Beyond the Hype: Why Do Data-Driven Projects Fail?

Tatiana Ermakova
Weizenbaum Institute,
Fraunhofer FOKUS,
Technical University of Berlin
tatiana.ermakova@fokus.fraunhofer.de

Julia Blume
HfT Leipzig,
HWR Berlin
Julia.blume@web.de

Benjamin Fabian
TH Wildau,
HfT Leipzig,
Humboldt-University of Berlin
benjamin.fabian@th-wildau.de

Elena Fomenko
Detecon International GmbH
Elena.Fomenko@detecon.com

Marcus Berlin
Detecon International GmbH
Marcus.Berlin@detecon.com

Manfred Hauswirth
Weizenbaum Institute,
Fraunhofer FOKUS,
Technical University of Berlin
manfred.hauswirth@fokus.fraunhofer.de

Abstract

Despite substantial investments, data science has failed to deliver significant business value in many companies. So far, the reasons for this problem have not been explored systematically. This study tries to find possible explanations for this shortcoming and analyses the specific challenges in data-driven projects. To identify the reasons that make data-driven projects fall short of expectations, multiple rounds of qualitative semi-structured interviews with domain experts with different roles in data-driven projects were carried out. This was followed by a questionnaire surveying 112 experts with experience in data projects from eleven industries. Our results show that the main reasons for failure in data-driven projects are (1) the lack of understanding of the business context and user needs, (2) low data quality, and (3) data access problems. It is interesting, that 54% of respondents see a conceptual gap between business strategies and the implementation of analytics solutions. Based on our results, we give recommendations for how to overcome this conceptual distance and carrying out data-driven projects more successfully in the future.

1. Introduction

With advances in computational power and the enormous amounts of data available nowadays, companies in almost every industry are increasingly exploiting data for business advantage [42], with top performers five times more frequently than lower performers, and half of the investigated companies in a study by MIT Sloan Management Review viewing this area as their top priority [30]. Interestingly, nearly six out of ten organizations define themselves through data-driven projects [30]. As examples for this trend, Google refines its core search and ad-serving algorithms by in-depth analysis based on search and access data, while Netflix uses this information to improve its recommendation system, and LinkedIn innovates data-driven products, features, and value-adding services, e.g., the "People You May Know" feature proposing new connections for a user [11].

Despite these success stories, data-driven projects often fail [34, 41, 51]. For instance, in a survey in 2014 with 226 respondents from Europe, North America, and APAC (Asia Pacific) and several industries, participants regarded only 27% of data-driven projects as successful. Only 13% of organizations had accomplished full-scale deployment with their data-driven implementations [7, 48].

Indeed, learning a generalizable model from the data might result in unintentionally selecting a spurious solution [21, 26]. In the cited data-driven model on Google Flu Trends, the flu propensity, as measured by the number of influenza-related doctor visits, could be estimated based on the number of flu-related Google search queries [18]; however, it was overestimated by more than a factor of two in two consecutive years [31]. Furthermore, research has been showing increasingly that other aspects are essential as well, such as people [3, 4, 6, 8, 12, 23, 24, 25, 27, 28, 34, 39, 52], data [1, 7, 8, 24, 27, 28, 29, 32, 34, 49], tools and technologies [8, 27, 28, 34, 39], as well as processes [8, 24, 28, 33, 34, 39, 47].

Partial practical insights on the challenges of data-driven projects were obtained in the expert rounds, including case studies [4, 9, 28, 33, 34, 39, 49, 51] and

HⁱCSS

expert interviews [1, 8, 24, 29, 32]. However, none of those findings were further evaluated with a larger expert sample. These issues motivated us to systematically investigate the research question of (1) what are the most crucial challenges of data-driven projects in practice, as perceived by experts. Based on a systematic survey of the related literature [40], we established the framing for the given problem statement to base our research project on as well as the motivation behind our research question. To design an effective survey for data experts, we followed the applied action design research approach [46]. Based on qualitative semi-structured interviews [38] with 13 experts representing different roles and competence areas such as strategic designers, data scientists, UX-Designers, or consultants, what took approximately eight weeks, we conducted a total of 20 rounds of design, validation and refinement of the survey draft. As a result of this phase, the focus of our research was further deepened to understand (1a) how successful experts perceived their previous data-driven projects and (1b) what roles distinct phases and challenges played in the non-successful data-driven projects. Finally, we evaluated the challenges in an online survey with 112 experts who had a data-related background and/or positions in data-related departments. The interviews and the survey were conducted between August and October 2019. We discuss the outcomes of this study and derive several implications for research and practice.

2. Background

Data science involves the extraction of informative patterns [42, 45] or generalizable knowledge from data [13, 42], e.g., in the form of testable explanations and predictions [22], or its meaningful transformation [10]. The "science" part of the term emphasizes the systematic [13, 22] or principled [42] enterprise/study, e.g., through the KDD routine (knowledge discovery in databases) including stages of data selection, preprocessing, transformation, data mining, interpretation and evaluation [14] or the CRISP-DM (Cross Industry Standard Process for Data Mining) [1, 5]. To compare with data science, data mining is related to the actual extraction of knowledge from data via technologies that incorporate these principles [14, 19, 42, 54].

