Breaking the vicious circle: A case study on why AI for software analytics and business intelligence does not take off in practice☆

Iris Figalist [a],[*], Christoph Elsner [a], Jan Bosch [b], Helena Holmström Olsson [c]

[a] Siemens Corporate Technology, Otto-Hahn-Ring 6, 81739 Munich, Germany
[b] Department of Computer Science and Engineering, Chalmers University of Technology, Hörselgången 11, 412 96 Göteborg, Sweden
[c] Department of Computer Science and Media Technology, Malmö University, Nordenskiöldsgatan, 211 19 Malmö, Sweden

ARTICLE INFO

ABSTRACT

In recent years, the application of artificial intelligence (AI) has become an integral part of a wide range of areas, including software engineering. By analyzing various data sources generated in software engineering, it can provide valuable insights into customer behavior, product performance, bugs and errors, and many more. In practice, however, AI for software analytics and business intelligence often remains at a prototypical stage, and the results are rarely used to make decisions based on data. To understand the underlying causes of this phenomenon, we conduct an explanatory case study consisting of and interview study and a survey on the challenges of realizing and utilizing artificial intelligence in the context of software-intensive businesses. As a result, we identify a vicious circle that prevents practitioners from moving from prototypical AI-based analytics to continuous and productively usable software analytics and business intelligence solutions. In order to break the vicious circle in a targeted manner, we identify a set of solutions based on existing literature as well as the previously conducted interviews and survey. Finally, these solutions are validated by a focus group of experts.

1. Introduction

Artificial intelligence (AI) has been known as a powerful tool to extract valuable information from data for quite some time. Consequently, this led to a growing interest of software-intensive companies in performing software analytics and business intelligence based on Artificial Intelligence (AI4SABI) in order to acquire meaningful and relevant information about their products and processes. Many advances have already been made in the analysis of software engineering data and data-driven product development (Olsson and Bosch, 2019; Buse and Zimmermann, 2012). For instance, customer behavior and product usage are being analyzed to provide insights for developers, software architects, product managers and even business-oriented roles such as sales (Figalist et al., 2019a,b).

However, if no measures or actions are initiated based on its results, the analysis of customer behavior remains insufficient (Butz and Goodstein, 1996). Prototypically implementing an analysis is simple, but in order to truly base decisions on it, an analysis must be delivered continuously and automatically. After providing AI4SABI solutions in Figalist et al. (2019a,b), and Figalist

et al. (2021), we observed on several occasions that analyses often get stuck at this point and are not pursued further.

To address this situation, our study focuses on identifying challenges and appropriate solutions for the application of data analytics and in particular AI in the context of software analytics and business intelligence. In order to identify the challenges and the underlying causes of our observation, we conduct an explanatory case study consisting of an interview study and a survey. Thereby we investigate two different perspectives, namely that of the (potential) users (e.g. product managers, software architects, sales) and that of providers (e.g. data scientists, data engineers, operations engineers, software architects) of AI4SABI solutions. This work builds on an earlier publication (Figalist et al., 2020) that only covers the challenges. In this paper we now add to it by additionally reviewing existing literature as well as the conducted interviews and survey for solutions that address the presented challenges. We categorize the identified solutions and conduct expert interviews for a validation of the same.

The contribution of this paper is three-fold. First, we identify the challenges for applying AI4SABI from two different perspectives, specifically the users and the providers of AI4SABI. This contributes to an improved mutual understanding. Second, we introduce the concept of a vicious circle consisting of key drivers why AI4SABI is rarely realized in a productized way. This circle supports involved stakeholders in comprehending why their

AI4SABI solutions might get stuck in a prototypical stage. Third, we present a set of solutions mapped to the key drivers in the vicious circle to support practitioners in taking a targeted approach at breaking the circle.

The remainder of the paper is structured as follows: First, the background of this study is provided in Section 2. The research method and study design is described in Section 3, while the findings of the interview study and the survey are presented in Sections 4 and 5. In Section 6 the derived challenges forming a vicious circle are outlined before investigating and presenting potential solutions in Section 7. The threats to validity are discussed in Section 8 followed by the conclusion in Section 9.

2. Background

The following subsections give an overview of the existing literature in the fields of *software analytics and data-driven software engineering* as well as the *challenges regarding data analytics and AI*. Both topics are already quite well researched as individual areas. However, in our study, we feel it is important to put both areas into a common context, since the products analyzed by software analytics or business intelligence are often not related to AI or data analysis, which generally impedes a successful implementation of AI4SABI.

2.1. Software analytics and business intelligence

The term software analytics (SA) depicts the application of data analysis to software data in order to generate insights for various stakeholders, from software engineers to managers, that ultimately support their decision making (Buse and Zimmermann, 2012; Menzies and Zimmermann, 2013). Related to this, the field of business intelligence (BI), sometimes also referred to as business analytics, utilizes data mining techniques on operational data to derive information for managerial decision making (Negash and Gray, 2008).

Analyzing customer behavior can be a valuable asset in making strategic decisions on product improvement and evolution (Marciuska et al., 2014; Olsson and Bosch, 2013). One of the most critical challenges is choosing the right metrics to serve a specific purpose or achieve a specific goal. Therefore, each metric must be carefully evaluated in terms of its suitability and explanatory power (Kaner and Bond, 2004).

At the same time, it is considered crucial, but also difficult, to ask the right questions and specify information needs (Olsson and Bosch, 2015; Figalist et al., 2019a). In order to really create impact by acting on the results, it is decisive to select and prioritize the right questions (Kim et al., 2016). Different roles have very different information needs, and while the information needs of technical stakeholders are well-researched, the needs of managers are less transparent (Buse and Zimmermann, 2012).

Moreover, the visualization of data and results in an interpretable way has proven to be quite complex, but nevertheless crucial to generate results in an understandable and explainable way. This is essential to establish a certain level of trust and comprehensibility required to act on the basis of the provided information (Buse and Zimmermann, 2012; Menzies and Zimmermann, 2018; Dam et al., 2018; Kim et al., 2016). Furthermore, there is still a lot of skepticism among stakeholders who do not trust the data and the analyses' outcomes, and feel like going data-driven is risky and time-consuming (Olsson and Bosch, 2019).

From data generation to data collection, data storage, data analysis, visualization, and ultimately to value creation and smart decision making, it is a long and complex process (Saggi and Jain, 2018). In software-intensive businesses, data is generated from various sources. It can be machine-generated (e.g. network data), human-generated (e.g. log or usage data) or business-generated (e.g.transactional data). The distribution of available data sources is often accompanied by a lack of sharing data among individual teams (Fabijan et al., 2016; Cito, 2016). In addition to that, all data sources have to be understood and quality issues must be resolved. Following this, a variety of preprocessing techniques are required (e.g. cleaning, integration, transformation) before an analysis can be performed that generates meaningful insights (Saggi and Jain, 2018; Hu et al., 2014). Throughout this entire process it can be challenging to deal with complex types and structures of data from multiple sources, to filter, process and store it in a usable format, and to select appropriate algorithms to analyze it (Saggi and Jain, 2018).

2.2. Artificial intelligence — Definition and challenges

Artificial intelligence can be defined as "the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings. The term is frequently applied to the project of developing systems endowed with the intellectual processes characteristic of humans, such as the ability to reason, discover meaning, generalize, or learn from past experience" (Copeland, 2020). Machine learning (ML) is "a type of artificial intelligence in which computers use huge amounts of data to learn how to do tasks rather than being programmed to do them" (Machine Learning, 2021). ML can be applied to various problems, such as text classification, natural language or speech processing, or computer vision (Mohri et al., 2018; Jordan and Mitchell, 2015). Standard learning tasks of ML include classification, regression, ranking, and clustering (Mohri et al., 2018).

