DEVELOPMENT AND EVALUATION OF AN [CUSTOMIZABLE MOBILE] APPLICATION FOR BEHAVIORAL RESEARCH IN DATA ANALYTICS

Master Thesis
of

MAX DARMSTADT

July 10, 2023

Matriculation Number
1820000

Submitted at the Chair of Enterprise Systems
University of Mannheim

Reviewer: Prof. Dr. Hartmut Höhle
Supervisor: Jan Schilpp

Contents

List of Figures

List of Tables

List of Abbreviations

DSC	Design Science Research
IEEE	Institute of Electrical and Electronics Engineers
ISO	International Organization for Standardization
IT	Informationstechnologie

Abstract

Over the past few years, data analytics has become increasingly important for companies across all industries. With the massive amount of data that is now available, companies can use data analytics to gain valuable insights into consumer behavior, market trends, and internal operations, among other things. As a result, data analytics has become a critical tool for companies looking to gain a competitive edge in today's rapidly evolving business environment. However, while data analytics has become an essential tool for businesses, there has been relatively little research done in the area of behavioral research. Specifically, there is a lack of research on the decision-making process involved in data analytics, and how individuals and organizations use data analytics to inform their decisions. One of the major challenges in conducting research in this area is the high cost of developing custom applications for each study. The development of such applications can be time-consuming, expensive, and often requires specialized expertise. To address this challenge, this master thesis develops a generic application that streamlines the process of conducting studies in the field of data analytics. This application enables researchers to design, conduct, and analyze studies more efficiently and cost-effectively, allowing them to explore the field in greater depth. This will be accomplished by using the design science research approach. Firstly, the problem of a lack of behavioral research in data analytics is identified. Then, the objectives for a solution are defined through a literature review and the use of requirement engineering to gather requirements for the application. Next, the application is design, implemented prototypically and its functionality demonstrated. Finally, solution is evaluated through the usages of the requirements.

1 Introduction

1.1 Background and Motivation

The introduction and widespread usage of computers has proven to be disruptive for all industries. Entire industrial sectors have been reshaped, whole professions made obsolete and new career opportunities have been created. This shift towards the adaption of Informationstechnologie (IT) has been necessary for businesses to stay competitive in the fast changing economic environment of the twenty-first century. Nowadays, the digitalization of organizations is viewed as a prerequisite for a successful business, rather than being an endeavored state and computers are indispensable for all industries. This digital revolution enabled the emergence of the widespread creation and collection of data. The amount of data generated globally is rising yearly (Seagate, 2018) and the pressure to use these data volumes effectively in order to gain a business advantage rises. This new trend, often coined "big data" after the fact that never before seen amounts of data are generated and are available for processing, enables completely new business areas. This is reinforced, among other things, by the fact that many companies already view their data as a primary business asset (Redman, 2008). Simultaneously, the emergence of big data promises to completely reshape the decision-making process of traditional businesses through the adoption of data analytics. Although sales in the area of big data have risen significantly over the past years (BIS Research, 2018, Bitkom, 2018) and businesses already view big data as an important information technology trend (Bitkom, 2017) a lot of organizations struggle to effectivly utilize their data.

In their article, Amankwah-Amoah and Adomako study the influence of big data usage on business failure. They come to the conclusion that the mere possession of big data as an asset has no positive effects on an organization. In order to prevent business failure, big data must be used effectively (Amankwah-Amoah and Adomako, 2019). This, however, might be hindered by boundaries and conflicts that have arisen during an organization's existence. Technical challenges are only some part of the underlying problem. Boundaries that have surfaced due to behavioral aspects regarding these new technologies might also impact the effective usage of data analytics. This problem only becomes more apparent when new technologies which overlap with data analytics like machine learning and artificial intelligence

are taken into account. These technologies, often concidered to be the next big thing which has its origins in said increase data collection, are already gaining a lot of the attention in both research and businesses. Although these ”black box” technologies are even less explainable than more traditional big data and data analytics technologies, the focus on research clearly relies on technical apsects instead of behavioral boundaires and conflicts that might hinder the effectiveness of these technologies. Should corporate executives blindly rely on data analytics and machine learning for their decision making process or just use them as a support tool? If a senior manager has a different opinion on a product release than what is specified by data analytics processes, who's opinion should be prioritized and to what extend? The technical basis for these technologies has in many cases been laid, but the question of how to behave in relation to these new technologies is yet to be answered most of the time

A lack of research in these areas can not only be attributed to these technologies being new. Answering these questions and conducting extensive research in this field can be expensive and time-consuming.

Data-Value-Chain mit einbauen.

While some areas of research can be easily conducted and are only limited by funding or the number of researchers, the area of data analytics

1.2 Research Problem and Objectives

1.3 Methodology and Scope of the Study

Behavioral research can be conducted using a variety of quantitative and qualitative means. The experiment is probably the most time-consuming type of qualitative research, since it requires a great deal of time for the researcher to conduct and perform and a huge time commitment on the side of the participant. One of the major challenges in conducting expertiments in the area of data analytics is the high cost of developing custom applications. The development of such applications can be time-consuming, expensive, and often requires specialized expertise. Although there are already different applications for conducting studies of various kinds, such as survey tools, there is still a lack of applications that have been specifically developed to conduct behavioral research in the field of data analytics. Particu-