Along the process of data integration, "data-driven companies" implement data science approaches in the hope to add business value [20, 25]. "Data-centric companies" [15] such as Google, Amazon, or Microsoft go further and place data science at the core of their business [25, 53], whereas "data-informed companies" only use relatively simple data dashboards to support decisions [25]. For instance, Google bases on the search

and access data to refine its core search and ad-serving algorithms [11].

2. Methodology

2.1. Literature review

To shed light on the accumulated advancements of knowledge on the challenges of data-driven projects, we consulted the AIS (Association for Information Systems) research community, following a systematic literature review approach [40]. In our search for related scientific work via the AIS eLibrary, we discovered and extracted 1777 research studies based on the occurrence of the search terms "challenges", "failures" or "problems" jointly with the terms "data-driven" or "machine learning" or "deep learning" or "artificial intelligence" or "project" in their titles, abstracts, and subjects. When narrowing the number of search results, we first scanned the extracted articles' titles for indications of their relevance and conformity to the research scope to enable their meaningful reflection and comparison across studies. In the presence of an indication, we analyzed the abstract to determine whether the considered paper still meets the inclusion criteria. By implication, we excluded a paper from the review if no such indication was found. For a total of resulting 97 articles, we reviewed the content in its entirety. As a result of our manual screening, we identified 18 matching research studies.

Among the first ones, Butler and Sammon [4] speak about the significance of organizational and social factors, e.g., the end-user skills to leverage business-related data, experiential knowledge, and understanding of business-related data. In line with them, Bannerman et al. [3] suggest the importance of the right capabilities for project performance, illustrating them by both a classic case and a rather contemporary case from Australia. Hoxmeier and Lenk [23] mention such capabilities as technical knowledge, interpersonal communication skills, client, and project management skills. Geva and Saar-Tsechansky [16] show that experts can be ranked with respect to their unobserved decision quality, when applying a machine learning approach to the problem definition using one dataset with real expert decisions among others. Pflügler et al. [41] show that failure experiences exert a positive significant influence on project profitability.

Roughly identically to the approach of Wilson [51], the examination by Liu et al. [34] stresses the importance of identification of business needs, team building, identification of talents, skills and certifications needed, the involvement of stakeholders in the process, culture creation, identification of data and best statistical practices. In the fourth iteration of the

Data Value Map by Nagle and Sammon [39], developed over the time period of four years and evaluated by a long-lasting case study with 96 practitioners, the authors report a lacking organizational mental model, the need for a shared language, and over-emphasis on technology as the main barriers in developing a shared understanding for data-driven projects.

Taking the individual aspects of project scope, time, risk, human resources, procurement, and quality management into account enables Wilson [51] to compare two analogous IS projects, outsourced in Wellington, New Zealand, between 1992 and 1999, which were different in terms of failure or success.

Koronios et al. [28] form a list of potentially relevant factors from multiple data-driven projects and group them using ABC analysis into six critical success factors, including information strategy, attributable business value, top management support, project skills, information quality, security and integrity, and technological capability.

Li et al. [33] define key issues for a data-driven realization: From a technological perspective, these include data acquisition, data processing, the discovery of technological opportunities and identification technology. From a management view, these cover data-driven operation and decision-making, matched decision-making culture and appropriate process, overall planning of large data applications, and the target stage.

The exploratory case study by Dahlberg et al. [9] identifies further potential factors, using a plan-driven project as one example, which achieved its project goals despite the undelivered business value, and a change-driven (agile) project, which delivered the desired business value, but the resulting information system was lacking in terms of robustness.

According to the systematic literature review along with interviews with practitioners across various industries conducted by Baier et al. [1], potential challenges for project success could be assigned to three categories – pre-deployment, deployment and non-technical – and six clusters – structure of data, implementation, infrastructure, governance, customer relation, and economic implications.

As a result of the interviews conducted by Jensen et al. [24], challenges are supposed to be addressed through overall business case formulation and prioritization, organizational context appreciation, the definition of overall benefits and benefit measures, understanding of beneficial relationships across departments, measuring benefits and usefulness for end-users, management of missing benefits, the establishment of target end-users.

Based on a structured literature review and a study of 51 use cases, Volk et al. [49] propose a quantitative classification framework for big data-driven projects, covering the characteristics of data volume, variety, velocity, volatility, variability, and consistency. They show the suitability of the framework in the context of two scenarios of one consulting company located in Europe as an example.

The interviews by Cronholm et al. [8] unveil a different set of challenges, such as lack of a systematic process, problems related to data access, distrust of data, lack of appropriate digital tools, coupled with insufficient competence.

Based on the interviews by Lennerholt and van Laere [32], challenges can include access and use of data, i.e., lack of possibilities to gain access to data, multiple data sources in different environments, unknown data sources, long time required to request access, difficulties in making data available, as well as data quality, i.e., faulty data in standard reports, difficulties in changing faulty data, no common definition of data, no awareness of using faulty data.

Maass et al. [35] propose to connect the data-driven and theory-driven perspectives. The Data Insight Generator (DIG) by Kühne and Böhmann [29], a system built on expert interviews to complement the Business Model Canvas to design data-driven business models, connects the segments of key resources, i.e., data, and value propositions through the elements of data quality, and combination of datasets and pipes, analytics, and insights.