With its increase in popularity, artificial intelligence soon became an integral part of SA and BI solutions (Chen et al., 2012; Dam et al., 2018). In the context of our study, we define AI4SABI as *the application of artificial intelligence methods to software or operational data in order to address problems or generate insights in the area of software analytics or business intelligence*. In recent work, we have implemented a variety of use cases for AI4SABI, such as recommendation systems for API refactoring based on genetic algorithms or customer churn prediction models based on operational usage data using classification techniques (Figalist et al., 2021).

However, working with large amounts of data and applying complex algorithms to it often entails a number of challenges. Three types of AI-related challenges have been identified in the existing literature: development-, production- and organization-related challenges (Arpteg et al., 2018). For instance, a lack of transparency makes it difficult to understand machine learning models, debug code and estimate effort. Careful monitoring of frequent updates is required, especially with regard to changes or additions in data sources. Furthermore, building and using machine learning systems usually depends on a variety of components (e.g. infrastructure, data pipelines, visualizations). As a result, it often requires a collaboration of several teams and roles within teams that do not always share the same ideas and priorities. Therefore, it can be very complicated to align the teams and overcome cultural differences (Arpteg et al., 2018).

Similarly, a large-scale study of data scientists reveals that the key problems they face are mostly either data (quality, availability and preparation), analysis (scaling and evaluating different models) or humans (communicating results and convincing teams of their value) (Kim et al., 2017). Moreover, the code that analyzes data represents only a very small fraction (5%) of the overall system. A significant amount of effort is also put into configuration, data collection, feature extraction, data verification, infrastructure, monitoring and other activities (Sculley et al.,

2015). Prototypes in Data Science Projects Oftentimes, data science projects start out as prototypical implementations to try out new ideas and get some first results fast (Provost and Fawcett, 2013; Sculley et al., 2015). This also applies to data science projects in software analytics (Misirli et al., 2013) and business intelligence (Thierauf, 2001). A major downside of this is that the results of small-scale prototypes often do not reflect the results in reality (Sculley et al., 2015). After implementing several use cases of AI4SABI in Figalist et al. (2021), we observed that, on the one side, prototypes are a powerful tool to explain and convey ideas to stakeholders as an analysis becomes more tangible if there are some first results to look at. On the other side, the majority of analyses only provide the desired value if executed in an automated and continuous way (e.g. customer churn prediction models). However, the transformation of a prototypical analysis into an automated and deployed analysis is difficult and time-consuming (Misirli et al., 2013; Sculley et al., 2015). This results in "prototypes or early models that lack the final deployment step, and hence, any real-world impacts" (Misirli et al., 2013).

3. Research method & study design

In order to investigate the application of AI4SABI in a real-life contexts, case study research was selected as the research method for our study. We conducted an explanatory case study that was designed by following the guidelines presented by Runeson and Höst (2009). The case study consisted of a qualitative interview study and a quantitative survey. The combination of qualitative and quantitative research methods has proven to be a successful means of obtaining useful information that could not be obtained from individual methods alone (Seaman, 1999). For this reason, we applied a mixed-method approach in our study following a concurrent triangulation strategy (Easterbrook et al., 2008). The objective of the study is defined as (1) *the identification of challenges and key drivers why AI4SABI is rarely productized (automated and deployed) and utilized for decision-making in practice* and (2) *the identification of potential solutions to address these challenges*. As a result, the studied case is the application of AI4SABI within real-life product teams. We aim at addressing the following research questions:

(1) RQ1: What are the key drivers for analyses not being productized after the prototypical stage?
(2) RQ2: How can the identified challenges and key drivers be addressed?

After summarizing the results of the interview study and the survey, each of the research questions is addressed in a dedicated section (Sections 6 and 7).

The case company is a large, industrial company with around 293.000 employees and multiple subsidiary companies. The case company and its subsidiaries offer a diverse product portfolio in a variety of different domains, such as industry, infrastructure, mobility, healthcare, and energy. It is an international company with product teams that are spread all over the world (mostly in Europe, USA, India, and China).

In order to identify the challenges, we conducted both an interview and a quantitative survey as part of our case study. We then compared the results in order to confirm and cross-validate our findings.

In a next step, we reviewed existing literature and the previously conducted interviews and survey for potential solutions. These solutions were then categorized, linked to the challenges and validated by a focus group of experts. The overall research process is illustrated in Fig. 1.
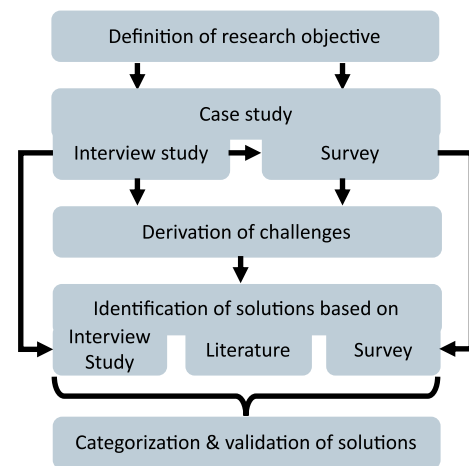

Fig. 1. Research process.

3.1. Interview study

The objective of the interview study is equal to the objective of the overarching case study. Semi-structured interviews were selected as a direct form of data collection. Specifically, we interviewed twelve stakeholders who are either users or providers of AI4SABI. Each participant either belonged to one of two products (*Product A* or *Product B*) or is a consultant to a number of product teams.

3.1.1. Product descriptions

Product A is a platform provider operating in the healthcare domain. The platform hosts several platform-internal and external medical applications for its $\sim$2700 customers. *Subproducts A_{1-3}* in our interviews represent internal applications of the platform. The platform and its internal applications are developed and operated by six modularized teams with sizes varying between 5 and 25 team members. Each team is either responsible for a specific platform component or a specific application that is hosted on the platform. The teams use Azure DevOps[1] and its services for the development and operations of their product. Moreover, they have a dedicated operations team that collects performance, usage, and sales data of the platform and its applications in order to create dashboards for different stakeholders and make it available to all other teams for further analysis. They use Microsoft Power BI[2] for the creation of dashboards and visualization of their analyses. They use both, business intelligence and software analytics, to generate insights for their stakeholders. The majority of their analyses are not AI-based. For example, as part of their business intelligence dashboards for their sales team, they provide statistics of the usage rates of customers clustered by the time to their subscription's expiration date. They use software analytics to generate information for their technical stakeholders, e.g. developers and architects. As an example, they analyze and visualize the navigation flows of customers to support the development teams in understanding whether their component or application is used as intended. AI-based examples that they have recently started working on are customer churn prediction models based on usage data and customer satisfaction classification based on service quality and web usage metrics. They have successfully productized their non-AI-based analyses and are now in the process of doing the same

---

Table 1
Interviewees grouped by products and function.

	Product A	Product B	Consultant (multiple products)
Users of AI4SABI	Product manager Product owner (Sub-product A_1) Product manager (Subproduct A_2) Product owner (Sub-product A_3)	Product owner Demand manager	Software architect (i.a. working for Product B)
Providers of AI4SABI	Product owner (for operations) Operations engineer Data scientist	Software architect	Software architect (i.a. working for Product B)

for their ML models. In the future, they aim at expanding their analyses and fully productizing their models.

