop on Human Interpretability in Machine Learning (WHI 2016)*. https://doi.org/https://arxiv.org/abs/1606.08813

GrandViewResearch. (2020). *Artificial Intelligence Market Size, Share & Trends Analysis Report By Solution (Hardware, Software, Services), By Technology (Deep Learning, Machine Learning), By End Use, By Region, And Segment Forecasts, 2020 - 2027*. Date Accessed: 10.01.2021. https://www.grandviewresearch.com/industry-analysis/artificial-intelligence-ai-market#:~:text=The global artificial intelligence market,42.2%25 from 2020 to 2027

Gray, C. M., Kou, Y., Battles, B., Hoggatt, J. & Toombs, A. L. (2018). The dark (patterns) side of UX design. *Conference on Human Factors in Computing Systems - Proceedings*, *2018-April*, 1–14. https://doi.org/10.1145/3173574.3174108

Grice, P. H. (1975). Logic and Conversation. In *Syntax and Semantics 3: Speech arts.* (Vol. 3, pp. 41–58). Academic Press, New York.

Guidotti, R., Monreale, A. & Ruggieri, S. (2018). *A Survey of Methods for Explaining Black Box Models*. *51*(5).

Gunning, D. & Aha, D. (2019). DARPA's Explainable Artificial Intelligence Program. *AI Magazine*, *40*(2), 44.58.

Hagras, H. (2018). Toward Human Understandable, Explainable AI. *IEEE COMPUTER SOCIETY*, *51*(9), 28–36. https://doi.org/10.1109/MC.2018.3620965

Hevner, A. R., March, S. T., Park, J. & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105.

Ho, T. K. (1995). Random Decision Forests. *3rd International Conference on Document Analysis and Recognition*, 278–282.

Holland, M. (2016). *US-Justiz: Algorithmen benachteiligen systematisch Schwarze.* Date Accessed: 10.03.2021. https://www.heise.de/newsticker/meldung/US-Justiz-Algorithmen-benachteiligen-systematisch-Schwarze-3216770.html

Holzinger, A., Biemann, C., Pattichis, C. S. & Kell, D. B. (2017). *What do we need to build explainable AI systems for the medical domain? Ml*, 1–28. http://arxiv.org/abs/1712.09923

Hyndman, R. (2021). *Cran Task View: Time Series Analysis*. Date Accessed: 10.03.2021.

Hyndman, R., Athanasopoulos, G., Bergmeir, C., Caceres, G., Chhay, L., O'Hara-Wild, M., Petropoulos, F. & Razbash, S. (2021). *Package "forecast."* Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/forecast/forecast.pdf

Hyndman, R. & Khandakar, Y. (2008). Automatic Time Series Forecasting: the forecast

Package for R. *Journal of Statistical Software*, *27*(3), 1–22. https://doi.org/http://dx.doi.org/10.18637/jss.v027.i03

IDC & Seagate. (2018). *Data Age 2025: The Digitization of the world.* https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf

Ihaka, R. (1998). *R: Past and Future History*. Date Accessed: 12.02.2021. https://www.stat.auckland.ac.nz/~ihaka/downloads/Interface98.pdf

Ihaka, R. & Gentleman, R. (1996). R: A Language for Data Analysis and Graphics. *Journal of Computational and Graphical Statistics*, *5*(3), 299–314.

Kapoor, A., Lee, B., Tan, D. & Horvitz, E. (2010). Interactive optimization for steering machine classification. *SIGCHI Conference on Human Factors in Computing Systems (CHI '10). Association for Computing Machinery,* New York, NY, USA, 1343–1352. https://doi.org/DOI:https://doi.org/10.1145/1753326.1753529

Kim, B., Varshney, K. R. & Weller, A. (2018). *Workshop on Human Interpretability in Machine Learning (WHI).*

Kirkpatrick, K. (2017). It's not the algorithm, it's the data. *Communications of the ACM*, *60*(2), 21–23. https://doi.org/https://doi.org/10.1145/3022181

Köchling, A. & Wehner, M. C. (2020). Discriminated by an algorithm: a systematic review of discrimination and fairness by algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*, 795–848. https://doi.org/https://doi.org/10.1007/s40685-020-00134-w

Kuechler, W. & Vaishnavi, V. (2008). The emergence of design research in information systems in North America. *Journal of Design Research*, *7*(1), 1–16. https://doi.org/10.1504/JDR.2008.019897

Kulesza, T., Burnett, M., Wong, W. & Stumpf, S. (2015). *Principles of Explanatory Debugging to Personalize Interactive Machine Learning*. 126–137.

Kulesza, T., Stumpf, S., Burnett, M., Yang, S., Kwan, I. & Wong, W. (2013). Too much, too little, or just right? Ways explanations impact end users' mental models. *IEEE Symposium on Visual Languages and Human Centric Computing*, 3–10. https://doi.org/doi: 10.1109/VLHCC.2013.6645235

Liaw, A. & Wiener, M. (2002). Classification and regression by randomForest. *R News*, *2*(3), 18–22.

Liaw, A. & Wiener, M. (2018). *Package 'randomForest.'* Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/randomForest/randomForest.pdf

Lim, B. Y. & Dey, A. K. (2010). Toolkit to support intelligibility in context-aware applications. *UbiComp'10 - Proceedings of the 2010 ACM Conference on Ubiquitous Computing*, 13–22. https://doi.org/10.1145/1864349.1864353

Lim, B. Y. & Dey, A. K. (2011a). Design of an intelligible mobile context-aware application. *Mobile HCI 2011 - 13th International Conference on Human-Computer Interaction with Mobile Devices and Services*, 157–166. https://doi.org/10.1145/2037373.2037399

Lim, B. Y. & Dey, A. K. (2011b). Investigating intelligibility for uncertain context-aware applications. *Proceedings of the 13th International Conference on Ubiquitous Computing (UbiComp '11). Association for Computing Machinery, New York, NY, USA,* 415–424. https://doi.org/DOI:https://doi.org/10.1145/2030112.2030168

Lim, B. Y. Dey, A. K., & Avrahami, D. (2009). Why and why not explanations improve the intelligibility of context-aware intelligent systems. *Conference on Human Factors in Computing Systems - Proceedings*, 2119–2128. https://doi.org/10.1145/1518701.1519023

Lipton, Z. C. (2018). The mythos of model interpretability. *Communications of the ACM*, *61*(10), 35–43. https://doi.org/10.1145/3233231

Louppe, G., Wehenkel, L., Sutera, A. & Geurts, P. (2013). Understanding variable importances in forests of randomized trees. In *Advances in Neural Information Processing Systems 26*.