To sum up, partial practical insights were gained in different expert rounds, i.e., both case studies [4, 9, 28, 33, 34, 39, 49, 51] and expert interviews [1, 8, 24, 29, 32]. The gained insights relate the challenges of data-driven projects to people [3, 4, 6, 8, 12, 23, 24, 25, 27, 28, 34, 39, 52], data [1, 7, 8, 24, 27, 28, 29, 32, 34, 49], tools and technologies [8, 27, 28, 34, 39], as well as processes [8, 24, 28, 33, 34, 39, 47]. However, none of the extracted findings were further evaluated with a larger sample of data experts. Previous large-scale evaluations in this field only observed the role of the unobserved decision quality in the ranking of experts [16] as well as the role of failure experiences in project profitability [41]. Moreover, only some of them are actually introduced as challenges of data-driven projects, e.g., misalignment between data stakeholders [39], data access, insufficient competence, the lack of appropriate digital tools, and the lack of a systematic process [8].

2.2. Expert interviews and choice of challenges

To design an effective survey for data experts, which is understandable, short, focused, to the point, and clearly structured, we followed the applied action design research methodology [46]. Based on qualitative

semi-structured interviews [38] with 13 experts representing different roles and competence areas such as strategic designers, data scientists, UX-Designers, or consultants, we conducted a total of 20 rounds of design, validation and refinement of the survey draft [46] to address the right issues effectively, e.g., in terms of wording, design, and ordering of questions. The process took approximately eight weeks. Some of the experts were involved twice to review the course of the interview drafts.

Known issues such as data access [7, 8, 32], data volume [1, 49], data quality [24, 27, 29, 32], legal issues [1], security issues [28], lack of technical expertise [12, 23] were adopted and supplemented through other ones. As for a systematic process [8], unstructured project execution was included. Related to misalignment between data stakeholders [39], the lack of organizational alignments/agility and conceptual distance between business strategies and implementation of analytics solutions were formulated. Insufficient competence [8] was concretized in the lack of hard skills (e.g., programming language, software knowledge), the lack of subject matter expertise, the lack of soft skills (communication, teamwork), and in the lack of understanding the way of making use of analytics. The lack of appropriate digital tools [8] was described as "required technology (including algorithms) was not available/ advanced enough". As a result of this phase, the focus could be deepened to understand (1a) how successful experts perceived their previous data-driven projects and (1b) what roles distinct phases and challenges played in the non-successful data-driven projects.

2.3. Study design

Our primary interest were experts who had been involved in at least one data-driven project. To target them, invitations were sent to known experts with some data background and/or positions in data-related departments. Experts were contacted via e-mail with the URL to the questionnaire and a confidentiality note, asking them to participate in the survey and to further distribute the link within their business social networks via e-mail, LinkedIn, and Twitter.

Respondents were briefly instructed only about the main objective of the questionnaire formulated as challenges in data-driven projects and were provided with the working definition of data-driven projects and some formal instructions to follow. For comparability reasons, data-driven projects were defined as "*projects working with large amounts of data that have to be processed with dedicated analytics software (not Excel) and coding skills*", as well as "*projects in which data is analyzed for solving business problems*". Further, we asked them to consider all data-driven projects the respondent has ever worked in, not specifically in the company or industry they currently work in.

When starting the survey, respondents were first asked about the industries for which their data-driven projects were carried out and their most frequent roles in these data-driven projects. The considered industries included manufacturing, public sector, information media and telecommunications, wholesale/retail, transport and logistics, banking and insurance, consulting, health care, and professional, scientific and technical services [2]. As for the expert roles, we included data scientist, developer, data architect, project manager, field/subject matter expert, business developer, and sales expert.

The next question focused on the percentage of such projects that were considered successful by either management or customer. To detect the opinion of the respondent (in potential contrast to what the project was officially considered as), respondents were further expected to estimate the percentage of data-driven projects that added real business value (e.g., process optimization, new data-based service) and were implemented in operations. We generally assume that data-driven projects can be considered to have added real business value when being implemented in operations or being used as a proof-of-concept (PoC).

Regarding the non-successful data-driven projects, i.e., those which did not add real business value and/or where the solution was not implemented in operations, respondents were asked to rate the impact of different phases, i.e., project management, user need understanding, business understanding/feasibility, data understanding, data preparation, modeling, business case evaluation, deployment (based on CRISP-DM [1, 5]) on the non-success of such data-driven projects. Finally, participants were asked to either rate the impact or give their implicit agreement to the impact of different challenges on the non-success of such data-driven projects.

4. Results

Our sample of 112 survey respondents consisted of data scientists with 27% in the first role and 6% in the second role, project managers (21% and 18%), field matter experts (10% and 12%), data architects (8% and 10%), developers (7% and 10%), business developers (7% and 10%), and sales experts (1% and 3%) (see Figure 1).