Product B is a platform provider operating in the industrial domain. The platform serves ~800 customers and hosts multiple company-internal and external applications for industrial device management and analysis. Similar to *Product A*, the platform and its applications are being developed by modularized and distributed teams. In total, around 330 individuals are involved in the development of the platform with similar team sizes as *Product A*. The teams are following DevOps principles for the development and operation of their platform and applications. The platform provider setup a dedicated team for logging, processing and storing all requests users make to the platform and its applications. Their goal is to enable the other teams to extract continuous feedback based on the usage of their platform component or application. The platform's sales team works with a team of data scientists that creates BI dashboards displaying statistics of the customers' usage and sales data. In addition to that, a customer churn prediction model was trained based on the customers' usage data. However, the latest predictions of the model are not yet integrated in the dashboard. In the area of SA, the platform's security team worked with a company-internal research team who applied pattern detection to the usage data (specifically the requests authentication headers) in order to detect wrongly assigned permissions. To use the analyses on a continuous basis, the stakeholders would like to automate and deploy the analyses in the near future.

3.1.2. Interviewees

The interviewees can be categorized into two groups: users or providers of AI4SABI. In order to identify the interviewees for our study, we sent out a description of the study and an invitation to participate to contact persons of various platforms and applications within the case company who shared it among their teams. We received twelve responses from ten different teams distributed across two platform providers. The interviewees were selected due to their interest in either using or providing AI4SABI solutions. All interviewees from the provider group had already experience in applying AI in the context of SA or BI. In the user group, not all participants had hands on experience in using AI4SABI. However, this was not an exclusion criteria for us since the reasons for this were also relevant for our case study. From the user perspective we interviewed two product managers, three product owners, one demand manager, and one software architect while the provider perspective was investigated by interviewing one product owner (for operations), one operations engineer, one data scientist, and two software architects. An overview of the interviewees is summarized in Table 1.

Before the interviews were conducted, two interview guides were prepared, one for each group (users and providers). In addition, each interview guide was slightly adapted to the respective role of the interviewee. The interview guides for the

users of AI4SABI covered the following topics: current process of decision making in their job; current status of using data as feedback and how data and information is collected and shared within and across teams; and experiences with and views on data-driven feedback and decision making. The interview guides for the providers included: the process of building and providing AI4SABI solutions (including data collection and processing, implementation, automation and deployment of analysis, preparation of results); the organizational and human-related challenges they face; and technical difficulties they encounter.

3.1.3. Interview process

At the beginning of each interview, the interviewee was asked for permission to record the interview. The interviews lasted between 35 min and one hour, and were transcribed afterwards. We applied a thematic analysis approach (Maguire and Delahunt, 2017) for analyzing and coding the interview transcripts. The objective was to identify key challenges in working in a data-driven way that prevent a successful application of AI4SABI. The high-level themes used for thematic coding were (1) current decision-making processes; (2) current state of data collection; (3) current state of data analysis; (4) perceived usefulness of AI4SABI; (5) mindset towards data-driven decision-making; (6) people-related challenges (e.g. skepticism, interpretability); (7) organizational challenges (e.g. priority, resources); and (8) data-related challenges (e.g. data model, data quality). The coding itself was performed by the first author. The results were shared and discussed with the other authors to avoid any misinterpretations. Finally, the findings of each interview were summarized and sent to each interviewee for further validation.

We achieve triangulation by (1) interviewing multiple roles with opposite points of view (users and providers of AI4SABI); (2) conducting interviews within multiple independent product teams; and (3) enriching the results by an additional quantitative data source (survey). In addition to that, the transcript and summaries were also shared and discussed among the researchers to get several interpretations and perspectives in order to avoid misunderstandings and to make sure to cover all important aspects.

3.2. Survey

For the execution of the survey we followed the guidelines of Kitchenham and Pfleeger (2001-2003) and Ciolkowski et al. (2003). In order to achieve the overall research goal, the survey focused on the current state of applying data-driven approaches in real-world projects, as well as on the experiences and challenges in providing and using customized AI4SABI solutions. The target population of the survey were stakeholders in the field of software engineering who are either (potential) users or providers of AI4SABI. To conduct the survey, we created

an online questionnaire with a total of 29 questions that were divided into three parts: (1) personal background; (2) current status of projects with regard to data-driven approaches and analytics tools; and (3) experiences and challenges in implementing, automating, deploying, and leveraging AI4SABI. The first part consisted of three multiple choice questions regarding the participants' role, their experience, and the tasks they perform as part of their job. The second part comprised 8 questions in total, two of them multiple choice (e.g. "Do you or others in your (main) project's team use any analytics tools to gain insights from usage and/or runtime data?"), three single choice (e.g. "On a five point scale, how satisfied are you or the ones using it with the preparation and presentation of the results [of these analytics tools]?"), and three open questions (e.g. "If applicable and not mentioned above, what additional types of data are being collected?"). The third part consisted to 18 questions in total. One multiple choice question: "Do you or others in your (main) project's team manually analyze data?". The answers to this question were divided into who in the team is manually analyzing data (the respondent or someone else in the team) and, if applicable, into the type of analytics that is being applied, e.g. basic analytics, such as simple statistics, or advanced analytics techniques, such as machine learning. In addition to that, the respondents were asked to rate certain challenges or the difficulty of certain tasks on a 5-point Likert scale (15 single choice questions, e.g. "On a five point scale, based on your experience, how difficult is the transformation of a prototypical analysis into an automated, continuous analysis?"). Finally, we included two open questions to gain insights in challenges or difficulties that we might have missed (e.g. "If you consider it hard or extremely hard, what do you consider particular difficult?").

In the third part, the results of the interview study were partially used as a basis for the survey questions, especially for rating the adverse factors on the application of data analytics in the respondent's projects that were preciously identified in the interviews and literature. The majority of the questions were based on 5-point Likert scales. As the backgrounds of the participants are very diverse, there was always the option to abstain if no knowledge in this area was available. We decided to send out the same questionnaire to both groups since a clear separation of the two is often not feasible. In fact, before sending out the questionnaire, we collected feedback on the questions within a department of the case company and learned that people oftentimes take on multiple roles. Therefore, they could not clearly be assigned to either the users or providers of AI4SABI since that depends on the individual role.

The link to the survey was sent via email to 1596 individual stakeholders of several divisions across our case company. We received a total of 230 responses which equals a response rate of 14%. The survey results were investigated from two angles. First, we described the distribution of answers for all single-choice and multiple-choice questions. The majority of questions were based on 5-point Likert scales (e.g. *not at all*, *slightly*, *moderately*, *very*, or *extremely* difficult/influential) which made them easy to compare. Second, we examined the responses in the free text fields and extracted and categorized relevant statements similar to the coding done in the interview study. For conducting the survey, we had to use the company-internal survey tool. Unfortunately, this tool does not allow an evaluation of results across questions but only provides the answers for each question independently. For this reason, we were not able to analyze the dependencies between questions and factors.

3.3. Derivation of challenges and key drivers

In order to identify the challenges and key drivers why AI4SABI is rarely productized and utilized for decision-making in practice, we investigated and compared the results of the interview study and the survey. In a first step, the high-level themes identified during coding of the interviews were examined in more detail. All statements relevant for addressing our research questions were extracted for each of the themes. In a next step, we grouped the statements within each theme into further sub-themes (e.g. from data-related challenges to data quality related challenges). Most of the challenges were identified in the themes *people-related challenges*, *organizational challenges*, and *data-related challenges*. However, other themes sometimes provided further explanations for certain situations and problems (e.g. in *perceived usefulness of AI4SABI* or *mindset towards data-driven decision-making*). In these cases, the relations were also documented.

In a second step, we extracted and categorized the findings of the survey. For example, we examined the ratings of the twelve adverse factors and calculated the mean and median values of each response category across all factors. The mean values for the response categories *very* or *extremely* influential were 29.8% and 8.6%, while their median values were 28% and 9.5% respectively. Based on this, we set the thresholds for relevant factors to 30% and 10% for the categories *very* and *extremely* influential. Factors that exceeded one or both of the thresholds were highlighted as especially influential. In addition to that, the coded responses of the free text fields were grouped into subcategories similar to the sub-themes of the interview study.