larly there is a lack of applications that enable researchers to create studies on a wide range of use cases, without the need to develop tailor-made applications for each specific studies. Moreover, follow-up research is made more difficult to conduct, as these custom applications are usually not publicly available and in most cases unflexible to customize for further studies. This thesis objective is to support the process of behavioral research in the field of data analytics. Due to the aforementioned reasons, an artefact is developed which can be used to conduct experiments in the field of data analytics with the focus of behavioral research. This artefact is supposed to streamline and cheapen research. In order to accomplishe this, the Design Science Research (DSC) Methodology is used which contains the six steps of *Identification of the Problem*, *Definition of Objectives for a solution*, *Design and Dev of artefacts*, *Demonstration of the Artifact*, *Evaluation of the solution*, *Communication*. The identification of the problem is conducted through a literature search with the goal of identifying the degree of research done in the area of data analyitics. To be more specific the current state of research in the field of boundaries and conflicts is investigates. In order to accomplish this, the identified literature is categorized using the so called information-value-chain. This is done in order to show the current state of research on the processing of data and to identify possible literature gaps. After identifying the problem, the next step of the DSC Methodology *Definition of Objectives for a solution* utilizes a second literature search to collect an detailed overview about what kind of studies have been conducted in the field of data analytics, with a special focus on, whether these studies have used custom applications to conduct experiments. These findings are then used to develop requirements using the requirement engineering framework, which is in of itself made up of four different steps *Elicit requirements*, *Requirements specification*, *Verification and validation* and *Requirements management*. These requirements are then used in the *Design and Dev of artefacts steps* of the DSC framework to develope a technical artefact as a solution for the afformentioned problem. Subsequently, the resulting artefact is demonstrated and Evaluation using the afformentioned requirements, in the *Demonstration of the Artifact* and *Evaluation of the solution* steps. The last step of the DSC framework, which focusses on communicating the results to its stakeholders, is ensured by this thesis itself. The last step is thus already fulfilled by this theis itself and will not be considered further.

In summary the objective of this thesis is to develop a solution in the form of an artefact

which enables researchers to design, conduct, and analyze studies more efficiently and cost-effectively, allowing them to explore the field of behavioral research in data analyitics in greater depth.

2 Theoretical foundations

This section represents the theoretical fundamentals of this elaboration by defining the terms “data analytic”, “information value chain”, and “boundaries and conflicts” as they are used in the context of this literature review.

2.1 Definition of terms

2.2 Data Analytics

The term “data analytics” originated in the early 2000s and describes an interdisciplinary field that combines areas such as statistics, machine learning, pattern recognition, system theory, operations research and artificial intelligence (Runkler, 2020). It can be generally defined “[...] as the application of computer systems to the analysis of large data sets for the support of decisions.” (Runkler, 2020). This definition showcases the broadness of the topic, as most computer systems process some amount of data and theoretically allow for some kind of decision making. Due to this broad definition, data analytics can cover slightly different subject areas depending on the context it is discussed in. In this elaboration, data analytics refers to the processing of large amounts of data, also referred to as “big data”, through mathematical procedures or machine learning methods with the goal of creating new knowledge. In summary, processes that merely prepare or show data are not considered data analytics, but only processes that process data in such a way that new knowledge can be derived from it. This distinction is made to differentiate data analytics from traditional data processing areas like business intelligence. The goal of data analytics, as is discussed in this literature review, is to retrieve some kind of previously unknown knowledge from a set of data. This process can be generally described using the “information value chain” model. In their research, Abbasi et al. analyze this model in the context of big data in an effort to create an inclusive research agenda for big data in information system research (Abbasi et al., 2016).

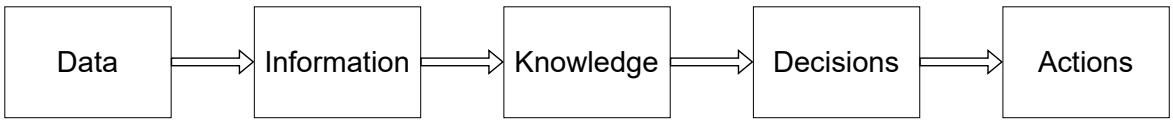

Figure 1: Information Value Chain

2.3 Information Value Chain

The information value chain (figure 1) is a set of phases that define the transformation of raw data to information and eventually into knowledge. "Data" describes raw facts without any structuring. Once organized, the processed data represents "information". This "information" is then used to find patterns and draw conclusions. At this time, the information becomes knowledge (Fayyad et al., 1996a), Fayyad et al., 1996b. This knowledge is then used to make "decisions" and take corresponding "actions" (Sharma et al., 2014). Each phase of the information value chain also includes a different set of technologies and methodologies. For example, the "data" phase contains technologies and actions regarding the basic storage of data like database systems or data warehouses (Abbasi et al., 2016). The conventional version of this information value chain represents an approach that generally explains the processing of data. The main steps of this information value chain are also applicable for big data (Abbasi et al., 2016). This general structure of processing data is also supported by literature from the data analytics field (Runkler, 2020). In addition, the information value chain contains the further phases "decisions" and "actions", which deal with the influence of the processed data. These phases reflect the impact of data analytics, since data analytics is primarily a technology for the decision-making process (Runkler, 2020). For this reason, the information value chain is a suitable model to structure different phases in the processing of data in the context of data analytics.

2.4 Boundaries and Conflicts in Organizations

This literature review uses the terms boundary and conflict interchangeably. In order to include as much literature as possible, the criteria for boundaries are kept very general. Prior to conducting the literature review, there was no formal definition of boundaries in the context of data analytics used for the selection of literature. Generally, boundaries are described as

"[...] a real or imagined line that marks the limits or edges of something and separates it from other things or places [...]" (Hornby, 2015). Based on this general description, the term boundary is defined in the context of this elaboration as any circumstance that leads to a reduction in the effectiveness or efficiency of an organization. Boundaries and conflicts are therefore used to describe any circumstance that hinders an organization from being perfectly productive. An example of such boundaries or conflicts would be communication issues between different departments, which lead to a reduction of productivity.

2.5 Design Science Research Methodology

2.6 Requirement Engineering

Die Methodik, um die Anforderungen der Endanwender in dieser Arbeit zu bestimmen, orientiert sich an dem sogenannten "Requirement Engineering" Institute of Electrical and Electronics Engineers (IEEE) Standard für die Analyse und Evaluation von Anforderungsspezifikationen. (Vgl. Alain Abran, James W. Moore, 2004, S.2) Der Begriff des "Requirement Engineering", frei übersetzt mit "Anforderungsentwicklung", ist dabei ein englischer Begriff aus der Systemanalyse und wird zur Analyse und Evaluierung von Endanwenderanforderungen genutzt (Vgl. Sommerville, 2011, S.82-111) Der IEEE Standard 610.12-1990 definiert eine Anforderung (englisch "Requirement") dabei folgendermaßen:

> ***"(1) Eine Bedingung oder Fähigkeit, die ein Benutzer benötigt, um ein Problem zu lösen oder ein Ziel zu erreichen. (2) Eine Bedingung oder Fähigkeit, die ein System oder eine Systemkomponente erfüllen oder besitzen muss, um einen Vertrag, eine Norm, eine Spezifikation oder andere formell auferlegte Dokumente zu erfüllen. (3) Eine Dokumentendarstellung einer Bedingung oder Fähigkeit wie in 1 oder 2."*** (IEEE, 1990, S.62)