They participated in data-driven projects in various industries [2] such as manufacturing (23%), public sector (1%), information media and telecommunications (77%), wholesale/retail (14%), transport and logistics (30%), banking and insurance (22%), consulting (23%),

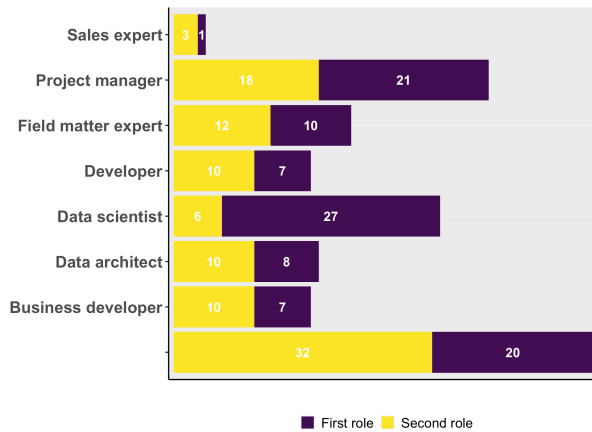health care (4%), and professional, scientific and technical services (1%).

Figure 1. First and second roles of study participants (x-axis: percentage of respondents)

Out of the 112 survey respondents, 18% and 23% stated that less than 25% and 25-50% of their data-driven projects created real business value in the past, respectively. Smaller proportions of our study participants, i.e., 12% and 16%, thought that less than 25% and 25-50% of their data-driven projects were considered successful by management or customer. Slightly more than three in ten (34%) say that their managers or customers regarded more than 75% successful (see Figure 2). Only two in ten (21%) of our respondents claim that more than 75% of their data-driven projects added a real business value. It seems that from the employee's perspective, more data-driven projects are considered successful by management or customer than really do provide business value.
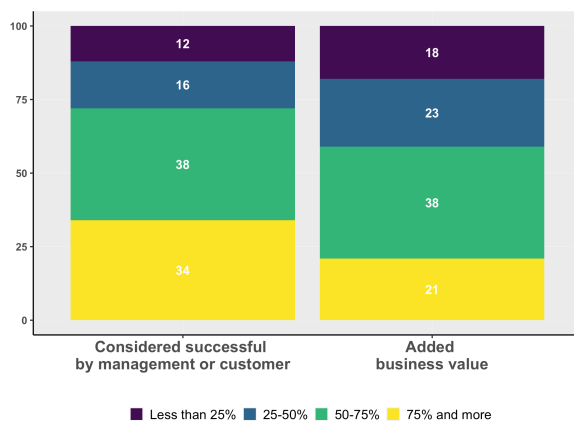

Figure 2. Difference between perceived official success and business value (y-axis: percentage of respondents)

Around three in ten study participants (29%) stated that 10-25% of data-driven projects were implemented in operations. The same applies to 25-50% (32%) and 50% and more data-driven projects (29%). One in five (22%) experienced this in less than 10% of data-driven projects. At this point, it should be stated that not all project solutions, more specifically data-driven solutions, aim to be implemented in operations. Sometimes, a solution's project goal is a proof-of-concept (PoC) that simply demonstrates the feasibility of a concept.
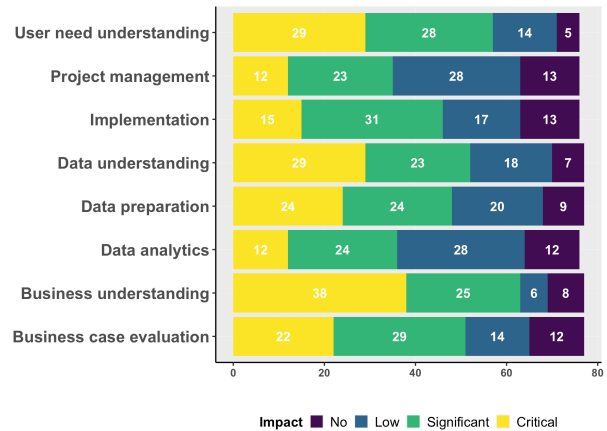

Figure 3. Perceived impact of project phase on non-success (x-axis: percentage of respondents)

Regarding the perceived role of distinct phases in the non-successful data-driven projects, business understanding is considered of critical and significant impact by 38% and 25% of respondents, respectively, followed by – concerning critical impact – understanding user needs (29% and 28%), data understanding (29% and 23%), data preparation (both 24%), business case evaluation (22% and 29%), implementation (15% and 31%), data analytics (12% and 24%), and project management (12% and 23%) (see Figure 3).

As for challenges, critical and significant impacts are mostly attributed to data quality (34% and 29%), data access (30% and 24%), budget/time (21% and 31%), cultural resistance (17% and 27%), lack of soft skills (16% and 24%), unstructured project execution (14% and 28%), communication with customer (12% and 35%), transfer to the customer (9% and 24%), and lack of hard skills (11% and 21%) (see Figure 4).
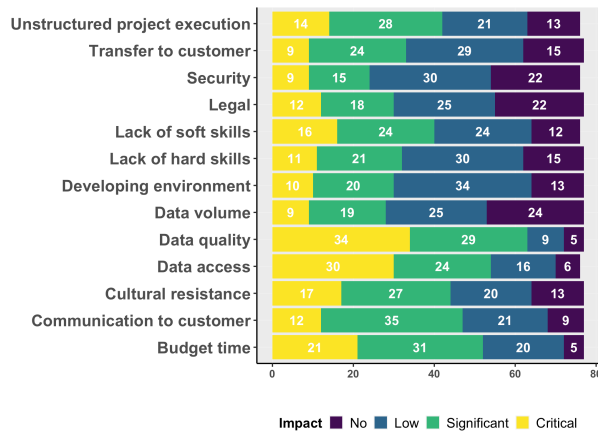
Figure 4. Perceived impact of challenges on non-success (x-axis: percentage of respondents)