Finally, we compared and mapped the results of the interview study to the results of the survey in order to cross-validate our findings. During the analysis, we observed that there are certain key challenges mentioned in the interviews and the survey that act as key drivers influencing each other. Based on this observation, we were able to identify a vicious circle that impedes a successful application of AI4SABI.

3.4. Identification and categorization of solutions

For the identification of potential solutions we used three different sources of information: existing literature, the interviews conducted as part of the interview study, and the survey responses.

In a first step, we queried common scientific libraries (IEEXplore, ACM Digital Library, ScienceDirect, Springer Link) using search terms related to the previously identified challenges. This process was conducted as a systematic literature review, but as a literature search in order to identify a list of key publications. Out of these publications, we extracted concrete actions that authors describe to improve situations related to the presented challenges as potential solutions and assigned each solution to the corresponding challenge.

Even though the interviews were primarily designed for identifying the challenges faced by practitioners, it also invited the interviewees to share their experience and expertise with regards to the actions and solutions they had found useful to address these challenges. Therefore, we examined the interview transcripts from an additional angle and extracted the proposed actions and solutions before also linking them to the respective challenges.

Similarly to the interview study, several survey respondents used the free text fields to share their experiences and solutions to specific problems. These solutions were also extracted and assigned to the challenges.

In a next step, we grouped the identified solutions of each challenge by their similarity and assigned related solutions to overarching solution categories.

Table 2
Overview of expert interviewees.

Interviewee	Product	Qualification
Data scientist	Product A	Analyzes data & communicates with stakeholders
Operations engineer	Product A	Provides infrastructure for AI4SABI
Operations manager	Product A	Builds dashboards for customer insights & communicates with stakeholders
Product owner for operations	Product A	Intermediate between data scientists and stakeholders
Software architect (1)	Product B	Builds AI4SABI solutions (infrastructure analysis)
Software architect (2)	Product B	Builds AI4SABI solutions (infrastructure analysis)

3.5. Expert interviews

In order to validate the solutions we identified, we interviewed six experts who have been working on different topics in the area of AI4SABI for multiple years. Four of them have already participated in the interviews of the previously conducted interview study. An overview of the interviewees and their qualification is provided in Table 2.

In the beginning of each interview the interviewees were given a short introduction into the background of our study and the challenges we had identified. Next, we presented the categorized solutions and explained how they relate to the respective challenges. Based on this, the experts gave their assessment of the solutions' usefulness. After discussing the solutions in detail, the experts were also asked to rate the usefulness of the solution categories on a scale from 1 (not useful at all) to 5 (very useful). Each interview lasted between 30 min and one hour.

4. Interview study results

In order to identify the underlying causes of why AI-based analyses for SA and BI often get stuck after a prototypical implementation, we conduct twelve interviews with stakeholders involved in the software engineering process. The following subsections outline the perspectives of two groups (users and providers) and the challenges they face when trying to establish data-driven ways of working.

4.1. Group I — Using AI4SABI solutions

The first group of interviewees consists of one software architect, one demand manager, two product managers, and three product owners. All of them are working in different teams of the two products. In addition to *Product B*, the software architect also works as a consultant for additional projects.

4.1.1. Recognizing value

All of the interviewees are very open minded towards AI4SABI and can see the value and benefits in developing software in a data-driven manner. For instance, the product manager of *Product A* explains that "sometimes what the stakeholders [within the product teams] are asking for is not what the customer really wants". Therefore, data-driven feedback of customers would be a valuable asset for her to make decision on product or feature evolution. Furthermore, the product owner and demand manager

of *Product B* both emphasize the difficulty of receiving early feedback from customers and, therefore, see the value in additionally collecting and analyzing data-driven feedback. A very similar statement is also given by the product owner of *Subproduct A_3* and the product manager of *Subproduct A_2* who would like to have more information on how customers are using the product's features.

> **Findings IV.1**: Stakeholders can see value and benefit in AI4SABI (6/7); current need for data-driven feedback exists (5/7)

4.1.2. Changing mindsets

Although the benefits and value of using data-driven approaches are being acknowledged by increasingly more people, including our interviewees, many others (e.g. colleagues of the interviewees) are rather skeptical and have yet to be convinced. Precisely this is very decisive as the product owner of *Subproduct A_1* states: "in order to enable data-driven approaches, the entire community would have to support and stand behind the idea and then assign a certain priority to this topic". In addition to that, the product manager of *Product A* clarifies that "all areas of the ecosystem would be affected" by the implementation of data-driven concepts.

However, according to the product owner of *Subproduct A_1* it "is difficult to gain acceptance of the new ways of working". The demand manager of *Product B* notes that "people need to be motivated to want to use it". Additionally, the product owner of *Product B* emphasizes that this will require a change of mindset for people to get used to it. Moreover, three of our interview partners stress that it is particularly challenging to convince management of the value of data-driven approaches because it usually requires additional resources to process and analyze data, build and maintain an infrastructure around the data, and prepare and visualize the results.

> **Findings IV.2**: Challenging to convince others of value (4/7); change of mindset required (3/7); difficult to convince management to provide resources (3/7)

4.1.3. Priorities & resources

According to six of the interviewees, the priority of introducing data-driven practices is still quite low partly due to the management's skepticism. The product owner of *Subproduct A_1* states that "at the moment this kind of topic does not get any priority due to the pressure of constantly developing new features". Related to this, the demand manager of *Product B* explains that it is a "challenge to integrate it in the timeline as the time-to-market expectations are very short".

As a consequence of the low priority assigned to data-driven approaches, the interviewees feel that the available resources are not sufficient to drive the topic forward. Without resources to develop some initial solutions, it is difficult to demonstrate and prove the value. However, only by clearly conveying the value and benefits of data-driven practices, the priority of these topics will be increased.

> **Findings IV.3**: Low priority due to pressure of developing new features (5/7); Low priority due to management's scepticism (3/7); not enough resources provided to drive topic forward (3/7)

4.2. Group II — Building AI4SABI solutions

In order to investigate an additional perspective of why data-driven techniques like AI4SABI are only rarely successfully realized in practice, the second group of our interview study consists

of roles responsible for implementing and building these types of solutions, specifically two software architects, one product owner (for operations), one data scientist, and one operations engineer. They are all experienced in developing and providing AI4SABI solutions to software engineering stakeholders, either by specifying requirements and analysis goals, collecting, processing and analyzing data, or building infrastructure to automate these processes. Though the interview guides are slightly tailored to the respective roles, each interview consists of the following elements: (1) generic processes and approaches when building or implementing AI4SABI solutions; (2) organizational- and people-related challenges; and (3) technical challenges.

4.2.1. Data model, processing & quality

In order to obtain valuable insights for stakeholders, it is often unavoidable to combine data from different sources. (e.g. usage data, log data, sales data, performance data, bug reports etc.). Therefore, a number of different stakeholders who provide the data must be involved. The data scientist of *Product A* explains that this often "leads to different views on the data which can cause a lot of confusion". For that reason, according to the operations engineer of *Product A*, it is very important to define a unified data model that is applicable to all data sources in order to allow an easy mapping. As the amount of data is usually quite large (100 GB per day in *Product B*), the software architect of *Product B* is responsible for decompressing the data, removing data that is not needed and store it in a usable format.

Furthermore, the operations engineer of *Product A* notes that due to the variety of data sources, the data cleaning process is very multifaceted, and hence "becomes very complicated very fast". When automating the data collection it is crucial "to check and if necessary enforce consistency". Analogously, the software architect of *Product B* notes that it is critical to "make sure pipeline elements that are doing the preprocessing are always up and triggered at appropriate times, make sure the transformation tasks are working properly, [...] and having retry mechanisms in place in case they are failing".