Die hier verwendete Vorgehensweise orientiert sich allerdings nur in Teilen an diesem Standard, da das "Requirement Engineering" grundsätzlich ein Teil des Systementwicklungsprozesses ist und sich diese Arbeit nur mit dem Analysieren und Evaluieren von Standardinhaltsanforderungen in einem existierenden System befasst. Deshalb orientiert sich die Definition

im Kontext dieser Arbeit an Punkt (1) des IEEE Standard 610.12-1990 mit dem Ziel, dem Benutzer eine verbesserte Usability zu gewährleisten. Dabei kann das "Requirement Engineering" in die vier Teilbereiche der Anforderungserhebung, Anforderungsanalyse, Anforderungsspezifikation und Anforderungsbewertung aufgeteilt werden. Bei der Anforderungserhebung werden zunächst die Anforderungen an ein neues Softwaresystem definiert, diese dann analysiert und ausgearbeitet, um sie dann zu dokumentieren und zu evaluieren. (Vgl. Sommerville, 2011, S.82-111) Die Anforderungserhebung der Endanwender an das System wird in dieser Arbeit mithilfe einer Umfrage durchgeführt. Diese Umfrage wird anonymisiert mit Teilnehmern des SAP Enable Now Round Tables stattfinden. In der Anforderungsanalyse wird die momentane Situation der SAP Marketing Cloud betrachtet und mithilfe der Umfrageergebnisse verglichen, um die Anforderungen der Endanwender abzuleiten. Diese Anforderungen werden dann in einem Proof-of-Concept beispielhaft umgesetzt und dazu verwendet, die Anforderungsbewertung durchzuführen. In der Anforderungsbewertung wird dabei der International Organization for Standardization (ISO) 9241 Standard(ISO Org., 2019) genutzt, um die Verbesserung der Usability zu verdeutlichen. Das Ziel der Umfrage und dieser Arbeit ist es also herauszufinden, wie der SAP Enable Now Webassistent von den Endanwendern genutzt wird und wie der dafür bereitgestellte Standardinhalt in der SAP Marketing Cloud verbessert werden kann.

3 Identification of the Problem

3.1 Previous Studies and Gaps in the Literature

In order to identify constrains on the research on data analytics a literature search is conducted. The main objective of it is to analyze the existing literature to find research gaps, particularities and interrelationships between literature. This is supposed to give insights into the current state of research and to find out which part of the research process on data analyitcs still has room for improvements.

The literature search itself is conducted in the field of boundaries and conflicts that might hinder the usage of data analytics.Consequently, relevant literature was identified and reviewed. Afterwards, the identified literature was categorized and analyzed.Initially, it was assumed that the topic of boundaries and conflicts in data analytics lies both in the field of information systems and business (Abbasi et al., 2016, Levina and Vaast, 2005). For this reason, the literature search was mainly conducted in literature databases that focussed on these topics. Table 1 shows the databases that were used.

Online database	Subject Focus
ABI/INFORM Collection	Business and management
Business Source Premier	Accounting, business, economics, management
EconBiz	Business and economics
ProQuest One Business	Business
AIS Electronic Library	Informatics
MIS Quarterly Website	Business informatics
Web of Science	Multiple databases that provide access to different academic topics
Google Scholar	Web search engine for scholarly literature across an array of disciplines

Table 1: Databases Used in the Literature Search

The literature search was conducted using a keyword search. The used keywords consist of phrases like “Data Analytics”, “Data” and “Boundary”. A full list of keywords that were used is included in the appendix.

In order to ensure the quality of the identified literature initially, only publications from certain journals were considered. These journals consist of the *Senior Scholars' Basket*

of Journals and the *UT Dallas Top 100 Business School Research Rankings*. The former includes journals in the area of information systems and the later includes journals in the area of business administration. A full list of included journals is listed in the appendix. Furthermore, only peer-reviewed articles were taken into account. This was done to ensure the quality of the found publications and to additionally exclude book reviews, editorials and opinion statements. Moreover, other 'non-scholarly' texts or publications that did not meet scientific requirements were also not considered in the search. Secondly, the abstracts of the particular articles were inspected to narrow the search further. Consequently, literature that did not meet the topic of boundaries in data analytics was excluded from the search. The literature found in the search was then used for a backward and forward search. During a backward search, all cited sources of an article are examined and during a forward search all the literature that cites the original article is examined (Webster and Watson, 2002). The backward search was conducted using Google Scholar. In addition to this, articles from other journals were, in a second step, reviewed and included as well if they met the scientific requirements, were officially published and relevant to the topic. This process yielded 35 research publications. The results were then assigned to different phases of the aforementioned information value chain, their content best represents. This was done to find literature gaps in the general process of data processing. Additionally, the identified literature was categorized by their research methodology and by their respective industry and involved departments in order to find patterns and similarities in the literature. For example, the overaverage occurrence of boundaries in certain industries could indicate that certain businesses are more susceptible to the emergence of boundaries. In a second in-depth analysis, the different boundaries and possible solutions that were proposed by these articles were categorized and summarized in order to draw overarching conclusions.

The findings of the literature search were structured in multiple ways in order to draw further conclusions than the individual publications allow for. Therefore, the literature search was structured according to the information value chain, the methodology used and the industry covered. Subsequently, the commonalities in content of the literature and the conclusion of the discussed articles are presented.

As stated before, the information value chain consists of the phases "data", "informa-

tion", "knowledge", "decisions" and "actions". The found literature was assigned to these phases, in order to structure and analyze the findings. By mapping the literature found, parts of the data processing process that are over- or under-represented may become visible. From this, conclusions can be drawn about the current state of research. Furthermore, the categories "overspanning" and "other" were introduced to represent literature that either fits multiple phases of the information value chain or none. Using this method leads to the results shown in the "First Search" column of table 2.