54% of our respondents agree with the conceptual distance between business strategies and the implementation of analytics solutions being influential when it comes to the non-success in data-driven projects (see Figure 5). The statement that was the second most agreed to is the lack of organizational alignments/agility (49%), followed by the lack of subject expertise (42%), lack of analytics understanding (38%), non-involvement of data scientists into problem definition (34%), and lack of technical expertise (29%). Tools and technology are the least critical reasons for non-success, i.e., usage of inappropriate tools (21%), required technology unavailable (26%), and "the project was not deployable from the beginning" (25%).
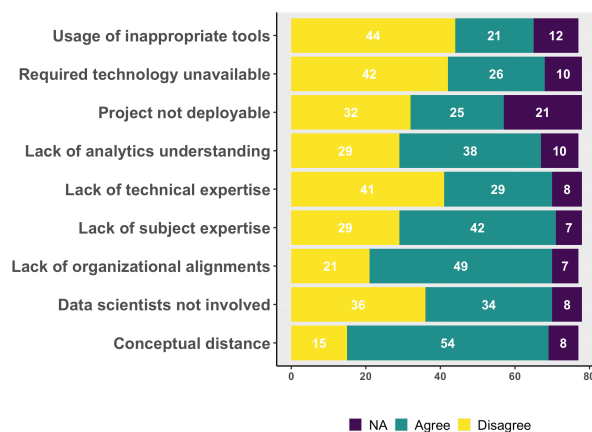

Figure 5. Perceived role of challenges on non-success (x-axis: percentage of respondents)

5. Discussion

The present work provided a systematically established frame of reference for the given problem statement, demonstrating the body of knowledge on the challenges of data-driven projects. Known issues such as data access [7, 8, 32], data volume [1, 49], data quality [24, 27, 29, 32], legal issues [1], security issues [28], and lack of technical expertise [12, 23] could be adopted without changes and supplemented through those previously reflected in the challenges of misalignment between data stakeholders [39], data access, insufficient competence, the lack of appropriate digital tools, and the lack of a systematic process [8]. Taken together, this gave rise to unstructured project execution, the lack of organizational alignments/agility, conceptual distance between business strategies and implementation of analytics solutions, the lack of hard skills (e.g., programming language, software knowledge), the lack of subject matter expertise, the lack of soft skills (communication, teamwork), the lack of understanding the way of making use of analytics, and not available and/or advance required technology (including algorithms).

Overall, all challenges derived and evaluated in our survey can be confirmed as being in line with the aspects previously reported in the literature. This is applicable to data access [7, 8, 32], data volume [1, 49], and data quality [24, 27, 29, 32], as well as legal issues [1], security issues [28], budget/time issues [47], and transfer to customer [23], communication issues with customer [23]. Further, usage of inappropriate tools, non-availability or non-advancement of the required technology (also including algorithms) [27, 34], or the lack of appropriate digital tools [8] have been already reported in prior work.

Next, the lack of subject matter expertise has been integrated, previously observed through the lack of business knowledge [12] and/or understanding of business-related data [4]. The lack of soft skills, e.g., communication, teamwork, revolves around interpersonal communication skills, and client and project management skills [23]. The lack of hard skills (e.g., programming language, software knowledge), insufficient competence [8] and the lack of technical expertise [12, 23] are caught within the aspects of the identification of talents, skills and certifications needed [34], experiential knowledge and skills to leverage business-related data [4], and data science expertise [6, 27, 37].

The lack of understanding of making use of analytics appears to be meant equated with the lack of understanding of how analytics can support achieving strategic goals [25]. Lack of organizational alignments and/or agility, as well as the conceptual distance between business strategies and implementation of analytics solutions could have been grasped in the misalignment between data stakeholders and be associated with the postulated lack of shared

understanding, lack of an organizational shared model, and lack of a shared language [39]. Non-involvement of data scientists into problem definition and/or business understanding completes the picture emerging from the lack of the right framing and data purpose [52] and carries specific reasons for the identification of business needs and involvement of stakeholders into the process [34], and established data user [24].

Finally, unstructured project execution can be found reflected in the lack of a systematic process [8], and/or matched decision-making culture and appropriate process and overall planning of data-driven projects [33, 47]. Cultural resistance within organizations can be found shown by the notions of team building and culture creation [34], however, also seen in part justified through the lack of an organizational mental model and the lack of a shared language [39].

In our study, 34% of respondents reported that more than 75% were considered successful by management, whereas only 21% found that more than 75% of data-driven projects added real business value. Potentially, managers considered data-driven projects rather successful in order to maintain their good image. These percentages cannot be compared directly due to ranges of percentages in this research and absolute percentages in other studies. Nevertheless, still in line with our findings, according to Capgemini in 2014, 73% of big-sized data-driven projects were not successful [7]. According to Gartner in 2018, 85% of big-sized data-driven projects are not expected to deliver business value [27]. In 2019, their predictions concerned 80% of data-driven projects in the area of artificial intelligence in 2020 and 80% of analytic insights in 2022 [50]. VentureBeat did not expect 87% of data science projects to achieve the production stage [48].