An additional challenge mentioned by the software architect working as a consultant for multiple products, including *Product B*, is an insufficient quality of data provided by stakeholders. He explains that "due to low quality of data, the results [of the analysis] are sometimes not good enough". As a result, he needs to "convey to the team that the quality of data needs to improve" even if that requires additional effort on the stakeholders sides'. Otherwise the analysis will no longer be taken up and will consequently get stuck in a prototypical stage.

> **Findings IV.4**: Challenge of combining multiple sources into one data model (4/5); ensure automated pipelines running correctly (3/5); ensure sufficient quality of data (3/5)

4.2.2. Mindset & cultural gap

Even after results are generated and presented to stakeholders, the acceptance of and believe in results is quite low. According to the product owner for operations of *Product A* the "results of analyses sometimes do not correspond with their beliefs" and they "would rather decide themselves what's good and bad".

Both, the software architect of *Product B* and the software architect working as a consultant, describe it as a problem of mindset because on the one side "people are still skeptical and don't see the benefit yet", and one the other side "to improve the quality of data [required for a sufficient quality and usefulness of results], process changes in their day-to-day work are inevitable".

The operations engineer of *Product A* believes that there is some kind of psychological effect because the stakeholders often "feel like an analysis generates additional effort and expense,

so it hurts more than it helps". In addition to that, the product owner of *Product A* states that stakeholders often feel "pressure to explain oneself instead of perceiving it as something useful". Due to this, stakeholders often have a negative attitude towards analyzing software engineering data and "feel like data is being used against them", according to the data scientist of *Product A*.

Another big challenge is the interpretability and understandability of analysis results. The data scientist of *Product A* states that stakeholders often do not have any "motivation to engage with something if they don't understand the technical side". For this reason, it is very "important to take the time to explain the analysis and results". However, in order to convey the technical knowledge necessary to convince stakeholders of the value of an analysis, some form of translation between roles is required. The data scientist brings up the example that even AI experts will not be able to communicate the required knowledge if they cannot express it in a way that is understandable to non-data-science stakeholders. Furthermore, it is important to consider that people are different. For instance, some people are very structured while others are more creative. As a result, different kinds of characters may need different kinds of support to take up AI4SABI. This should be taken into account when explaining an analysis and its results.

The operations engineer of *Product A* reports that even stakeholders who recognize the value in analyzing their data, find it difficult to specify their information needs and "to come up with any ideas themselves, they simply lack the time to think about it". Furthermore, the software architect of *Product B* notes that "people who are showing some interest, have to prove the value of the analysis to the 'higher' stakeholders who are making the decisions on financing such analyses". Consequently, the AI4SABI users do not only need to understand it, they also need to be able to defend or even explain it successfully to their higher stakeholders to succeed. This can be very challenging as the software architect working as a consultant explains: "some managers are very skeptical and it can be difficult to convince them that it's useful or worth investing time or people".

> **Findings IV.5**: Impression by providers that acceptance and belief in usefulness by users of AI4SABI is still low (5/5); Users' perception that it hurts more than it helps (3/5); explanations and translation very important to increase understandability and interpretability (3/5); difficulties in specifying use cases for analysis (2/5); challenging to convince others of value (2/5)

4.2.3. Priority, time & resources

The aforementioned attitude towards AI4SABI and other data-driven approaches has a direct impact on the priority, time and resources invested in driving these topics forward. The software architect working as a consultant explains that stakeholders are often "so much involved in the day-to-day activities that they don't have time to evaluate or try out something innovative". The data scientist of *Product A* notes that "it's becoming more present in managers' minds but they lack of understanding how to deal with the results".

Moreover, the software architect of *Product B* states that "the 'higher' stakeholders often have different priorities, stabilizing the product for example". For that reason, the software architect working as a consultant tries to keep the effort on the stakeholders' side as low as possible, although this is often difficult to achieve when it comes to data access and improvement of data quality. Oftentimes, analyzing data for one stakeholder requires certain actions from other stakeholders who do not directly benefit from the results and are, therefore, reluctant to spend their time on such topics.

Ultimately, it is very important to get "commitment from management" in order to assign a certain priority and the required resources to the topic.

Table 3
Overview of sample size, respondents and experience levels.

	Sample size		No. of respondents		
Total	1596		230		
Management roles	657		138		
Technical roles	939		121		
Experience levels in current role (in years)	<1	>1–3	>3–5	>5–10	>10–15	>15
% of respondents	4%	32%	19%	19%	10%	16%

Table 4
Tasks performed by the participants.

Business analysis	40%
Concept development	73%
Feature definition	67%
Staff and budget management	19%
Communication with customers	57%
Development/Implementation	61%
Infrastructure setup	26%
Database set up and/or maintenance	11%
Operations	19%
Data analysis	30%

Findings IV.6: Low priority assigned by management (3/5); stakeholders too occupied with their daily work (3/5); additional workload for stakeholders (3/5);

5. Survey results

The survey was sent to 1596 individual stakeholders working in software engineering, 230 (14%) of whom completed the questionnaire. Participants in the survey include both technical (60%) and management roles (53%), whereas each person can also be assigned to both roles. Table 3 presents a summary of the sample size, the respondents and their level of experience. In addition to that, the different types of tasks carried out by the respondents are listed in Table 4.

5.1. Current status in projects

When asked about the current status of collecting and analyzing (usage and/or runtime) data in the respondents' projects, 45% report that they are already collecting data, whereas 19% state that data is being collected but not yet analyzed. In the near future, an additional 19% plan to collect and analyze their data. The most common type of data being collected is log data (59%), followed by usage data (54%) and incident and bug reports (53%). An overview of the types of data being collected is given in Fig. 2.

Oftentimes, off-the-shelf tools are used to analyze and visualize collected data in the participant's projects. In particular, 52% of the respondents use tools like Microsoft Power BI, Tableau, Azure Log Analytics or Application Insights to extract some first information and visualize their data. Out of the 52% who are using these tools, 54% are either satisfied (52%) or very satisfied (2%) with the preparation and presentation of data and analysis results by the tools. Another 31% are neither satisfied nor dissatisfied, while 12% of users are dissatisfied and 3% very dissatisfied.

In addition to that, we asked the participants who are using off-the-shelf tools how well they are able to interpret the analysis results provided by the tools. Only one third of the respondents claims that they are able to interpret the results very (32%) or extremely (4%) well, while 64% of the respondents state that they are only moderately (49%), slightly (11%) or not at all (4%) able to interpret the results.

Table 5
Using results for decision making, productizing analysis and setting up an infrastructure.

Using results for decision making				
Never	Seldom	Sometimes	Often	Almost always
4%	23%	36%	34%	3%
Difficulty of transforming a prototypical analysis into a productized (automated & deployed) analysis?				
Not at all	Slightly	Moderately	Very	Extremely
2%	5%	48%	37%	8%
Effort of setting up an infrastructure for AI4SABI				
Very low	Low	Average	High	Very high
1%	8%	39%	37%	14%

Furthermore, the respondents are asked about types of information or insights they wish for but are not able to retrieve from the tools they are currently using. The majority of comments are related to additional data sources that respondents would like to include in analyses and visualizations. However, a significant number of participants also mention the application and use of AI techniques such as pattern detection and prediction models.

Findings V.1: The majority of respondents is open towards using AI4SABI, many already apply data-driven approaches; many use off-the-shelf analytics tools, though some struggle to interpret the results; the type of additional customized information respondents ask for underpins the usefulness of AI4SABI

5.2. Challenges and experiences in AI4SABI

In the third part of the questionnaire, respondents are asked to share their experiences with AI4SABI and the challenges they are facing during implementation or usage. In 47% of cases the respondents themselves and in 37% of cases colleagues or team members of the participants are applying basic analytic techniques (e.g. simple statistics) to manually analyze data. Only 18% of respondents report that they are successfully applying advanced analytics (e.g. complex statistics, machine learning etc.) in their projects.