March, S. T. & Smith, G. F. (1995). Design and Natural Science Research on Information Technology. *Decision Support Systems*, *15*(4), 251–266. https://doi.org/10.1016/0167-9236(94)00041-2

Meth, H., Mueller, B. & 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. https://doi.org/10.1016/j.artint.2018.07.007

Montgomery, D. C., Jennings, C. L. & Kulahci, M. (2015). *Introduction to Time Series Analysis and Forecasting*. John Wiley & sons. https://books.google.ch/books?id=Xeh8CAAAQBAJ&dq=Time+Series+forecasting+in+R&lr=&hl=de&source=gbs_navlinks_s

Moody, J. (2019). *What does RMSE really mean?* Date Accessed: 20.02.2021. https://towardsdatascience.com/what-does-rmse-really-mean-806b65f2e48e

Nielsen, J. (1994). *Enhancing the Explanatory Power of Usability Heuristics*. 152–158. https://doi.org/10.1145/191666.191729

Oxborough, C., Rao, A., Cameron, E. & Westermann, C. (2018). *Explainable AI*. Date Accessed: 30.11.2020. https://www.pwc.co.uk/audit-assurance/assets/pdf/explainable-artificial-intelligence-xai.pdf

Panetta, K. (2017). *Gartner Top 10 Strategic Technology Trends for 2018*. Date Accessed: 30.12.2020. https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2018/

Peffers, K., Tuunanen, T., Rothenberger, M. & Chatterjee, S. (2007). A Design Science Research Methodology for Information Systems Research. *Journal of Management Information Systems*, *24*(3), 45–77.

Perrier, V., Meyer, F., Granjon, D. & Fellow, I. (2021). *shinyWidgets: Custom Inputs Widgets for Shiny*. Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/shinyWidgets/index.html

Pu, P. & Chen, L. (2006). *Trust Building with Explanation Interfaces*.

Pu, P., Chen, L. & Hu, R. (2012). *Evaluating recommender systems from the user ' s perspective : survey of the state of the art*. 317–355. https://doi.org/10.1007/s11257-011-9115-7

R Core Team. (2021). *The R Stats Package*. Date Accessed: 15.03.2021. https://doi.org/https://stat.ethz.ch/R-manual/R-devel/library/stats/html/00Index.html

Resnizky, H. G. (2015). *Learning Shiny*. Date Accessed: 12.12.2021. https://books.google.ch/books?id=P_1_CwAAQBAJ&dq=shiny+R+introduction&hl=de&source=gbs_navlinks_s

Ribera, M. & Lapedriza, A. (2019). Can we do better explanations? A proposal of user-centered explainable AI. *CEUR Workshop Proceedings*, *2327*.

RStudio. (2020a). *Shiny*. Date Accessed: 15.03.2021. https://shiny.rstudio.com/

RStudio. (2020b). *Shiny Gallery*. Date Accessed: 10.11.2020. https://shiny.rstudio.com/gallery/

RStudio. (2020c). *Welcome to Shiny*. Date Accessed: 15.03.2021. https://shiny.rstudio.com/tutorial/written-tutorial/lesson1/

Sajja, S., Aggarwal, N., Mukherjee, S., Manglik, K., Dwivedi, S. & Raykar, V. (2020). *Explainable AI based Interventions for Pre-season Decision Making in Fashion Retail*.

Date Accessed: 12.01.2021. http://arxiv.org/abs/2008.07376

Schaffer, J., Tobias, H., Jones, D. & Donovan, J. O. (2015). *Getting the Message ? A Study of Explanation Interfaces for Microblog Data Analysis*. 345–356.

Schlegel, U., Arnout, H., El-Assady, M., Oelke, D. & Keim, D. A. (2019). Towards a rigorous evaluation of XAI methods on time series. *Proceedings - 2019 International Conference on Computer Vision Workshop, ICCVW 2019*, *Ml*, 4197–4201. https://doi.org/10.1109/ICCVW.2019.00516

Scornet, E., Biau, G. & Vert, J. P. (2015). Consistency of random forests. *The Annals of Statistics*, *43*(4), 1716–1741.

Shmueli, G., Kenneth, C. & Lichtendahl, J. (2016). *Practical Time Series Forecasting with R: A Hands-On Guide [2nd Edition]*. Axelrod Schnall Publishers.

Shneiderman, B., Plaisant, C., Cohen, M., Jacobs, S. & Elmqvist, N. (2016). *Designing the User Interface: Strategies for Effective Human-Computer Interaction: Sixth Edition.* Pearson.

Smith-Renner, A., Rua, R. & Colony, M. (2019). Towards an explainable threat detection tool. *CEUR Workshop Proceedings*, *2327*.

Sundararajan, M., Xu, S., Taly, A., Sayres, R. & Najmi, A. (2019). Exploring principled visualizations for deep network attributions. *CEUR Workshop Proceedings*, *2327*.

Takeda, H., Veerkamp, P. & Yoshikawa, H. (1990). Modeling Design Process. *AI Magazine*, *11*(4), 37–37. https://doi.org/https://doi.org/10.1609/aimag.v11i4.855

Tintarev, N. & Masthoff, J. (2007). *A Survey of Explanations in Recommender Systems*. 801–810.

Turek, M. (n.d.). *Explainable Artificial Intelligence (XAI)*. Date Accessed: 10.11.2020. https://www.darpa.mil/program/explainable-artificial-intelligence

Tyralis, H. & Papacharalampous, G. (2017). Variable Selection in Time Series Forecasting Using Random Forests. *Algorithms*, *10*(4).

van der Maaten, L. & Hinton, G. (2008). Visualizing Data using t-SNE. *Journal of Machine Learning Research*, *9*, 2579–2605.

Vidal, T. & Schiffer, M. (2020). Born-Again Tree Ensembles. Proceedings of the 37th International Conference on Machine Learning. *Proceedings of Machine Learning Research.*, *119*, 9743–9753. Date Accessed: 12.01.2021. http://proceedings.mlr.press/v119/vidal20a.html

Vig, J., Sen, S. & Riedl, J. (2011). Navigating the tag genome. *In Proceedings Ofthe 16th International Conference on Intelligent User Interfaces*, 93–102.

Wachter, S., Mittelstadt, B. & Russell, C. (2018). Counterfactual explanations without opening the black box : automated decisions and the GDPR. *Harvard Journal of Law & Technology*, *31*(2), 1–52.

Weller, A. (2017). Transparency: Motivations and Challenges. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *11700 LNCS*(Section 2), 23–40. https://doi.org/10.1007/978-3-030-28954-6_2

Wickham, H. (2019). *stringr*. Date Accessed: 15.03.2021. https://www.rdocumentation.org/packages/stringr/versions/1.4.0

Wickham, H., Francois, R., Henry, L. & Mueller, K. (2021). *Package 'dplyr.'* Date Accessed: 15.03.2021. https://cran.rstudio.com/web/packages/dplyr/dplyr.pdf

Wickham, W., Henry, L., Pedersen, T. L., Takahashi, L., Wilke, C., Woo, K., Yutani, H. & Dunnington, D. (2020). *Package 'ggplot2.'* Date Accessed: 15.03.2021. https://cran.r-project.org/web/packages/ggplot2/ggplot2.pdf

Wierzynski, C. (2018). *The Challenges and Opportunities of Explainable AI*. Date Accessed: 12.01.2021. https://www.intel.com/content/www/us/en/artificial-intelligence/posts/the-challenges-and-opportunities-of-explainable-ai.html