Learning from failed projects was found highly promising to succeed in the future [41, 43]. As for the perceived role of distinct phases in the non-successful data-driven projects, business understanding is regarded of critical and significant impact by 38% and 25% of respondents, respectively, followed by – with respect to critical impact – understanding user needs (29% and 28%). This indicates a necessity for considering the additional steps of business and user needs understanding when conducting data-driven projects. As part of UX (User Experience) research, the Empathy Map can be applied, distinguishing between what the user says out loud, thinks throughout the experience, how she or he behaves, and the emotional state [17]. A data scientist from Airbnb even goes so far as claiming that including a Design Thinking Leader to the traditional CRISP-DM will increase the success rate of a data science project by five to ten times [44].

Critical and significant impacts on the non-success of data-driven projects are most frequently attributed to *data-related* challenges such as data quality (34% and 29%) and data access (30% and 24%). Regarding the perceived role of distinct phases in the non-successful data-driven projects, data understanding (29% and 23%) and data preparation (both 24%) are seen of similar critical and significant impact by approximately every third respondent. Indeed, a data-driven project largely depends on data quality [26]. Capgemini [7] also regarded data access as the biggest challenge in data-driven projects, as the data sources of every 8 in 10 organizations have not been completely integrated across the departments. Another reason for this could be the current that data becomes increasingly heterogeneous and unstructured [13]. In the light of the BI&A (business intelligence and analytics) evolution, Chen et al. [6] observed a shift from structured (DBMS-based) content, to unstructured (web-based) content, to mobile and sensor-based content.

Furthermore, critical and significant impacts on the non-success of data-driven projects are attributed to rather *process-related* challenges, such as budget/time (21% and 31%), cultural resistance (17% and 27%), and unstructured project execution (14% and 28%). Stein (2015) [47] also states that the outcome and timing of data-driven projects are hard to plan precisely. Companies can further consider adopting a well-established uniform framework for handling data projects, e.g., KDD [14], CRISP-DM [1], Microsoft TDSP (Microsoft Team Data Science Process), or Design Thinking [52].

Finally, critical and significant impacts on the non-success of data-driven projects are also attributed to *people-related* challenges, including lack of soft skills (16% and 24%) and lack of hard skills (11% and 21%). One third agreed on the lack of technical expertise (29%) being critical. McKinsey Global Institute predicted alarming shortages within data scientists by 2019: The United States alone will face a deficit of 140,000 to 190,000 specialists with deep analytical skills and a shortage of 1.5 million data-savvy managers with the expertise to analyze Big Data to come to adequate decisions [6, 37]. Many companies have identified the need for hiring data scientists, while universities are rushing to initiate data-science programs [42].

Despite multiple insights from the systematic assessment presented in this paper, we see limitations to our current work due to limited time and space. First, our study is by design confined to the expert perceptions related to their overall experience in data-driven projects. Researchers might be willing to move beyond this "boundary" and focus on a specific sample of experiences. Our sample of data experts might be seen

somewhat skewed making our findings potentially not applicable to other samples of those practitioners significantly different than this one, e.g., working in countries or industries less advanced in a data-driven sense. Researchers might be interested to examine the specified issues based on a more representative sample of data experts and conduct comparisons across countries and industries. We observe that one of our core questions, i.e., "the project was not deployable from the beginning" has by far the most explicit not assessable answers (21%), followed by "data volume", "legal", and "security", which were found of no impact by 24%, 22% and 22% of study participants, respectively. This might be the result of an unclear formulation of the questions. Although the stated limitations neither impaired the purpose of the present study nor diminished its theoretical and practical value, we see these aspects as relevant for further examination.

6. Conclusions

This study examined the most crucial challenges that lead to the non-success in data-driven projects, as perceived by data experts. The current challenges in these kinds of projects were elicited and explored using a survey of the literature, qualitative expert interviews, and an online survey among experts of different industries and responsibilities. From a theoretical perspective, the findings enrich the current understanding of this rapidly growing research field and can be further used to build a foundation for constructs and operationalizations [36]. From a practical perspective, the findings create awareness for a detailed pallet of challenges and benefit organizations interested to overcome them.

The respondents evidently emphasize the clarification of the business goal as well as an understanding of user needs in order to carry out data-driven projects successfully. The perceived role of business understanding and user needs understanding are considered of critical and significant impact in the non-successful data-driven projects by larger proportions of participants and around every six in ten experts. We revealed a necessity for implementing the additional steps of user needs and business understanding in the process of finding data-driven solutions.

Further, our research shows that this understanding is of high relevance, especially in the beginning, when it comes to supporting the project team in clarifying the different objectives and making the solution valuable for the user. More than half of our respondents agree with the influence of the conceptual distance between business strategies and the implementation of analytics solutions on the non-success in data-driven projects.

There are also some clues derived that different stakeholders might differ in seeing data-driven projects as successful.

Only if addressing all three aspects of understanding, i.e., the user, the data, and the business, data-driven solutions will become more successful and valuable in the future.