Furthermore, the participants are asked to rate the frequency of using these analyses for decision-making in practice. A total of 37% of participants respond that they either often (34%) or almost always (3%) use such analyses for decision making, while 36% use it sometimes and 27% either seldom (23%) or never (4%).

In addition to that, we ask the respondents how difficult they experience the transformation of a prototypical analysis into a productized (automated & deployed) analysis. Only 7% assess it as not at all (2%) or slightly (5%) difficult, while 48% experience it as moderately difficult and 45% as either very (37%) or extremely (8%) difficult.

Based on this, the participants are asked how they experience the effort of setting up an infrastructure for continuous data analysis. 14% of the participant rate the effort as very high and 37% as high. 39% have experienced an average amount of effort and only 9% claim the effort to be either low (8%) or very low (1%). A structured overview of the results is presented in Table 5.

In order to identify and understand the factors that negatively influence a successful application and usage of AI4SABI, we extracted a list of factors from existing literature and the interviews presented in Section 4. As part of the survey, the respondents are asked to rank how much these factors negatively impact the application and usage of AI4SABI in their projects on a 5-point
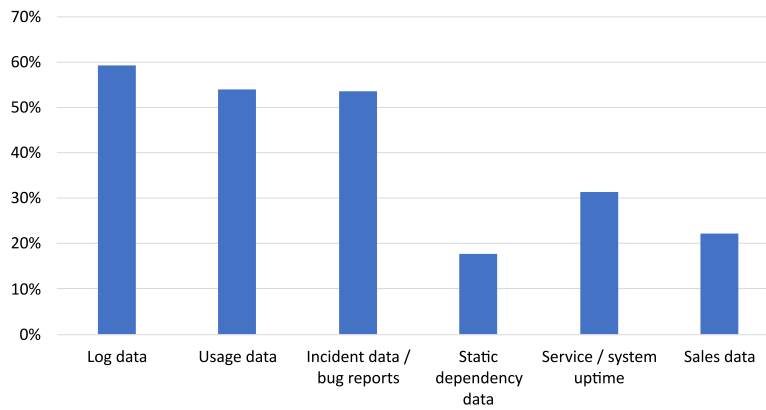
Fig. 2. Types of collected data sources.

Table 6
Rating of adverse factors on the application of data analytics in the respondents' projects; the bold numbers indicate which factors were ranked as either very (at least 30%) or extremely (at least 10%) influential by the respondents.

	Not at all	Slightly	Moderately	Very	Extremely
Lack of time and priorities	3%	9%	20%	**53%**	**14%**
Data privacy and security concerns	12%	21%	27%	28%	**11%**
Lack of knowledge of data sources within projects	12%	23%	34%	23%	8%
Lack of expertise in building infrastructure for data analytics	6%	21%	30%	**32%**	**11%**
Lack of expertise in data engineering (i.a. data collection, data processing)	9%	19%	30%	**36%**	6%
Difficulty of accessing, integrating and processing data from different sources	4%	18%	36%	28%	**14%**
Lack of expertise in data analytics (i.a. selecting and applying algorithms)	8%	20%	30%	**34%**	7%
Lack of data quality	8%	18%	33%	27%	**14%**
Effort of transforming a prototypical analysis to an automated and continuous analysis	4%	6%	31%	**48%**	**11%**
Difficulties in interpreting results	9%	25%	49%	17%	2%
Lack of trust in results (either due to or independent of interpretability)	16%	36%	28%	17%	3%
Lack of belief in usefulness of results	26%	24%	33%	15%	2%

Likert scale. There is always the possibility to skip and not rate a factor if participants consider that factor to be out of their area of expertise. Table 6 gives an overview of the factors and their rating.

A total of 67% of the survey's participants perceive the lack of time and priorities as an either very (53%) or extremely (14%) important factor that impedes the implementation and usage of AI4SABI in practice. Furthermore, a lack of expertise in data analytics (41%), data engineering (42%) and in building infrastructure for data analytics (43%) is considered either very or extremely challenging by the respondents. The difficulty of accessing, integrating and processing data from different sources is also rated either very or extremely high by 42% of the participants. The same applies to the effort of transforming a prototypical analysis into an automated and continuous analysis (59%). The plot in Fig. 3 presents an additional illustration of the results. The negative impact of the individual factors is visualized from light (not at all) to dark (extremely). The larger the dark fraction in the bars is, the greater is also the negative influence of the respective factor.

To conclude the survey, we offer a free text field for the respondents to share further experiences and challenges that impede a successful application and use of AI4SABI. Some participants consider it particularly difficult to change the mindset of others regarding AI4SABI and data-driven practices in general. One respondent claims that "the main problem is the people not being convinced that using data makes sense". Another one states "the concrete mindset is often missing. Only a training course is not sufficient". Related to this, one respondent feels like there is a "lack of management support, which is partly because it is difficult to provide 'use cases' that a manager will find valuable for the organization". For that reason, it is "difficult to continuously get the budget and staffing". From the opposite viewpoint a participant explains: "I am limited by cost savings at present, so it requires a step by step approach". Even "if time is spent on analysis, oftentimes there's considerably less (sometimes: no) time to act on its result".

Moreover, the specification of use cases and the collaboration of different roles to achieve this can be very challenging. One respondent states that "data analysts answer questions, but [...] do not proactively provide ideas". Another one mentions that there is a "cultural difference between domain experts and data scientists".
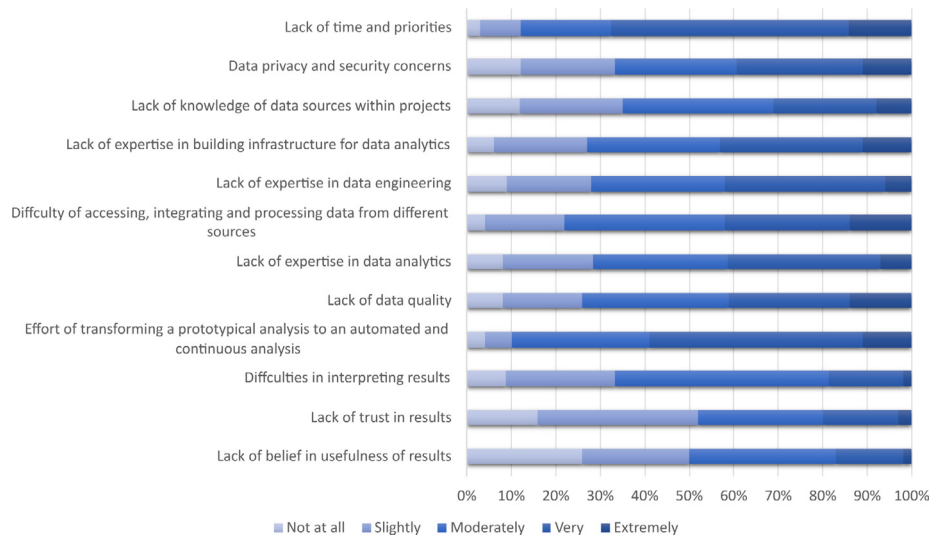
Fig. 3. Rating of adverse factors on the application of data analytics in the respondents' projects.

From a technical perspective respondents claim that it is "difficult to get access to data from all the individual databases and sources" and ensuring data quality at the same time. Another participant adds that "prototypes are cheap, [but] the next step takes time and effort" because "models need validation, maintenance and automation" which consequently requires "setting up the right infrastructure" for it.