Wilson, A. G., Kim, B. & Herlands, W. (2016). *Proceedings of NIPS 2016 Workshop on Interpretable Machine Learning for Complex Systems.*

Yang, R. & Newman, M. W. (2013). Learning from a learning thermostat: lessons for intelligent systems for the home. *In Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '13). Association for Computing Machinery,* New York, NY, USA, 93–102. https://doi.org/DOI:https://doi.org/10.1145/2493432.2493489

Zornoza, J. (2020). *Explainable Artificial Intelligence Discover one of the biggest trends in Machine Learning and AI*. Date Accessed: 12.12.2020. https://towardsdatascience.com/explainable-artificial-intelligence-14944563cc79

# Appendix

## A. Documentation of Literature Review

### A.1. Design Principles found in the Literature

| **Design Principles for XAI - Post Hoc Explanation** | Ribera & Lapedriza, 2019 | Kulesza et al., 2015 | Zornoza, 2020 | Fiddler.ai, 202) | Oxborough et al., 2018 | Sajja et al., 2020 | Dudley & Kristensson, 2018 | Hagras, 2018 | Wachter, Mittelstadt & Russell, 2018 |
|---|---|---|---|---|---|---|---|---|---|
| Added Layer of Interpretability | x | x | x | | | | x | x | |
| Brief, plain Language; Easy understandable | x | x | | | | | x | x | |
| Understandable, easy Visualization | | | | x | | | x | | |
| Interactivity | | x | | | | | x | | |
| Ease of Use / Consistency | | x | | | | | | | |
| Backtracing / Reversing Steps | | x | | | | | x | | |
| User has control | | | | | | | | | |
| Feature / Factor importance | | | | x | x | x | | | |
| Counterfactual Explanations | x | | | | | x | | | x |
| Swift Execution Times | | x | | | | | x | | |
| What-If Scenarios | | | | x | | | | | |
| If-Then Rules | | | | | | | | x | |
| Multiple Time Horizons | | | | | | x | | | |
| Correlated Input Features | | | | | x | | | | |
| Type of Prediction | | | | | x | | | | |
| Problem Domain | | | | | x | | | | |

| **Design Principles for XAI - Post Hoc Explanation** | Smith-Renner et al., 2019 | Sundararajan et al., 2019 | Chromik, Völkel, Eiband & Buschek, 2019 | Bunt et al., 2012 | Lim & Dey, 2011a | Vig, Sen, & Riedl, 2011 | Pu & Chen, 2006 | Pu et al., 2012 | Tintarev & Masthoff, 2007 |
|---|---|---|---|---|---|---|---|---|---|
| Added Layer of Interpretability | x | x | x | x | x | | x | x | x |
| Brief, plain Language; Easy understandable | | x | x | x | x | x | x | x | x |
| Understandable, easy Visualization | | x | x | | x | | | x | |
| Interactivity | x | | | | | | | x | x |
| Ease of Use / Consistency | | x | | | x | | | | x |
| Backtracing / Reversing Steps | | | | | x | | | | |
| User has control | | | | | x | | | | x |
| Feature / Factor importance | | | | | x | | | | |
| Counterfactual Explanations | | | | | | | | | |
| Swift Execution Times | | | | | | | | | |
| What-If Scenarios | | | | | x | | | | |
| If-Then Rules | | | | | x | | | | |
| Multiple Time Horizons | | | | | | | | | |
| Correlated Input Features | | | | | | | | | |
| Type of Prediction | | | | | | | | | |
| Problem Domain | | | | | | | | | |

| **Design Principles for XAI - Post Hoc Explanation** | Babich, 2019 | Shneiderman et al., 2016 | Nielsen, 1994 | Schaffer, Tobias, Jones, & Donovan, 2015 | **Total** | This Thesis |
|---|---|---|---|---|---|---|
| Added Layer of Interpretability | x | x | x | x | 17 | x |
| Brief, plain Language; Easy understandable | x | x | x | x | 16 | x |
| Understandable, easy Visualization | x | x | x | | 9 | x |
| Interactivity | x | x | x | | 8 | x |
| Ease of Use / Consistency | x | x | x | | 7 | x |
| Backtracing / Reversing Steps | x | x | x | | 6 | |
| User has control | x | x | x | x | 6 | x |
| Feature / Factor importance | | | | | 4 | x |
| Counterfactual Explanations | | | | | 3 | |
| Swift Execution Times | | x | | | 3 | |
| What-If Scenarios | | | | | 2 | |
| If-Then Rules | | | | | 2 | |
| Multiple Time Horizons | | | | | 1 | |
| Correlated Input Features | | | | | 1 | |
| Type of Prediction | | | | | 1 | |
| Problem Domain | | | | | 1 | |
| Comparability | | | | | 0 | x |

## A.2. Consolidation of Design Principles

### Overview of Design Principles from Literature Review

**Overview of Design Principles**

| Design Principles for XAI - Post Hoc Explanation | Total Citations in Literature Review N=22, Multiple responses possible |
|---|---|
| Added Layer of Interpretability | 17 |
| Brief, plain Language, easy Understandable | 16 |
| Understandable, easy Visualization | 9 |
| Interactivity | 8 |
| Ease of Use / Consistency | 7 |
| Backtracing / Reversing Steps | 6 |
| User has Control | 6 |
| Feature / Factor Importance | 4 |
| Counterfactual Explanations | 3 |
| Swift Execution Times | 3 |
| What If Scenarios | 2 |
| If Then Rules | 2 |
| Multiple Time Horizons | 1 |
| Correlated Input Features | 1 |
| Type Of Prediction | 1 |
| Problem Domain | 1 |

**Overview of consolidated Design Principles**

| Consolidated Design Principles | Total Mentions in Literature Review |
|---|---|
| Ease of Use (incl. Consistency / Swift Execution times) | 35 |
| Interactivity (incl. User has control, What-if Scenarios, If-Then rules) | 18 |
| Added Layer of interpretability | 17 |
| Backtracing / Reversing Steps | 6 |
| Feature / Factor Importance (incl. Correlated Input Features) | 5 |

# B. Documentation of Expert Interviews

## B.1. List of Expert Interviews

| Name | Position | Date |
|---|---|---|
| Eric Meier | Junior Sous Chef, Guarda Val | 3rd of November 2020 |
| Michaela Frank | Chef de Partie, Restaurant Wart | 5th of November 2020 |
| Thomas Betzl | Chef & Owner, Restaurant “kleine Kappel” | 6th of November 2020 |
| Jessica Svahn Bold | Management Division G, University of St. Gallen | 10th of November 2020 |
| Michael Remus & Maik G. | Manager, HSG Mensa<br>Team Leader Kitchen, HSG Mensa | 25th of November 2020 |
| Anonymous | Head of Gastronomy Management, Large University of Applied Sciences in Switzerland | 30th of November 2020 |
| Doris Vögeli | Chef, Paul Scherrer Institute | 1st of December 2020 |

## B.2. Interview Guidelines

University of St.Gallen

# Principles for the creation of a digital UI

**Name**
**Function**

Philipp Schmelzer, Master's thesis

[Date]

### AGENDA

**The interview process**

- Welcome and round of introductions
- Duration of the interview
- Recording the interview
- Planned procedure for the interview

**Recording the interview**

**15-20 min**

**The interview**

- Presentation of the initial situation
- Background of the interlocutor
- Current demand calculation in the company
- Potential for improvement in current demand planning
- Visualisation of the demand planning
- Other aspects of a demand planning tool
- Conclusion of the interview

### MY PERSON / INITIAL SITUATION

**Presentation of the Master's thesis topic**

- Programming of predictive models that calculate the demand for dishes and food.
- Converting this data into understandable "language".
- Develop a visualisation or user interface to present these results in a usable way.
- Derive design principles that underpin this transformation and help the end user understand the data.