References

[1] Baier, L., F. Jöhren, and S. Seebacher, "Challenges in the Deployment and Operation of Machine Learning in Practice", *Proceedings of the 27th European Conference on Information Systems (ECIS)*, (2019).

[2] Bain&Company, "Industry Insights", *Bain*, 2020. https://www.bain.com/insights/industry-insights/

[3] Bannerman, P.L., and P.W. Yetton, "Why Project Performance Varies: A Capability-Based Explanation", *PACIS 2011 Proceedings*, (2011).

[4] Butler, T., and D. Sammon, "Life in the Data Warehouse: A Case Study of Second Phase Implementation Problems and Solutions", *AMCIS 2000 Proceedings*, (2000).

[5] Chapman, P., J. Clinton, R. Kerber, T. Khabaza, C. Shearer, and R. Wirth, *CRISP-DM 1.0*, 1999.

[6] Chen, H., R.H.L. Chiang, and V.C. Storey, "Business Intelligence and Analytics: From Big Data to Big Impact", *MIS Quarterly 36*(4), 2012, pp. 1165–1188.

[7] Colas, M., I. Finck, J. Buvat, R. Nambiar, and R.R. Singh, "Cracking the Data Conundrum: How Successful Companies Make Big Data Operational", *Capgemini UK*, 2015. https://www.capgemini.com/gb-en/resources/cracking-the-data-conundrum-how-successful-companies-make-big-data-operational/

[8] Cronholm, S., H. Göbel, and P. Rittgen, "Challenges Concerning Data-Driven Innovation", *ACIS 2017 Proceedings*, (2017).

[9] Dahlberg, T., and A. Lagstedt, "There Is Still No 'Fit for All' IS Development Method: Business Development Context and IS Development Characteristics Need to Match", *Proceedings of Hawaii International Conference on System Sciences 2018 (HICSS-51)*, (2018).

[10] datascienceassn.org, *About Data Science*, Data Science Association, 2020.

[11] Davenport, T.H., and D.J. Patil, "Data Scientist: The Sexiest Job of the 21st Century", *Harvard Business Review*, 2012. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

[12] Debortoli, S., O. Müller, and J. vom Brocke, "Comparing Business Intelligence and Big Data Skills - A Text Mining Study Using Job Advertisements", *Business & Information Systems Engineering 6*(5), 2014, pp. 289–300.

[13] Dhar, V., "Data Science and Prediction", *Communications of the ACM 56*(12), 2013, pp. 64–73.

[14] Fayyad, U., G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases", *AI Magazine 17*(3), 1996, pp. 37–37.

[15] Fernández, A., "Data-Driven versus Data-Centric", *Stratio*, 2018. https://www.stratio.com/blog/datadriven-versus-datacentric/

[16] Geva, T., and M. Saar-Tsechansky, "Who's A Good Decision Maker? Data-Driven Expert Worker Ranking under Unobservable Quality", *ICIS 2016 Proceedings*, (2016).

[17] Gibbons, S., "Empathy Mapping: The First Step in Design Thinking", *Nielsen Norman Group*, 2018. https://www.nngroup.com/articles/empathy-mapping/

[18] Ginsberg, J., M.H. Mohebbi, R.S. Patel, L. Brammer, M.S. Smolinski, and L. Brilliant, "Detecting Influenza Epidemics Using Search Engine Query Data", *Nature 457*(7232), 2009, pp. 1012–1014.

[19] Han, J., M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Haryana, India; Burlington, MA, 2011.

[20] Hartmann, P.M., M. Zaki, N. Feldmann, and A. Neely, *Big Data for Big Business? A Taxonomy of Data-driven Business Models used by Start-up Firms*, University of Cambridge, Cambridge, 2014.

[21] Hastie, T., R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction, Second Edition*, Springer-Verlag, New York, 2009.

[22] Heilbron, J.L., *The Oxford Companion to the History of Modern Science*, Oxford University Press, Oxford ; New York, 2003.

[23] Hoxmeier, J., and M. Lenk, "Service-Learning in Information Systems Courses: Community Projects that Make a Difference", *Journal of Information Systems Education 14*(1), 2003, pp. 91–100.

[24] Jensen, M., P.A. Nielsen, and J.S. Persson, "Managing Big Data Analytics Projects: The Challenges of Realizing Value", *Proceedings of the 27th European Conference on Information Systems (ECIS)*, (2019).

[25] Kampakis, S., "What are the differences between data-driven, data-informed and data-centric?", *The Data Scientist*, 2018. https://thedatascientist.com/data-driven-data-informed-data-centric/

[26] Karpatne, A., G. Atluri, J. Faghmous, et al., "Theory-Guided Data Science: A New Paradigm for Scientific Discovery from Data", *IEEE Transactions on Knowledge and Data Engineering 29*, 2017, pp. 2318–2331.

[27] Kidd, C., "Why does Gartner predict up to 85% of AI projects will 'not deliver' for CIOs? – BMC Blogs", 2018. https://www.bmc.com/blogs/cio-ai-artificial-intelligence/

[28] Koronios, A., J. Gao, and S. Selle, "Big Data Project Success – A Meta Analysis", *PACIS 2014 Proceedings*, (2014).