> **Findings V.2**: Results are sometimes used for decision making; the effort to set up an infrastructure is perceived as high; the transformation from prototypes to continuous & automated analysis is difficult; it is difficult to access, integrate and process data from different sources; Critical adverse factors: lack of time and priorities, lack of expertise in data engineering, data analytics & building infrastructure, and lack of data quality

6. Identification of challenges and key drivers

This section aims at addressing RQ1: *What are the key drivers for analyses not being productized after the prototypical stage?* To achieve this, we analyze and compare the results of the interview study and the survey. We were able to identify multiple interdependent challenges that often prevent AI4SABI and other data-driven approaches from being successfully applied and used in practice. We observed that several of these challenges act like drivers that impact and partially amplify each other, effectively producing a vicious circle. Table 7 gives an overview of the identified challenges and key drivers, and indicates the source (interview study or survey) of each challenge. While there are additional challenges and factors potentially influencing a successful application of AI4SABI (e.g. data privacy and security concerns), we only selected interrelated challenges that are contributing to the key drivers of the vicious circle. A description of these drivers is provided in the following subsections before the vicious circle is explained and presented.

6.1. Lack of priority, time and resources

In order to successfully implement and use AI4SABI, a certain priority and consequently time and resources need to be assigned to the topic. The majority of participants in our study report that the priority assigned by the management is insufficient for driving these topics. Therefore, not enough resources are provided to build new or improve existing AI4SABI solutions.

6.2. Low data quality

On multiple occasions, interviewees and respondents in our study mention the lack of data quality in their projects. In addition to that, the data sources in the context of AI4SABI are often distributed and spread across multiple teams. In order to extract valuable insights, the data sources need to be combined. This often leads to the challenge of storing and integrating data from various sources in a consistent and accessible way. As a consequence, the data are often prone to be of lower quality. In addition to that, the lack of priority, time and resources impedes the improvement of data quality.

6.3. Inability to cross cultural gap

Based on the available data, possible use cases and topics of interest for an analysis need to be identified and discussed. In this context, we encountered a conflict during the evaluation of results. On the one hand, providers of AI4SABI claim that it is difficult to obtain requirements or specifications from the stakeholders who use the analysis in the future. On the other side, AI4SABI users feel overwhelmed by this task and, for that reason, expect the data analysts to provide inspiration and concrete suggestions. As a result, both stakeholders feel frustrated and not well understood by the other. To solve this, more time would need to be dedicated to more intensive exchanges and discussions between both parties. Data analysts need to better understand the stakeholders' domain-specific needs, while stakeholders need to acquire a technical understanding of the applied analyses and their results (e.g. by participating in trainings). However, as mentioned by one of our interviewees "single introduction sessions will not be sufficient", but a willingness to engage in a long-term collaboration will be required.

6.4. Ineffective prototypical analysis

Eventually, AI4SABI providers start with a prototypical implementation of a given use case. Mostly these initial implementations are based on data snapshots to produce some first results. However, in order to actually make use of the results, an analysis needs to be based on high-quality data and it should be executed in a continuous and automated way. The findings of our case study indicate that the results are often not good enough due to a low quality of data. Additionally, it is complex, takes a lot

Table 7
Key drivers and challenges derived from interview study (I) and survey (S).

Key drivers	Challenges	I	S
Lack of priority, time, and resources	Low priority due to management's skepticism	×	×
	Low priority due to new feature development and short time-to-market expectations	×	
	Insufficient resources (partially due to low priority)	×	×
	Challenging to convince management to invest time or people	×	×
Low data quality	Difficulty of accessing, integrating and processing data from various sources	×	×
	Insufficient data quality (improving quality requires additional effort on the stakeholders' side)	×	×
	Complexity of automating data collection and preprocessing	×	
	Lack of expertise in data engineering		×
Inability to cross cultural gap	Change of mindset is required — difficult to establish mindset	×	×
	Difficult to gain acceptance and motivate people to want to work in a data-driven way	×	
	Skepticism towards analyses/analyses not perceived as useful (generates additional effort, "pressure to explain themselves", "data is being used against them")	×	×
	Translation between roles is required/single trainings not sufficient	×	×
	Different expectations of specifying information needs	×	×
Ineffective prototypical analysis	Results not good enough due to low data quality	×	
	Lack of understanding how to deal with results/not time to act on results	×	×
	Lack of expertise in data analytics		×
	Analyses need to be automated/executed continuously to provide value → Complexity, high effort, and lack of expertise in model validation, maintenance and automation, and setting up the required infrastructure		×
Inability to prove value	Challenging to convince management (additional resources are required, difficult to identify valuable use cases for them)	×	×
	Acceptance of and belief in results low if results do not correspond with stakeholders' believes	×	
	Challenging to enable stakeholders to interpret and understand analysis results (explanations need to be adapted to stakeholders' domain and language)	×	
	Challenging for stakeholders to prove value to "higher" stakeholders who are making decisions on financing and resources	×	

of effort, and requires expertise to automate analyses. This leads to two challenges: On the one side, users of AI4SABI may not be able to recognize the value of an analysis and its results. On the other side, additional resources need to be acquired to realize the productization of the analysis.

6.5. Inability to prove value

Finally, before required resources can be allocated, the value of AI4SABI or other data-driven practices needs to be proven to relevant stakeholders. That includes not only the users of AI4SABI as primary stakeholders but also indirect stakeholders such as managers or other affected colleagues. This turns out to be a difficult task because, especially in the short-term, the advantages and benefits of AI4SABI are not directly perceived by customers. As a result, the pressure of implementing new features is usually higher than analyzing existing ones. Therefore, the former will in many cases be prioritized by the management.

In order to prove the value, oftentimes a change of mindset is required since people need to get used to a new way of working. It is important to establish trust and support stakeholders in understanding and interpreting the results. Otherwise, they are going to lose interest.

6.6. Vicious circle

In summary, the low priority assigned to the topic results in two consequences: On the one side, an inadequate amount of time is spent on dealing with data quality. On the other side, the discussions between users and providers of AI4SABI are not sufficient to overcome communication problems, create a mutual understanding and define relevant use cases. This cultural gap, along with poor quality of data, leads to a prototypical implementations whose results cannot deliver the desired value. In turn, if the value is not clearly visible, it is challenging to convince others of that value. So the vicious circle (see Fig. 4) is complete as this again will result in not increasing the priority of the topic.

7. Identification of potential solutions

This section aims at addressing RQ2: *How can the identified challenges and key drivers be addressed?* Based on they key drivers presented in the vicious circle, we review a list of key publications as well as the conducted interviews and survey for potential solutions that can support practitioners in taking a targeted approach at breaking the circle.

7.1. Solutions proposed in literature

One of the key challenges we identified in our study was a *low quality of data* in the interviewees' and survey respondents' projects. The heterogeneity of data and tools makes it difficult to reuse work, including cleaned or preprocessed data, of others. To address this, a large-scale survey among data scientists revealed that centralizing data and the corresponding definition as well as agreeing on and using only a subset of available tools can decrease the engineering effort across teams and lead to a better reusability of components (Kim et al., 2017).

Furthermore, data scientists at Microsoft use qualitative channels for the validation of quantitative data to ensure the meaningfulness of measurements (Kim et al., 2016). This can, for instance, be done by surveying a subset of people who are responsible for the data (e.g. users or customers, if feasible) (Kim et al., 2016) or by involving experts or stakeholders who are familiar with the data (Kim et al., 2017; Passi and Jackson, 2018). A triangulation of multiple data sources increases the confidence in results (Kim et al., 2016).