Objective for this interview: **Requirements** and **needs of end users** in daily use.

### BACKGROUND OF THE INTERLOCUTOR

Goal: Gain an understanding of the person's history, experience and position in the organisation.

- How long have you been working in gastronomy / in your business?
- What are their tasks (apart from the actual cooking)?
- Is purchasing / demand planning part of your field of activity?

### CURRENT CALCULATION OF DEMAND IN THE COMPANY

Goal: Gain an understanding of the current demand planning and purchasing of the different dishes.

- How does the current demand planning of dishes / food work?
- How does the demand planning process work?
- Is there currently an overview showing what needs to be purchased and in what quantity?
- What other information is important in demand planning?
- What other elements are essential to consider for demand planning?

### POTENTIAL FOR IMPROVEMENT CURRENT DEMAND PLANNING

Goal: Identify the optimal process for needs-based demand planning.

- What is currently in need of improvement in demand planning?
- What could be simpler in your demand planning?
- Are there any wishes, suggestions, potential for improvement?

## VISUALISATION OF THE DEMAND PLANNING

Goal: Gain an understanding of how to use a user-friendly and easy-to-understand visualisation.

- Is it more crucial for you to receive an explanation of how the algorithm works, or one that explains the recommendation currently being made?
- What information must be integrated on a visualisation for demand planning?
- How should the visualisation be designed to add the most value? (show examples)
- Which (visual) elements should not be missing from a user interface for demand planning?
- Are there any particular preferences about the type of visualisation or other needs you would like to add?

## OTHER ASPECTS OF A DEMAND PLANNING TOOL

Goal: Understand problems and limitations of data use.

- How important is it to you that you can understand the tool's suggestions?
- What elements would such a tool need to have in order for you to trust its recommendation?
- Would an interactive "What If..." function that can simulate different scenarios?
- To what extent would different time horizons need to be considered because food has different shelf lives or because of seasonality, fluctuations in demand, etc.?
- Must it be possible to exclude individual dishes or other factors, e.g. (exclusion of certain foods due to allergies or taste preferences, for seasonal dishes, etc.)?
- Would it be helpful if recommendations in the sense of a "recommender system", which gives suggestions for other dishes, were integrated?
- What other functions would a court demand planning tool need to have?

## CONCLUDING QUESTIONS

Goal: Clarify other questions on the topic.

- Do you have any tools or applications in your company that work with machine learning or AI?

## B.3. Summaries of Interviews

### Interview with Eric Meier on 3 November 2020

**Eric Meier**
**Junior Sous Chef,**
**Guarda Val**

**Derived design principles:**

1 Added Layer of Interpretability

2 Ease of Use

3 Interactivity

4 Comparability

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Explanation must be designed in such a way that no PC expertise is necessary, preferably in the local mother tongue and in simplified language.
- Simple, clear, concise structure of the app and use, no excessive demands due to details
- Concise, high-quality explanations in pop-up windows or separately so that the window looks not crowded.
- The menu planning should be based on numbers, the reports visually represented with graphs
- Quick familiarisation must be possible, no time-consuming functions

**Functionalities that support the credibility of an app:**
- Interactivity for the user, possibility of customisability of windows and elements
- Individual adjustments to the forecast by the user possible
- Comparison of the forecast with different figures e.g. customers, previous year, previous week, reservations
- Scenarios simplify familiarisation and increase the user's understanding

**Potential for further development**
- Recommender Systems
- Monthly reports on consumption, purchasing, turnover, profit
- Integration of a supplier view and order overview
- Integration of shelf life and storage system

### Interview with Michaela Frank on 5 November 2020

**Michaela Frank**
**Chef de Partie,**
**Restaurant Wart**

**Derived design principles:**

1 Added Layer of Interpretability

2 Ease of Use

3 Interactivity

4

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Very simply designed, "foolproof", visually designed
- Proposals are based on the algorithm and the decision-making process must therefore be explainable and understandable
- Explanations are important, less is more, a lot of text overwhelms, keyword-like
- Familiarisation period: learning by doing

**Functionalities that support the credibility of an app:**
- Interactivity: User has control, can give input back
- Feedback loops by the user
- Scenario calculations for throughput of the canteen / restaurant and walk-in customers

**Potential for further development**
- Integration of recommender systems, collection of ideas, suggestions, menu documentation, photos
- Integration of break even point calculation per dish, batch sizes, portion sizes and minimum order quantities
- Reminder, the order will expire soon or go bad

### Interview with Thomas Betzl on 6 November 2020

**Thomas Betzl**
**Chef & Owner,**
**Restaurant little Kappel**

**Derived design principles:**

1 Ease of Use

2 Added Layer of Interpretability

3

4

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Algorithm cannot replace a chef's gut feeling and wealth of experience
- Quick to learn, easy to use, clear and concise visualisations, not overwhelming, reduced to the core functions, must increase productivity in everyday life and not slow it down.

**Functionalities that support the credibility of an app:**
- Possibility to add new menus / products and delete old ones -> interactivity in the app, user has control over the app
- Does not need interactive gimmicks, must save time in everyday life and not eat up time

**Potential for further development**
- Linking with invoicing tool and supplier systems
- Inventory list, what stock do I have, monthly retrievable

## Interview with Jessica Svahn Bold on 10 November 2020

**Jessica Svahn Bold**
**Management Division G, HSG**

**Derived design principles:**

1. Added Layer of Interpretability
2. Factor Importance
3. Ease of Use
4. Interactivity

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Good / bad weather -> seasonalities, customer needs, offering different scenarios, weather, university schedule, special events at the HSG, demand for individual products based on specific events, discount promotions
- Little and easy to understand text, work with pictures

**Functionalities that support the credibility of an app:**
- Interaction with the tool very important, user must have control
- Possibility to plan different scenarios, customise menus, etc.