Information Value Chain	First Search	Additional Search
Data	4	
Information	3	
Knowledge	21	
Decisions	4	0
Actions	0	0
Overspanning	0	3
Other	3	
Total	35	3

Table 2: Results Assigned to the Information Value Chain

Table 2 shows an overabundance of literature that got assigned to the "knowledge" phase of the information value chain. Among other things, this is due to the fact that the content of this literature deals with the construction and exchange of knowledge within certain groups. The context of this literature is mostly not directly written within the context of data analytics, but nonetheless deals with boundaries in a relevant context.

The significantly fewer entries for the other phases could be explained due to these phases being researched less. However, it cannot be concluded that this underrepresentation is due to the fact that these phases are less susceptible to boundaries. For this, more literature would have to exist confirming that these areas are less prone to boundaries. The underrepresentation of the phases "data" and "information" could also be explained by the fact that these phases are more technology driven and therefore less researched in the context of boundaries. In fact, the corresponding literature, which was assigned to these phases mainly consists of publications researching the application of big data. Their main research object does not directly consist of the identification or resolution of boundaries. Nonetheless, in total, seven individual publications could be found that fit into these two phases. In addition, these two

phases (“data” and “information”) are mostly considered together in the further elaboration, since the literature which was assigned to these phases lies thematically very closely together.

Only four publications were assigned to the “decisions” phase and none to the “actions” phase. These results in particular call into question if the topic of boundaries in data analytics has been extensively researched. The reason for this is the fact that data analytics is primarily a decision support method (Runkler, 2020). Therefore, an overabundance of literature delineating the decision-making process of data analytics should likely exist. This is compounded by the fact that no literature could be found that addressed overspanning issues, as no overarching theories could exist for an insufficiently studied topic. In order to ensure that the ratio of the literature found is based on the research state and not on the keyword search being biased in any way, a second literature search was conducted focussed on finding more literature that could be assigned to the “decisions” or “actions” phase. This was only done for these phases as these two are most relevant in the context of data analytics and because, in total, the least literature could be assigned to them (viewing “data” and “information” together). This second keyword search was conducted with the goal of finding more literature that could be assigned to the phases “decisions” and “actions”. Therefore, a new set of keywords including “decision”, “decision making” and “action” were added to the existing set of keywords. The full list of keywords is included in the appendix. Furthermore, the abstracts were examined with an emphasis on the aforementioned goal. The results of this second keyword search are represented in the “Additional Search” column of table 2. A total number of three additional publications were identified using this second search. These three publications were all assigned to the “overspanning” category. Consequently, no additional literature that could be assigned to the phases “decisions” or “actions” could be identified. This further indicates the fact that the topic of boundaries in data analytics is not researched extensively.

A total number of 38 publications were identified in these two searches and analyzed further.

In order to further analyze the literature and to potentially draw further conclusions, the found literature was also categorized regarding the research method that was used. This categorization is presented in table 3.

Research approach	Method	Number
Qualitative (22)	Case Study	13
	Interviews	4
	Experiments	2
	Observation	3
Quantitative (16)	Survey	12
	Data Analysis	6

Table 3: Research Approach Used in the Literature

The distribution presented in table 3 does not show any significant results. Methods such as case studys and surveys are used more often, this however might be due to the fact that these research methods are easier to implement or more common. Experiments for example might be harder to justify and less effective than surveys, in the context of data analytics.

As already indicated in section **??**, much of the literature found could be assigned to the "knowledge" phase of the information value chain. This literature mostly consists of publications which research boundaries in certain professional groups, but not within the context of data analytics. In addition, literature can be identified which was assigned to the information value chain phases "data" and "information". This literature mostly consists of technical publications, whose main goal is to research areas of application and advantages of data analytics. These publications, are for the most part, not concerned with potential boundaries that could arise within data analytics. These two research areas, which do not quite fit the context of boundaries in data analytics, reinforce the assumption that this field is not researched extensively. Other indications that have already been mentioned are the lack of literature that overspans the topic as a whole and the lack of literature that can be assigned to the "decisions" and "actions" phases of the information value chain.

3.2 Applications (Anwendungen) for Behavioral Research

4 Definition of Objectives for a solution

4.1 Literature Review Studies in Data Analytics and General

4.2 requirements elicitation

4.2.1 Functional and non-functional requirements

Damit eine Anwendungsmodernisierung durchgeführt und eine modernisierte Architektur konzeptioniert werden kann, ist die Bestimmung der Anforderungen wichtig.(Vgl. Seacord et al., 2003, Kapitel 3) Anforderungen können nach ISO/IEC 25000, beziehungsweise dem Qualitätsmodell aus ISO/IEC 25010, als Qualitätskriterien an Software und Systeme klassifiziert werden.(Vgl. ISO/IEC 25010, 2011) Das IEEE definiert Anforderungen als eine Bedingung oder Eigenschaft, welche von einem System oder einer Systemkomponente erfüllt werden muss, um eine Problemstellung oder Zielsetzung eines Nutzers oder formalen Dokuments zu erfüllen.(Vgl. IEEE, 1990, S.62) Anforderungen können nach diesen beiden Definitionen als zu erfüllende Eigenschaften oder Qualitätskriterien einer Software oder eines Systems definiert werden. Aus diesem Grund werden die Anforderungen an den SAP Lean Catalog aus den von der SAP beschriebenen Produkteigenschaften und Merkmalen des SAP Lean Catalogs abgeleitet. Zusätzlich werden allgemeine Anforderungen an eine Softwaremodernisierung berücksichtigt, welche erfüllt sein müssen, um eine erfolgreiche Migration durchzuführen. Diese Anforderungen an den SAP Lean Catalog und an eine Softwaremigration im Allgemeinen werden weiter in funktionale und nichtfunktionale Anforderungen eingeteilt.

Eine funktionale Anforderung beschreibt eine Funktion oder Fähigkeit eines Systemes, die konkret von einem System oder einer Softwarekomponente durchgeführt werden können muss.(Vgl. IEEE, 1990, S.35) Ein Beispiel für eine funktionale Anforderung wäre die Berechnung des Bestellpreises in Euro und in Dollar. Nichtfunktionale Eigenschaften beschreiben hingegen Verhaltensweisen des Systems(Vgl. Seacord et al., 2003, Kapitel 3) und gehen damit über die funktionalen Eigenschaften hinaus. Damit beschreiben funktionale Anforderungen was ein System können muss und nichtfunktionale Anforderungen wie es funktionieren soll. Nichtfunktionale Anforderungen beschreiben außerdem häufig die Qual-

ität der Funktionen und können mehrere andere Anforderungen beeinflussen.(Vgl. Balzert, 2011, S.109ff) Ein Beispiel für nichtfunktionale Eigenschaften wäre, dass die Umrechnung von Euro in Dollar in “wenigen Sekunden” durchgeführt werden muss.