[29] Kühne, B., and T. Böhmann, "Data-Driven Business Models - Building the Bridge Between Data and Value", *Proceedings of the 27th European Conference on Information Systems (ECIS)*, (2019).

[30] LaValle, S., E. Lesser, R. Shockley, M.S. Hopkins, and N. Kruschwitz, "Big Data, Analytics and the Path From Insights to Value", *MIT Sloan Management Review*, 2011. https://sloanreview.mit.edu/article/big-data-analytics-and-the-path-from-insights-to-value/

[31] Lazer, D., R. Kennedy, G. King, and A. Vespignani, "Big Data. The Parable of Google Flu: Traps in Big Data Analysis", *Science (New York, N.Y.) 343*(6176), 2014, pp. 1203–1205.

[32] Lennerholt, C., and J. van Laere, "Data Access and Data Quality Challenges of Self-Service Business Intelligence", *Proceedings of the 27th European Conference on Information Systems (ECIS)*, (2019).

[33] Li, M., Z. Zhang, and Z. Hu, "Big Data-Driven Technology Innovation: Concept and Key Problems", *WHICEB 2017 Proceedings*, (2017).

[34] Liu, Yi., H. Han, and J. DeBello, "The Challenges of Business Analytics: Successes and Failures", *Proceedings of Hawaii International Conference on System Sciences 2018 (HICSS-51)*, (2018).

[35] Maass, W., J. Parsons, S. Purao, V. Storey, and C. Woo, "Data-Driven Meets Theory-Driven Research in the Era of Big Data: Opportunities and Challenges for Information Systems Research", *Journal of the Association for Information Systems 19*(12), 2018.

[36] MacKenzie, S.B., P.M. Podsakoff, and N.P. Podsakoff, "Construct Measurement and Validation Procedures in MIS and Behavioral Research: Integrating New and Existing Techniques", *MIS Quarterly 35*(2), 2011, pp. 293–334.

[37] Manyika, J., M. Chui, B. Brown, et al., *Big Data: The Next Frontier for Innovation, Competition, and Productivity*, McKinsey, 2011.

[38] Myers, M.D., and M. Newman, "The Qualitative Interview in IS Research: Examining the Craft", *Information and Organization 17*(1), 2007, pp. 2–26.

[39] Nagle, T., and D. Sammon, "The Data Value Map: A Framework for Developing Shared Understanding on Data Initiatives", *Proceedings of the 25th European Conference on Information Systems (ECIS)*, (2017), 1439–1452.

[40] Okoli, C., "A Guide to Conducting a Standalone Systematic Literature Review", *Commun. Assoc. Inf. Syst.*, 2015.

[41] Pflügler, C., T. Malzer, T. Jäschke, M. Wiesche, and H. Krcmar, "'Do I Want to Have Losers In My Team?' - A Quantitative Study of Learning from IT Project Failure", *Proceedings of Hawaii International Conference on System Sciences 2018 (HICSS-51)*, (2018).

[42] Provost, F., and T. Fawcett, "Data Science and its Relationship to Big Data and Data-Driven Decision Making", *Big Data 1*(1), 2013, pp. 51–59.

[43] Rhaiem, K., and N. Amara, "Learning from Innovation Failures: A Systematic Review of the Literature and Research Agenda", *Review of Managerial Science*, 2019.

[44] Roumeliotis, G., "Design Thinking for Data Scientists", 2015. https://www.linkedin.com/pulse/design-thinking-data-science-george-roumeliotis

[45] Saar-Tsechansky, M., "Editor's Comments: The Business of Business Data Science in IS Journals", *MIS Quarterly 39*(4), 2015, pp. iii–vi.

[46] Sein, M., O. Henfridsson, S. Purao, M. Rossi, and R. Lindgren, "Action Design Research", *Management Information Systems Quarterly 35*(1), 2011, pp. 37–56.

[47] Stein, R.M., "Why Managing Data Scientists Is Different", *MIT Sloan Management Review*, 2015. https://sloanreview.mit.edu/article/why-managing-data-scientists-is-different/

[48] VentureBeat, "Why do 87% of data science projects never make it into production?", 2019. https://venturebeat.com/2019/07/19/why-do-87-of-data-science-projects-never-make-it-into-production/

[49] Volk, M., S. Hart, S. Bosse, and K. Turowski, "How much is Big Data? A Classification Framework for IT Projects and Technologies", *AMCIS 2016 Proceedings*, (2016).

[50] White, A., "Our Top Data and Analytics Predicts for 2019", 2019. https://blogs.gartner.com/andrew_white/2019/01/03/our-top-data-and-analytics-predicts-for-2019/

[51] Wilson, D., "A Tale Of Two Projects: Why IT Projects Fail (And Why They Sometimes A", *ACIS 2004 Proceedings*, (2004).

[52] Woods, R., "A Design Thinking Mindset for Data Science", *Medium*, 2019. https://towardsdatascience.com/a-design-thinking-mindset-for-data-science-f94f1e27f90

[53] Wykes, N., "Why you should have a data centric approach", 2017. https://www.linkedin.com/pulse/why-you-should-have-data-centric-nicolas-wykes

[54] Zaki, M.J., and J. Wagner Meira, *Data Mining and Analysis: Fundamental Concepts and Algorithms*, Cambridge University Press, 2014.