**Potential for further development**
- Personnel calculation
- Turnover overview
- Integration storage system

## Interview with Michael Remus and Maik G on 25 November 2020

**Michael Remus**
**Manager, HSG Mensa**
**Maik G.**
**Team Leader Kitchen, HSG Mensa**

**Derived design principles:**

1. Factor Importance
2. Added Layer of Interpretability
3. Comparability
4. Interactivity

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Comparison to previous year, explanation of figures, show factors that play a role, credibility through comparison to previous year
- Explanations Icon based, weather, seasonality, guest volume, terrace open / closed, previous year's comparison, special university events.
- Text as short and simple as possible

**Functionalities that support the credibility of an app:**
- Menu with the individual quantities, what are the influencing factors? Previous year's figures
- Value of the menus (much sought-after menus vs. little sought-after menus with stars)
- Integration total menu number
- Interactions, icons editable, add icons to the menu plan yourself, adjust values, make tools usable, menu predictions editable (e.g. quantity required)
- Need for menus per week / day and number per menu
- Where are we right now in the demand curve (Year or month)

**Potential for further development**
- Recommender system with focus on seasonal products
- Integration of the recipes and automatic generation of an order list from the forecasts

## Interview with Anonymous on 30 November 2020

**Anonymous**
**Head of gastronomy management, Big University of Applied Sciences in Switzerland**

**Derived design principles:**

1. Added Layer of Interpretability
2. Comparability
3. Ease of Use
4. Interactivity

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

[X] Explanations about the functionality [ ] No explanations necessary

Explanation of the algorithm should be done in this way:

[X] Post-hoc [ ] Intrinsic

**Factors that make an explanation comprehensible:**
- Comparability with the previous year increases plausibility and credibility
- Comparison with total menus previous year
- Explanations: less detailed, icon-based, clearly presented and simply explained, usable by everyone

**Functionalities that support the credibility of an app:**
- Preview of at least 2 weeks
- Evaluation: How much food was sold?
- Interactivity: Data input by user to further train prediction
- Interactive menu planning, adaptations by the user possible

**Potential for further development**
- Automatic reconciliation with the cash register system

**Interview with Doris Vögeli on 1 December 2020**

**Doris Vögeli**
**Chef,**
**Paul Scherrer Institute**

**Derived design principles:**

1. Added Layer of Interpretability
2. Ease of Use
3. Interactivity
4. Factor Importance

**Evaluation of the interview questions**

Do you need explanations for the comprehensibility / trust in the app?

☒ Explanations about the functionality ☐ No explanations necessary

Explanation of the algorithm should be done in this way:

☒ Post-hoc ☐ Intrinsic

**Factors that make an explanation comprehensible:**
- Factors explaining the forecast (Special events, date, day of the week, season, weather, etc.)
- Explanations brief and to the point, no technical language
- Explanations must be intuitively understandable and clear
- Structure of the app must be simple and clear and not overwhelming

**Functionalities that support the credibility of an app:**
- Comparison with historical data
- Forecast directly for the whole week for comparability
- Interactivity and different input fields
- Individual adjustments to the forecast possible

**Potential for further development**
- Deposit of the recipes and the sales price
- Training the algorithm with new data

## B.4. Adapted Design Principles

### Overview of Derived Design Principles

| Design Principles | Eric Meier | Michaela Frank | Thomas Betzl | Jessica Svahn Bold | Michael Remus & Maik G. | Anony-mous | Doris Vögeli | Total Count of Mentions |
|---|---|---|---|---|---|---|---|---|
| Added Layer of Interpretability | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 |
| Ease of Use | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 6 |
| Interactivity | 1 | 1 | 0 | 1 | 1 | 1 | 1 | 6 |
| Comparability | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 3 |
| Factor Importance | 0 | 0 | 0 | 1 | 1 | 0 | 1 | 3 |
| Backtracing | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |

# C. Documentation of the Design Workshop

University of St.Gallen

## Design workshop
### Prototyping a user-centric UI

Michaela Frank, Chef de Partie, Restaurant Wart
Eric Meier, Junior Sous Chef, Hotel Guarda Val

December 3, 2020

Master thesis, Philipp Schmelzer

### Coordinates

| | | |
|---|---|---|
| **Targets** | The participants have<br>• understood the basic design principles of developing a user-centric UI.<br>• understood the prepared elements for prototyping and the underlying functions and know what they mean.<br>• collected ideas on how a user interface to visualize a demand forecast for system catering could look like.<br>• have created several prototypes of the visualization. | |
| **Date and time** | Thursday, December 3, 2020, 1:00 - 2:00 p.m. | |
| **Location** | Hertensteinstrasse 28, Nussbaumen | |
| **Participant** | **Participant**<br>▪ Michaela Frank<br>▪ Eric Meier | **Host**<br>▪ Philipp Schmelzer |

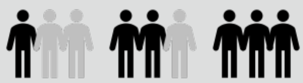
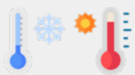
### Program and schedule of the design workshop

| Start | T | Topic |
|---|---|---|
| 13:00 | 5‘ | Huddle: Explanation of the objectives and the planned process of the workshop. |
| 13:05 | 10‘ | Explanation of the functions of the previously developed elements and icons |
| 13:15 | 10‘ | Explanation of the underlying design principles |
| 13:20 | 10‘ | Brainstorming about possible design forms of the user interface |
| 13:30 | 20‘ | Prototyping of the user interface in two rounds |
| 13:50 | 5‘ | Final huddle and farewell |