4.3 Requirements analysis

5 Design and Dev artefacts

5.1 System Architecture and Components

5.2 User Interface Design and Implementation

5.3 Prototype Development

6 Demonstration of the Artifact

7 Evaluation of the solution

7.1 Prototype Testing

7.2 Requirements validation

7.3 (App Performance and Usability / User Feedback and Satisfaction)

8 Conclusion

8.1 Summary of the Study

8.2 Contributions and Implications

8.3 Future Work and Recommendations

Bibliography

Abbasi, A., Sarker, S., & Chiang, R. H. L. (2016). Big data research in information systems: Toward an inclusive research agenda. *Journal of the Association for Information Systems*, *17*(2).

Abramson, C., Currim, I. S., & Sarin, R. (2005). An experimental investigation of the impact of information on competitive decision making. *Management Science*, *51*(2), 195–207.

Alain Abran, James W. Moore. (2004). *Swebok: guide to the software engineering body of knowledge*. IEEE Computer Society.

Allen, T. J., & Cohen, S. I. (1969). Information flow in research and development laboratories. *Administrative Science Quarterly*, *14*(1).

Amankwah-Amoah, J., & Adomako, S. (2019). Big data analytics and business failures in data-rich environments: An organizing framework. *Computers in Industry*, *105*.

Bag, S., Gupta, S., & Wood, L. (2022). Big data analytics in sustainable humanitarian supply chain: Barriers and their interactions. *Annals of Operations Research*, *319*(1), 721–760.

Balzert, H. (Ed.). (2011). *Lehrbuch der softwaretechnik: Entwurf, implementierung, installation und betrieb*. Spektrum Akademischer Verlag. https://doi.org/10.1007/978-3-8274-2246-0

Barrett, M., & Oborn, E. (2010). Boundary object use in cross-cultural software development teams. *Human Relations*, *63*(8), 1199–1221. https://doi.org/10.1177/0018726709355657

Bhatti, S. H., Hussain, W. M. H. W., Khan, J., Sultan, S., & Ferraris, A. (2022). Exploring data-driven innovation: What's missing in the relationship between big data analytics capabilities and supply chain innovation?. *Annals of Operations Research*, 1–26.

BIS Research. (2018). Umsatz mit big data im bereich healthcare weltweit nach anwendung in den jahren 2016 und 2025 (in milliarden us-dollar). https://de.statista.com/statistik/daten/studie/997352/umfrage/umsatz-mit-big-data-im-bereich-healthcare-nach-anwendung/

Bitkom. (2017). Welches sind die wichtigsten it-trends des jahres 2017? https://de.statista.com/statistik/daten/studie/675726/umfrage/die-wichtigsten-trends-in-der-itk-branche/

Bitkom. (2018). Umsatz mit big-data-lösungen in deutschland in den jahren 2016 und 2017 und prognose für 2018 (in milliarden euro). https://de.statista.com/statistik/daten/studie/257976/umfrage/umsatz-mit-big-data-loesungen-in-deutschland/

Carlile, P. R. (2002). A pragmatic view of knowledge and boundaries: Boundary objects in new product development. *Organization Science*, *13*(4), 442–455. https://doi.org/10.1287/orsc.13.4.442.2953

Chakraborty, I., Hu, P. J.-H., & Cui, D. (2008). Examining the effects of cognitive style in individuals' technology use decision making. *Decision Support Systems*, *45*(2), 228–241.

Chen, D. Q., Preston, S. D., & Swink, M. (2021). How big data analytics affects supply chain decision-making: An empirical analysis. *Journal of the Association for Information Systems*, *22*(5). https://doi.org/10.5194/gi-2016-11-RC2

Chen, L., Liu, H., Zhou, Z., Chen, M., & Chen, Y. (2022). It-business alignment, big data analytics capability, and strategic decision-making: Moderating roles of event criticality and disruption of covid-19. *Decision Support Systems*, *161*, N.PAG.

Chou, S.-W., & Chang, Y.-C. (2008). The implementation factors that influence the erp (enterprise resource planning) benefits. *Decision Support Systems*, *46*(1), 149–157.

Cross, R. L., & Parker, A. (2004). *The hidden power of social networks: Understanding how work really gets done in organizations*. Harvard Business School Press.

Currie, G., & Kerrin, M. (2004). The limits of a technological fix to knowledge management. *Management Learning*, *35*(1), 9–29. https://doi.org/10.1177/1350507604042281

Czekster, R. M., De Carvalho, H. J., Kessler, G. Z., Kipper, L. M., & Webber, T. (2019). Decisor: A software tool to drive complex decisions with analytic hierarchy process. *International Journal of Information Technology & Decision Making*, *18*(1), 65–86.

Demoulin, N. T., & Coussement, K. (2020). Acceptance of text-mining systems: The signaling role of information quality. *Information & Management*, *57*(1), N.PAG.

Donghyuk, S., Shu, H., Gene Moo, L., Whinston, A. B., Cetintas, S., & Kuang-Chih, L. (2020). Enhancing social media analysis with visual data analytics: A deep learning approach. *MIS Quarterly*, *44*(4), 1459–1492.

Du, Wnyu, Pan, S. L., Xie, K., & Xiao, J. (2020). Data analytics contributes to better decision-making beyond organizational boundaries. *MIS Quarterly Executive*, *19*(2).

Elgendy, N., & Elragal, A. (2016). Big data analytics in support of the decision making process. *Procedia Computer Science*, *100*, 1071–1084. https://doi.org/10.1016/j.procs.2016.09.251

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996a). From data mining to knowledge discovery in databases. *AI Magazine*, *17*(3).

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996b). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, *39*(11), 27–34. https://doi.org/10.1145/240455.240464

Fink, L., Yogev, N., & Even, A. (2017). Business intelligence and organizational learning: An empirical investigation of value creation processes. *Information & Management*, *54*(1), 38–56.

Foerderer, J., Kude, T., Schuetz, S. W., & Heinzl, A. (2019). Knowledge boundaries in enterprise software platform development: Antecedents and consequences for platform governance. *Information Systems Journal*, *29*(1), 119–144. https://doi.org/10.1111/isj.12186

Fosso Wamba, S., Queiroz, M. M., Wu, L., & Sivarajah, U. (2020). Big data analytics-enabled sensing capability and organizational outcomes: Assessing the mediating effects of business analytics culture. *Annals of Operations Research*, 1–20.

Ghasemaghaei, M. (2019). Does data analytics use improve firm decision making quality? the role of knowledge sharing and data analytics competency. *Decision Support Systems*, *120*, 14–24.

Ghasemaghaei, M., Ebrahimi, S., & Hassanein, K. (2017). Data analytics competency for improving firm decision making performance. *The Journal of Strategic Information Systems*, *27*(1), 101–113. https://doi.org/10.1016/j.jsis.2017.10.001

Ghasemaghaei, M., Hassanein, K., & Turel, O. (2017). Increasing firm agility through the use of data analytics: The role of fit. *Decision Support Systems*, *101*, 95–105.

Goodhue, D. L., Kirsch, L. J., Quillard, J. A., & Wybo, M. D. (1992). Strategic data planning: Lessons from the field. *16*(1).

Guven–Uslu, P., Blaber, Z., & Adhikari, P. (2020). Boundary spanners and calculative practices. *Financial Accountability & Management*, *36*(4), 439–460. https://doi.org/10.1111/faam.12266

Han, S., Datta, A., & Joshi, K. D., Chi, Lei. (2017). Innovation through boundary spanning: The role of it in enabling knowledge flows across technological and geographical boundaries. *International Journal of Knowledge Management*, *13*(4), 90–110. https://doi.org/10.4018/IJKM.2017100105

Hornby, A. S. (2015). *Oxford advanced learner's dictonary of current english* (9th). Cornelsen.

IEEE. (1990). Ieee standard glossary of software engineering terminology - ieee std 610.12-1990.

Inman, J. J., Winer, R. S., & Ferraro, R. (2009). The interplay among category characteristics, customer characteristics, and customer activities on in-store decision making. *Journal of Marketing*, *73*(5), 19–29.

Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of bi capabilities and decision environments. *Information & Management*, *50*(1), 13–23.

ISO Org. (2019). *Norm iso 9241*. Retrieved September 30, 2019, from https://www.iso.org/standard/77520.html

ISO/IEC 25010 (Ed.). (2011). *Systems and software engineering - systems and software quality requirements and evaluation (square)*. ISO/IEC. Retrieved March 23, 2021, from https://www.iso.org/standard/35733.html

Jha, A. K., Agi, M. A., & Ngai, E. W. (2020). A note on big data analytics capability development in supply chain. *Decision Support Systems*, *138*, N.PAG.

Ji, G., Yu, M., Tan, K. H., Kumar, A., & Gupta, S. (2022). Decision optimization in cooperation innovation: The impact of big data analytics capability and cooperative modes. *Annals of Operations Research*, 1–24.

Johnson, J. P., Lenartowicz, T., & Apud, S. (2006). Cross-cultural competence in international business: Toward a definition and a model. *Journal of International Business Studies*, *37*(4), 525–543.

Kankanhalli, A., Ye, H. (, & Teo, H. H. (2015). Comparing potential and actual innovators: An empirical study of mobile data services innovation. *MIS Quarterly*, *39*(3), 667–682. Retrieved July 6, 2023, from https://www.jstor.org/stable/26629625

Karhade, P., & Dong, J. Q. (2021). Innovation outcomes of digitally enabled collaborative problemistic search capability. *MIS Quarterly*, *45*(2), 693–718. https://doi.org/10.25300/MISQ/2021/12202

Kim, B. J., & Tomprou, M. (2021). The effect of healthcare data analytics training on knowledge management: A quasi-experimental field study. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(1), 60. https://doi.org/10.3390/joitmc7010060

Klingebiel, R., & Meyer, A. D. (2013). Becoming aware of the unknown: Decision making during the implementation of a strategic initiative. *Organization Science*, *24*(1), 133–153. Retrieved July 6, 2023, from http://www.jstor.org/stable/23362104

Korschun, D. (2015). Boundary-spanning employees and relationships with external stakeholders: A social identity approach. *Academy of Management Review*, *40*(4), 611–629. https://doi.org/10.5465/amr.2012.0398

Kotlarsky, J., Scarbrough, H., & Oshri, I. (2014). Coordinating expertise across knowledge boundaries in offshore-outsourcing projects: The role of codification. *MIS Quarterly*, *38*(2).

Krakowski, S., Luger, J., & Raisch, S. (2023). Artificial intelligence and the changing sources of competitive advantage. *Strategic Management Journal (John Wiley & Sons, Inc.)*, *44*(6), 1425–1452.

Lancelot Miltgen, C., Popovič, A., & Oliveira, T. (2013). Determinants of end-user acceptance of biometrics: Integrating the "big 3" of technology acceptance with privacy context. *Decision Support Systems*, *56*, 103–114.

Lebovitz, S., Lifshitz-Assaf, H., & Levina, N. (2022). To engage or not to engage with ai for critical judgments: How professionals deal with opacity when using ai for medical diagnosis. *Organization Science*, *33*(1), 126–148.

Lee, G. M., He, S., Lee, J., & Whinston, A. B. (2020). Matching mobile applications for cross-promotion. *Information Systems Research*, *31*(3), 865–891.

Lehmann, D., Fekete, D., & Vossen, G. (2016). *Technology selection for big data and analytical applications* (ERCIS Working Paper No. 27). Münster, Westfälische Wilhelms-Universität Münster, European Research Center for Information Systems (ERCIS). http://hdl.handle.net/10419/156084

Leidner, D. E., & Elam, J. J. (1995). The impact of executive information systems on organizational design, intelligence, and decision making. *Organization Science*, *6*(6), 645–664.

Levina, N., & Vaast, E. (2005). The emergence of boundary spanning competence in practice: Implications for implementation and use of information systems. *MIS Quarterly*, *29*(2).