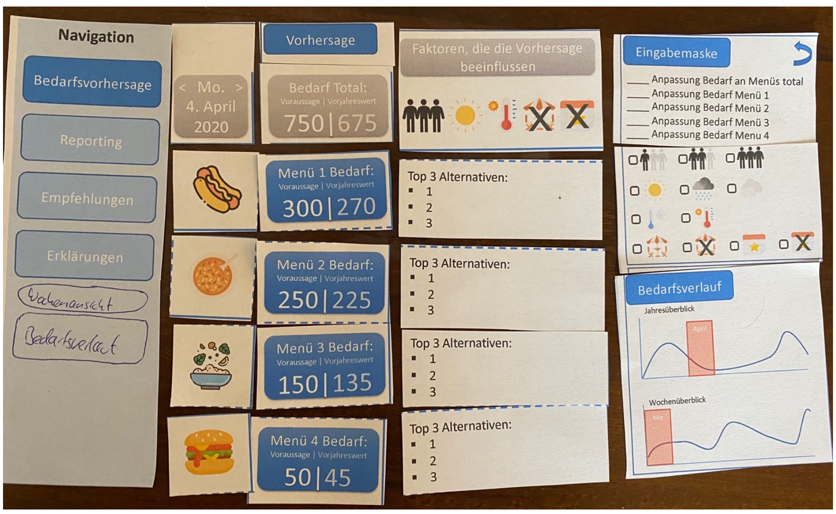
## Various elements for prototyping

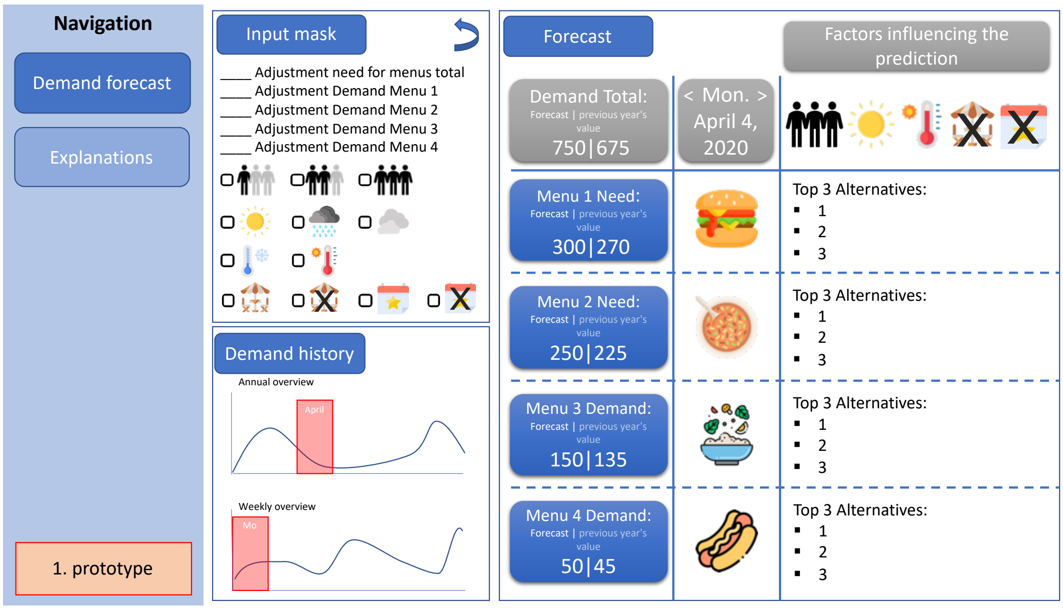

5

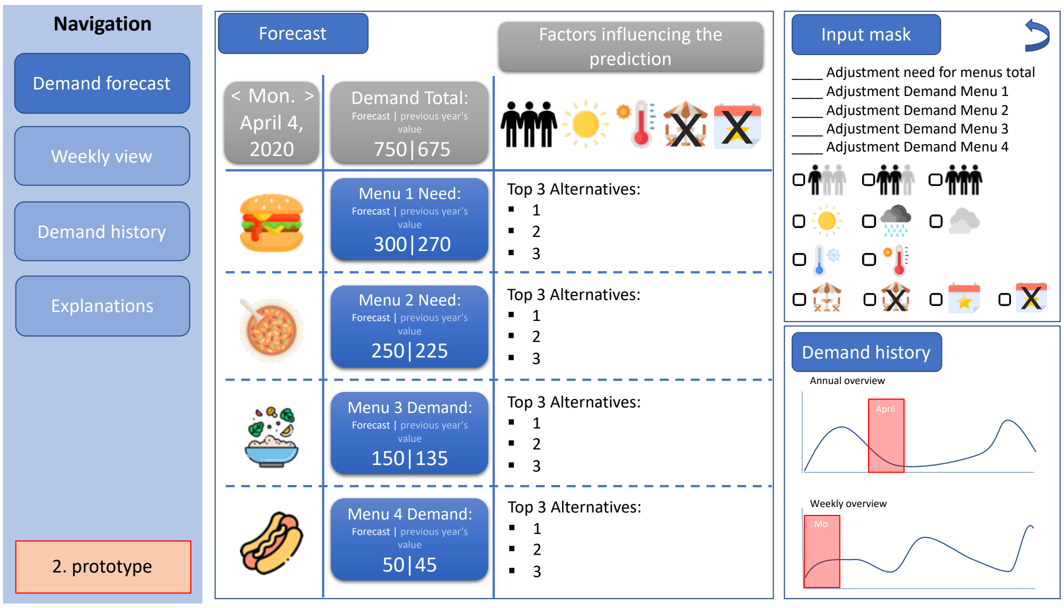
## Icons for visualization

| Icon | Meaning | Interpretation |
|---|---|---|
|  | Guest volume:<br>Low \| Medium \| High | Low: Forecast smaller than previous year's value<br>Mean: Forecast equal to previous year's value<br>High: Forecast higher than previous year's value |
|  | Season:<br>Cold \| Warm | Cold: Tends to be the same number of guests<br>Warm: Tends to be fewer guests, as consumption also takes place elsewhere |
|  | Weather:<br>Sunny, Rainy, Cloudy | Sunny: tends to fewer guests<br>Rainy: tends to be more guests<br>Cloudy: tends to be the same number of guests |
|  | Exterior:<br>Opened: Yes / No | If the outdoor area is open, guests consume more |
|  | Special event coming up:<br>Yes / No | Special events increase the need for menus<br>No: Need menus remains the same |