Levina, N., & Vaast, E. (2006). Turning a community into a market: A practice perspective on information technology use in boundary spanning. *Journal of Management Information Systems*, *22*(4), 13–37. https://doi.org/10.2753/MIS0742-1222220402

Levina, N., & Vaast, E. (2008). Innovating or doing as told? status differences and overlapping boundaries in offshore collaboration. *MIS Quarterly*, *32*(2).

Li, H., Lu, K., & Meng, S. (2015). Bigprovision: A provisioning framework for big data analytics. *IEEE Network*, *29*(5), 50–56.

Li, M. (, Huang, Y., & Sinha, A. (2020). Data-driven promotion planning for paid mobile applications. *Information Systems Research*, *31*(3), 1007–1029.

Liberatore, M. J., & Stylianou, A. C. (1995). Expert support systems for new product development decision making: A modeling framework and applications. *Management Science*, *41*(8), 1296–1316. Retrieved July 6, 2023, from http://www.jstor.org/stable/2632787

Lindgren, R., Andersson, M., & Henfridsson, O. (2008). Multi-contextuality in boundary-spanning practices. *Information Systems Journal*, *18*(6), 641–661. https://doi.org/10.1111/j.1365-2575.2007.00245.x

Lucas, H. C. (1981). An experimental investigation of the use of computer-based graphics in decision making. *Management Science*, *27*(7), 757–768. Retrieved July 6, 2023, from http://www.jstor.org/stable/2630917

Lukyanenko, R., Parsons, J., Wiersma, Y. F., & Maddah, M. (2019). Expecting the unexpected: Effects of data collection design choices on the quality of crowdsourced user-generated content. *MIS Quarterly*, *43*(2), 623–647. https://doi.org/10.25300/MISQ/2019/14439

Lurie, N. H., & Mason, C. H. (2007). Visual representation: Implications for decision making. *Journal of Marketing*, *71*(1), 160–177.

Majchrzak, A., More, P. H. B., & Faraj, S. (2012). Transcending knowledge differences in cross-functional teams. *Organization Science*, *23*(4), 951–970. https://doi.org/10.1287/orsc.1110.0677

Mäkelä, K., Barner-Rasmussen, W., Ehrnrooth, M., & Koveshnikov, A. (2019). Potential and recognized boundary spanners in multinational corporations. *Journal of World Business*, *54*(4), 335–349. https://doi.org/10.1016/j.jwb.2019.05.001

Marchena Sekli, G. F., & de La Vega, I. (2021). Adoption of big data analytics and its impact on organizational performance in higher education mediated by knowledge management. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(4), 221. https://doi.org/10.3390/joitmc7040221

Mell, J. N., Knippenberg, D., Ginkel, W. P., & Heugens, P. P. M. A. R. (2022). From boundary spanning to intergroup knowledge integration: The role of boundary spanners' metaknowledge and proactivity. *Journal of Management Studies*, *Not yet published*. https://doi.org/10.1111/joms.12797

Minbaeva, D., & Santangelo, G. D. (2018). Boundary spanners and intra-mnc knowledge sharing: The roles of controlled motivation and immediate organizational context. *Global Strategy Journal*, *8*(2), 220–241. https://doi.org/10.1002/gsj.1171

Montgomery, A. L., Li, S., Srinivasan, K., & Liechty, J. C. (2004). Modeling online browsing and path analysis using clickstream data [Copyright - Copyright Institute for Operations Research and the Management Sciences Fall 2004; Document feature - references; tables; graphs; equations; Last updated - 2023-07-03; CODEN - MARSE5]. *Marketing Science*, *23*(4), 579–595. https://www.proquest.com/scholarly-journals/modeling-online-browsing-path-analysis-using/docview/212288093/se-2

Mueller, O., Fay, M., & vom Brocke, J. (2018). The effect of big data and analytics on firm performance: An econometric analysis considering industry characteristics. *Journal of Management Information Systems*, *35*(2), 488–509. https://doi.org/10.1080/07421222.2018.1451955

Nutt, P. C. (1998). How decision makers evaluate alternatives and the influence of complexity. *Management Science*, *44*(8), 1148–1166. Retrieved July 6, 2023, from http://www.jstor.org/stable/2634692

Pawlowski, S. D., & Robey, D. (2004). Bridging user organizations: Knowledge brokering and the work of information technology professionals. *28*(4).

Peng, Y., & Sutanto, J. (2012). Facilitating knowledge sharing through a boundary spanner. *IEEE Transactions on Professional Communication*, *55*(2), 142–155. https://doi.org/10.1109/TPC.2012.2188590

Pil Han, S., Park, S., & Oh, W. (2016). Mobile app analytics: A multiple discrete-continuous choice framework. *MIS Quarterly*, *40*(4), 983–A42.

Po-An Hsieh, J. J., Rai, A., & Xin Xu, S. (2011). Extracting business value from it: A sensemaking perspective of post-adoptive use. *Management Science*, *57*(11), 2018–2039.

Popovič, A., Hackney, R., Coelho, P. S., & Jaklič, J. (2012). Towards business intelligence systems success: Effects of maturity and culture on analytical decision making. *Decision Support Systems*, *54*(1), 729–739.

Pour, M. J., Abbasi, F., & Sohrabi, B. (2023). Toward a maturity model for big data analytics: A roadmap for complex data processing. *International Journal of Information Technology & Decision Making*, *22*(1), 377–419.

Qiqi, J., Chuan-Hoo, T., Choon Ling, S., & Kwok-Kee, W. (2019). Followership in an open-source software project and its significance in code reuse. *MIS Quarterly*, *43*(4), 1303–1319.

Redman, T. C. (2008). *Data driven: Profiting from your most important business asset*. Harvard Business Review Press.

Richter, A. W., West, M. A., van Dick, R., & Dawson, J. F. (2006). Boundary spanners' identification, intergroup contact, and effective intergroup relations. *Academy of Management Journal*, *49*(6), 1252–1269. https://doi.org/10.5465/amj.2006.23478720

Runkler, T. A. (2020). *Data analytics*. Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-29779-4

Russell, S., Gangopadhyay, A., & Yoon, V. (2008). Assisting decision making in the event-driven enterprise using wavelets. *Decision Support Systems*, *46*(1), 14–28.