## First prototype of the visualization





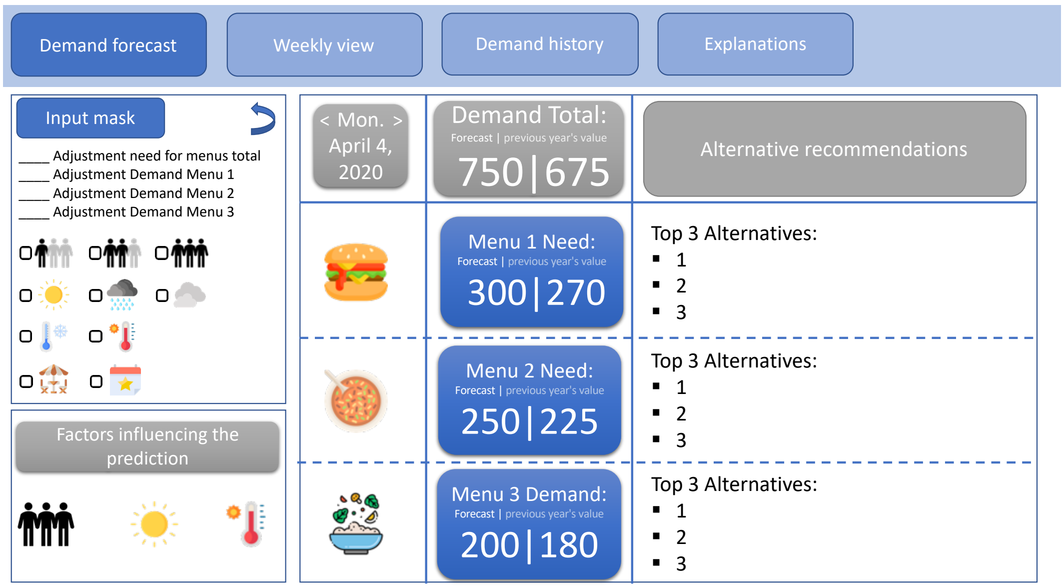



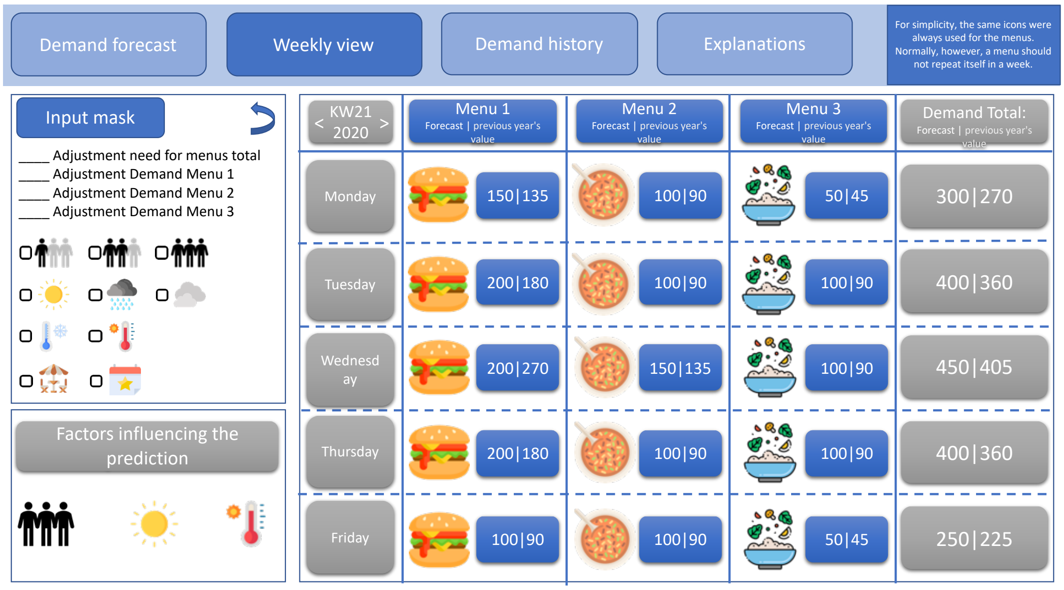

## Learnings after the first iteration

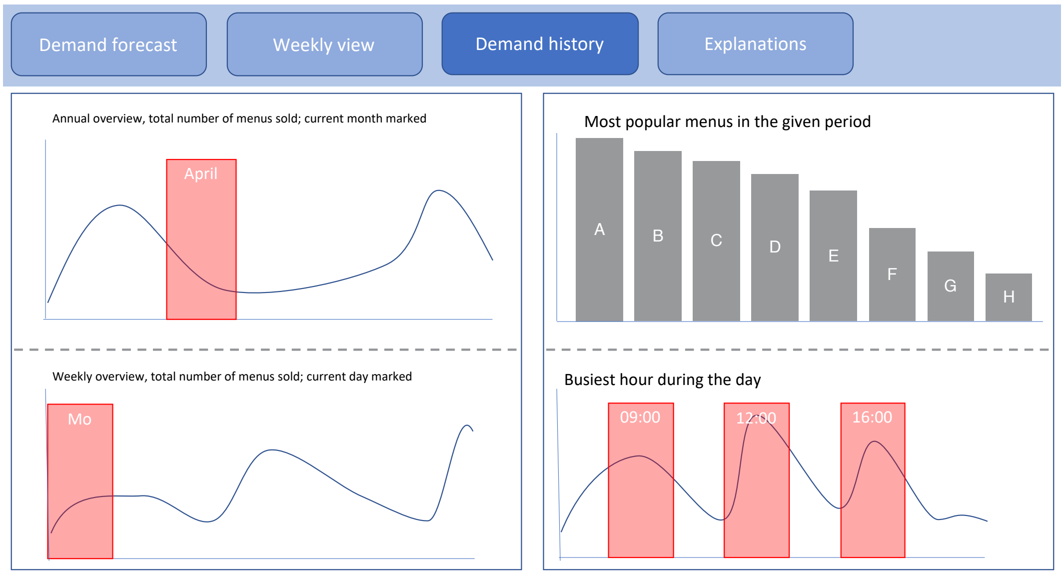
- Presentation is visual and simple, but sometimes a bit crowded
- To customize this you could move the demand history and reporting to a separate window.
- A weekly view would be very interesting
- Icons are so far well understandable, but need a short introduction -> in the explanation bar
- Icons are appropriate in number, the icon for "no outdoor operation" and "no special event" could be omitted
- All icons for weather sun, cloudy and rain important e.g. because of need for soups important
- Interactivity is adequate, it is important to be able to adjust the individual menus and specific factors afterwards
- It would be good if you could integrate a dropdown above the menu icons, where you could exchange the menus

2nd Prototype Variant incl. Subwindow


2nd Prototype Variant incl. Subwindow
The "Week view" subwindow is missing





- The forecast can be changed and adjusted to the needs by means of the icons and the input field, should a forecast not meet the wishes.
- So you can use the input mask, the total number of menus or the required number of menus 1 to 3 change and thus adjust the forecast.
- The arrow at the top right of the input mask acts as an "Undo" or "Step back" command and undoes the last action.
- Next to the selected day there are two buttons, with the help of which the predicted day or week (maximum 28 days into the future) will be displayed.
- The "factors influencing the prediction" are broken down in even more detail in terms of their importance on the right-hand side.
- The menu that generated the highest value of customers in the prediction is always displayed as the menu recommendation. The alternative variants are the three options that follow.

| Icon<br>The model is adjusted depending on the activated icon | Meaning | Interpretation |
|---|---|---|
| | Guest volume:<br>Low \| Medium \| High | Low: Forecast smaller than previous year's value<br>Mean: Forecast equal to previous year's value<br>High: Forecast higher than previous year's value |
| | Season:<br>Cold \| Warm | Cold: Tends to be the same number of guests<br>Warm: Tends to be fewer guests, as consumption also takes place elsewhere |
| | Weather:<br>Sunny, Rainy, Cloudy | Sunny: tends to fewer guests<br>Rainy: tends to be more guests<br>Cloudy: tends to be the same number of guests |
| | Exterior:<br>Opened: Yes / No | If the outdoor area is open, guests consume more |
| | Special event coming up:<br>Yes / No | Special events increase the need for menus<br>No: Need menus remains the same |

## Learnings after the second Iteration

- The new navigation arrangement is intended to create more space on the screen
- The subdivision into different windows is intended to streamline the visualization on the individual windows and make it clearer for the end user.
- It is usually sufficient to predict only three menus, since very few canteens or system restaurants offer four different menus in parallel.
- The recommendations should remain in the first view, unlike in the displayed image. Since the positioning of the content there allows at a glance to find out about potential alternatives of the daily program.
- Adding a similar feature was not done in the weekly overview, as this would generate even more information to be mapped and thus further limit readability.

## Potential for improvement - Design Workshop

- The elements brought along set a tremendous anchor for the two participants during prototyping. For a next iteration of such a workshop, it would be better to introduce the elements only in the second iteration of prototyping
- The participants strongly influenced each other's creativity. In the next iteration, care must be taken, in the sense of brainwriting, to let the two participants each first develop an independent prototype. Afterwards, the ideas can be exchanged and then a prototype can be developed together.
- The chefs both had some experience in using mobile devices. In the sense of a sophisticated development of the visualization, it would have been interesting to have had an older and not so digital-savvy chef in the workshop.