Ryder, B., Gahr, B., Egolf, P., Dahlinger, A., & Wortmann, F. (2017). Preventing traffic accidents with in-vehicle decision support systems - the impact of accident hotspot warnings on driver behaviour. *Decision Support Systems*, *99*, 64–74.

Seacord, R. C., Plakosh, D., & Lewis, G. A. (2003). *Modernizing legacy systems: Software technologies, engineering processes, and business practices*. Addison-Wesley. Retrieved February 21, 2021, from https://learning.oreilly.com/library/view/modernizing-legacy-systems/0321118847/

Seagate. (2018). Prognose zum volumen der jährlich generierten digitalen datenmenge weltweit in den jahren 2018 und 2025 (in zettabyte). https://de.statista.com/statistik/daten/studie/267974/umfrage/prognose-zum-weltweit-generierten-datenvolumen/

Sharda, R., Barr, S. H., & McDonnell, J. C. (1988). Decision support system effectiveness: A review and an empirical test. *Management Science*, *34*(2), 139–159. Retrieved July 6, 2023, from http://www.jstor.org/stable/2632057

Sharma, R., Mithas, S., & Kankanhalli, A. (2014). Transforming decision-making processes: A research agenda for understanding the impact of business analytics on organisations. *European Journal of Information Systems*, *23*(4), 433–441. https://doi.org/10.1057/ejis.2014.17

Sommerville, I. (2011). *Software engineering* (9th ed.). Pearson.

Song, P., Zheng, C., Zhang, C., & Yu, X. (2018). Data analytics and firm performance: An empirical study in an online b2c platform. *Information & Management*, *55*(5), 633–642.

Spiller, S. A., Reinholtz, N., & Maglio, S. J. (2020). Judgments based on stocks and flows: Different presentations of the same data can lead to opposing inferences. *Management Science*, *66*(5), 2213–2231.

Tang, H., Liao, S. S., & Sun, S. X. (2013). A prediction framework based on contextual data to support mobile personalized marketing. *Decision Support Systems*, *56*, 234–246.

Trieu, V.-H., Burton-Jones, A., Green, P., & Cockcroft, S. (2022). Applying and extending the theory of effective use in a business intelligence context. *46*(1).

van Osch, W., & Steinfield, C. W. (2016). Team boundary spanning: Strategic implications for the implementation and use of enterprise social media. *Journal of Information Technology*, *31*(2), 207–225. https://doi.org/10.1057/jit.2016.12

Webster, J., & Watson, R. T. (2002). Analyzing the past to prepare for the future: Writing a literature review. *MIS Quarterly*, *26*(2).

Wixom, B. H., & Watson, H. J. (2001). An empirical investigation of the factors affecting data warehousing success. *MIS Quarterly*, *25*.

Wook, M., Hasbullah, N. A., Zainudin, N. M., Jabar, Z. Z. A., Ramli, S., Razali, N. A. M., & Yusop, N. M. M. (2021). Exploring big data traits and data quality dimensions for big data analytics application using partial least squares structural equation modelling. *Journal of Big Data*, *8*(1). https://doi.org/10.1186/s40537-021-00439-5

Wright, R. T., Jensen, M. L., Bennett Thatcher, J., Dinger, M., & Marett, K. (2014). Influence techniques in phishing attacks: An examination of vulnerability and resistance. *Information Systems Research*, *25*(2), 385–400.

ZAVADSKAS, E. K., VAINIŪNAS, P., TURSKIS, Z., & TAMOŠAITIENĖ, J. (2012). Multiple criteria decision support system for assessment of projects managers in construction. *International Journal of Information Technology & Decision Making*, *11*(2), 501–520.

Zhang, Q., & Li, J. (2021). Can employee's boundary-spanning behavior exactly promote innovation performance? the roles of creative ideas generation and team task interdependence. *International Journal of Manpower*, *42*(6), 1047–1063. https://doi.org/10.1108/IJM-06-2019-0302

Zhao, Z. J., & Anand, J. (2013). Beyond boundary spanners: The 'collective bridge' as an efficient interunit structure for transferring collective knowledge. *Strategic Management Journal*, *34*(13), 1513–1530. https://doi.org/10.1002/smj.2080

Literature search: Boundaries and conflicts in data analytics

List of Keywords (First Search)

- Data Analytics
- Data AND Boundary
- Organization AND Data Analytics
- Big Data
- boundary
- boundary theory
- boundary spanning
- boundary objects
- boundary spanner

List of Keywords (Second Search)

- Data Analytics
- Data AND Boundary
- Organization AND Data Analytics
- Big Data
- boundary
- boundary theory
- boundary spanning
- boundary objects
- boundary spanner
- Decision
- Decision Making
- Action

Senior Scholars' Basket of Journals

- European Journal of Information Systems
- Information Systems Journal
- Information Systems Research
- Journal of AIS
- Journal of Information Technology
- Journal of MIS
- Journal of Strategic Information Systems
- MIS Quarterly

UT Dallas Top 100 Business School Research Rankings

- The Accounting Review
- Journal of Accounting and Economics
- Journal of Accounting Research
- Journal of Finance
- Journal of Financial Economics
- The Review of Financial Studies
- Information Systems Research
- Journal on Computing
- MIS Quarterly
- Journal of Consumer Research
- Journal of Marketing
- Journal of Marketing Research
- Marketing Science
- Management Science
- Operations Research
- Journal of Operations Management
- Manufacturing and Service Operations Management
- Production and Operations Management
- Academy of Management Journal
- Academy of Management Review
- Administrative Science Quarterly

- Organization Science
- Journal of International Business Studies
- Strategic Management Journal

Affidavit

I hereby declare that I have developed and written the enclosed master thesis entirely on my own and have not used outside sources without declaration in the text. Any concepts or quotations applicable to these sources are clearly attributed to them. This master thesis has not been submitted in the same or a substantially similar version, not even in part, to any other authority for grading and has not been published elsewhere. This is to certify that the printed version is equivalent to the submitted electronic one. I am aware of the fact that a misstatement may have serious legal consequences.

I also agree that my thesis can be sent and stored anonymously for plagiarism purposes. I know that my thesis may not be corrected if the declaration is not issued.

Mannheim, July 10, 2023

Max Darmstadt