# D. Documentation of Evaluation Interviews

## D.1. List of Evaluation Interviews

| Name | Position | Date |
|---|---|---|
| Eric Meier | Junior Sous Chef, Guarda Val | 18th of December 2020 |
| Jessica Svahn Bold | Management Division G, University of St. Gallen | 23rd of December 2020 |
| Michael Remus | Manager, HSG Mensa | 23rd of December 2020 |
| Michaela Frank | Chef de Partie, Restaurant Wart | 26th of December 2020 |
| Doris Vögeli | Chef, Paul Scherrer Institute | 21st of January 2021 |

## D.2. Interview Guidelines

## AGENDA

### The interview process

- Welcome
- Duration of the interview
- Recording the interview
- Planned procedure for the interview

**Recording the interview**

**10-15 min**

### The interview

- Goals of the interview
- Testing the Shiny web app
- Functionality evaluation
- Conclusion of the interview

## GOALS OF THE INTERVIEW

**Goals for the interview**

- Testing the functionality of the developed web application to see if it is suitable for daily use.
- Checking the application for various factors such as comprehensibility of results, trustworthiness of recommendations, ease of use, etc.
- Obtain feedback for the further development of the application.

Guiding question for this interview:

Does the application match the **requirements** and **needs of** the **end users** for daily use?

## EXPLANATION OF THE APPLICATION

Goal: Checking the design and structure of the app for various factors

- Link to the webapp: https://philipp-schmelzer-hsg.shinyapps.io/shiny/

**Brief explanation of the application**

- Three menu predictions for the next 28 days based on historical data and other factors
- Demand forecast for one or for five selected days
- Demand profile of the different menus offered
- Explanations describing in detail how the application works
- Input mask with which the date can be selected and the need for menus can be adjusted.

## TESTING THE APPLICATION

Goal: Checking the design and structure of the app for various factors

**The interviewee is to perform the following operations on the app:**

- Review of the proposed three menus on a daily and weekly basis
- Selection of a specific day / 5 days
- Adaptation menus
- Adjustment of the need for a menu
- Understanding the icons?
- Understanding the progression of needs?
- Changing the menus viewed in the course of demand
- Assessment of the declarations

## EVALUATION OF THE FUNCTIONALITY

Goal: Receive feedback on usability in daily business

| Factor to be assessed | Rating 1-5 |
|---|---|
| Clarity of the application | |
| Simplicity of the structure | |
| Simplicity of operation | |
| Comprehensibility of the explanation | |
| Credibility of the predictions | |
| Comprehensibility of the predictions | |
| Understandability of the predictions | |
| Usefulness in everyday life | |
| Usability in everyday life | |

## EVALUATION OF THE FUNCTIONALITY

Goal: Receive feedback on usability in daily business

| Question | Yes | No |
|---|---|---|
| To what extent do you trust the predictions of the application? | | |
| To what extent can you understand the predictions? | | |
| Do the predictions seem realistic to you? | | |
| Do the predictions seem credible to you? | | |
| Are the explanations in the app understandable for you? | | |
| Is the operation of the app intuitive for you? | | |
| Could you imagine using such an application for menu planning? | | |

## FEEDBACK FOR FURTHER DEVELOPMENT

Goal: Receive feedback for further development

**Potential for further development**

•

## CONCLUSION OF THE INTERVIEW

Goal: Clarification of final questions and thanking the interview partner

- Do you know of any other people who would be willing to be interviewed?

### D.3. Summary of the Evaluation Interviews

| Person | Feedback |
|---|---|
| Eric Meier | First familiarisation somewhat difficult, a short introduction would be helpful, step by step introduction<br>Add up the menu in the weekly overview, total sum<br>Numbers in the graphs and predicted value integrated in the graphs to have a comparison |
| Jessica Svahn Bold | Integration of an Instagram interface for advertising<br>For bar operation would need a slightly different version<br>Caution should be exercised when selecting the comparison period. With the current Corona year, there could be strong distortions.<br>The first time you open the app it would be good to get a little tour. Otherwise, you have to find out for yourself how everything works |
| Michaela Frank | Name current demand differently<br>Explanation of previous year's demand must be based on more variables than just 1 year in the past.<br>Menu selection sounds strange when products are suggested<br>Usage in everyday life is conceivable if other functions are added, e.g., coupling with the reservation system, or ordering and storage system |
| Michael Remus | Demand history: Show current values with comparative values<br>Explain guest volume icons in more detail (what is the icon based on, write that it depends on the guest volume of the weekdays) |
| Doris Vögeli | Keep it simple - keywords.<br>Explanatory tour would be very good for first time opening |

### D.4. Analysis of Feedback of the Evaluation Interviews

| Feedback | Mentions in Interviews |
|---|---|
| A step-by-step introduction to simplify the introduction to the functionalities. | 3 |
| Add forecast value to the graphs to see a simple comparison | 2 |
| "Demand previous year" but several factors and name them differently | 2 |
| Show total demand in the weekly overview | 1 |
| Show numbers on the graphs to increase comprehension | 1 |
| Integrate an Instagram interface for advertising | 1 |
| Alternative version for a bar operation with adapted functions | 1 |
| Rename "current demand" to a more comprehensible version | 1 |
| Insert explanation for "Demand previous year" in the explanations | 1 |
| Menu selection the wrong name, only products are currently displayed | 1 |
| Use in everyday life is conceivable if further functions are added, e.g. coupling with the reservation system, or ordering and storage system. | 1 |
| Explain guest volume icons in more detail | 1 |

## E. Documentation of R and Shiny Code

The dataset described in Chapter 3.3, the time series prediction model described in Chapters 3.4 and 3.5, and the Shiny implementation described in the 1st Cycle of Prototyping are stored in a separate GitHub repository. Additionally, the documentation of the design workshop and the expert and evaluation interviews is available.

Under the following link you will find the source code as well as a Read.Me file for the exact instructions to get the software running:

https://github.com/Philipp-Schmelzer/Master-Thesis

The web application is already online and can be accessed by using this link:

https://philipp-schmelzer-hsg.shinyapps.io/shiny/

However, due to the upload policy of shinyapps.io, the website may have been taken offline again due to inactivity.

## Declaration of authorship

I hereby declare

- that I have written this thesis without any help from others and without the use of documents and aids other than those stated above;
- that I have mentioned all the sources used and that I have cited them correctly according to established academic citation rules;
- that I have acquired any immaterial rights to materials I may have used such as images or graphs, or that I have produced such materials myself;
- that the topic or parts of it are not already the object of any work or examination of another course unless this has been explicitly agreed on with the faculty member in advance and is referred to in the thesis;
- that I will not pass on copies of this work to third parties or publish them without the University's written consent if a direct connection can be established with the University of St.Gallen or its faculty members;
- that I am aware that my work can be electronically checked for plagiarism and that I hereby grant the University of St.Gallen copyright in accordance with the Examination Regulations in so far as this is required for administrative action;
- that I am aware that the University will prosecute any infringement of this declaration of authorship and, in particular, the employment of a ghostwriter, and that any such infringement may result in disciplinary and criminal consequences which may result in my expulsion from the University or my being stripped of my degree.

St. Gallen, May 25th 2021

P.Schmelzer

Philipp Schmelzer