To overcome the *inability to cross the cultural gap* between data scientists and other stakeholders, a close and continuous interaction and collaboration is crucial (Kim et al., 2016). Moreover, stakeholder questions should be defined early on and refined in an iterative manner and data scientists should support the stakeholders in interpreting the data as well as the results (Kim et al., 2016). On the other side, data scientists should also consider
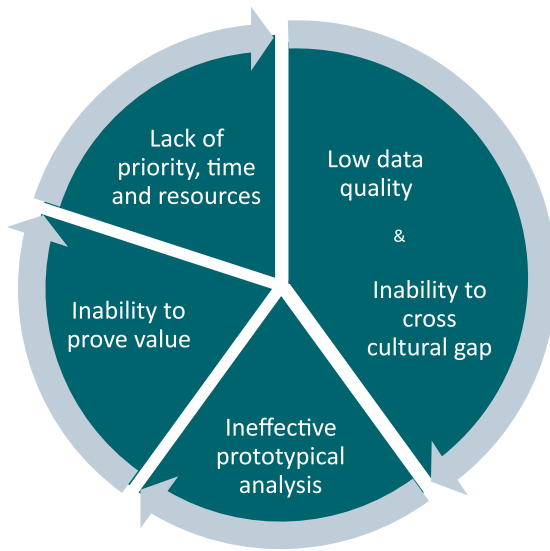
Fig. 4. Vicious circle consisting of key drivers that prevent a successful realization of AI4SABI.

and try to understand the business perspective since each analysis should be based on a set of predefined goals and business decisions that are to be supported by the analysis (Kim et al., 2017; Provost and Fawcett, 2013). By indicating how the business perspective is represented in the data and the model, an analysis can become more intuitive to stakeholders and the perceived usefulness of an analysis increases (Passi and Jackson, 2018).

Oftentimes, one-time analyses on snapshots of data are *ineffective* to show the desired value of an analysis. Therefore, it is important to "go the last mile" (Kim et al., 2016) and operationalize the analyses. In a case study at Microsoft, one of the interviewees explains that he or she spends 50% of the time on an analysis and the same amount of time on the integration into the product (Kim et al., 2016).

In order to clearly underline the *value* and benefit of an analysis, it is important to define and discuss concrete actions that can be taken based on the results (Kim et al., 2016). Moreover, explanations of analytics results should be kept as simple as possible to increase the understandability of the results and to show what stakeholders can achieve by using it (Kim et al., 2016; Passi and Jackson, 2018). After explaining an analysis, stakeholders should feel like they can use it by themselves without the need of constantly having a data scientist by their side (Kim et al., 2016).

To achieve this, data scientists should not use data science measures to explain the results, but rather translate their findings into the stakeholder's domain (e.g. "how much money can be saved?" or "how many customer calls can be prevented?") (Kim et al., 2016). It is important to consider that "stakeholders' trust in data science systems stems not only from model results and performance metrics, but also from some explanation or confidence in a model's inner working" (Passi and Jackson, 2018). Therefore, additional, stakeholder-targeted explanations minimize the risk of stakeholders misinterpreting a data science measure (e.g. accuracy score, precision, recall etc.) and build trust in results by explaining and visualizing results in a way that non-data-scientists can understand (e.g. showing probabilities of certain events instead) (Passi and Jackson, 2018). Thereby, explanations should be continuously evaluated and improved in collaboration with the stakeholders (Dam et al., 2018; Passi and Jackson, 2018).

Lastly, we were not able to find any solutions that address the last challenge in the vicious circle (*lack of priority, time and resources*) as the costs and financing of AI-based software analytics

or business intelligence solutions are being neglected in existing literature.

7.2. Solutions proposed in interviews

Related to the challenge of *low data quality* the operations engineer of *Product A* explains that they are using a "unified data model based on days to keep the effort to a minimum". Since all tables have the same granularity, they are easily extensible and connectible which can save a lot of time later on. In addition to that, the software architect working as a consultant for *Product B* explains that it is important to "identify which data sources are really important and to prioritize the data according to the stakeholders' needs"

In order to overcome the *inability to cross the cultural gap* between data scientist and other stakeholders, the data scientist of *Product A* highlights that while introduction sessions to machine learning are very important to help in understanding how the analyses work and how results can be interpreted, "single introduction sessions will not be sufficient". Therefore, shorter and more importantly iterative and continuous collaboration sessions are recommended. This is also stressed by the product owner and operations engineer of *Product A*. In order to provide meaningful AI4SABI solutions it is crucial to continuously work with the stakeholders, collect their feedback, and identify the problems they are facing. Otherwise, the stakeholders will stop trusting the analysis and will no longer want to use the AI4SABI solution.

The *inability to prove value* can be overcome by conducting sessions with the stakeholders to present and explain the results. Specifically, the software architect working as a consultant for *Product B* highlights the importance of explaining what kind of input data is used and what type of results or insights can be derived from the analysis. In addition to that, the software architect of *Product B* explains that they think about the presentation and visualization of data and information in the very beginning and even already define the UI elements as the visualization of results often supports stakeholders in comprehending what the results say and how they can use them. Moreover, the data scientist of *Product A* stresses that "trainings should not be on a high level but on a level that non-data-scientists can understand". Adding to that, the product owner for operations of *Product A* proposes to conduct role-specific trainings because he feels like "only few product managers understand machine learning" and he also emphasizes that "people need to understand the benefit and value of such analyses, otherwise even the coolest stuff is not worth anything".

For the remaining challenges *ineffective prototypical analysis* and *lack of priority time and resources* we were not able to extract any solutions out of the interviews.

7.3. Solutions proposed in survey

In regard to the *low quality of data*, one of the participants in the survey mentions that they are "enriching the current data source with more relevant data to get useful results". If the quality of data of one data source is insufficient, it can be helpful to include additional data sources.

Furthermore, the challenge of *ineffective prototypical analysis* is addressed by two of the survey's participants. One of the participants states that "AI is very powerful, but it seems it just gets applied to any problem case in order to solve it". Therefore, the respondent proposes to only apply AI to use cases where it really makes sense and to prioritize the use cases accordingly. Another survey participant notes that: "Currently, if time is spent on analysis, oftentimes there's considerably less (sometimes: no) time to act on its result. If one acts on an instinct rather than

Table 8
Identified solutions for key driver: *Low data quality*.

Solution	Category
Centralized storage of data & definition Unified data model (e.g. based on days) for same granularity of all tables → easily extensible and connectible Agree on a subset of tools for data storage & processing	Data storage & model definition
Triangulation of multiple data sources Enriching data with additional relevant data sources Use of qualitative channels for validation, e.g. involve stakeholders	Data validation

Table 9
Identified solutions for key driver: *Inability to cross cultural gap*.

Solution	Category
Close continuous collaboration with stakeholders (e.g. weekly data meetups)	
Iterative definition and refinement of questions	
Regular feedback sessions with stakeholders to support interpretation of data and results and identify problems faced by stakeholders	Collaboration
Introduction sessions are important but a single session will not be enough → rather multiple, short sessions on a regular basis instead of 1-day workshop	
Clearly specify and define goals and business decisions show how this is represented in the data	

Table 10
Identified solutions for key driver: *Ineffective prototypical analysis*.

Solution	Category
Plan fewer or less time-consuming analyses to have time for operationalization Allow time for both analysis and acting on its results Only apply AI to use cases where it really makes sense & prioritize use cases	Planning

profound analysis, there's more time to act and a better chance to actually have any result". Therefore, the respondent states that a crucial change is required to "allow time for both analysis and to act on its result". Moreover, the same participants also highlights the importance of prioritizing AI4SABI projects before starting and finalizing started projects.

Related to the *lack of priority, time, and resources* one of the survey's respondents explains that he or she is "trying to push this inside the team but due to time constraints nothing is taken into the backlog". Therefore, they are hiring students who can implement some first use cases in a cost-effective way. In addition to that, two further participants note that it is helpful to clearly show and discuss the tradeoff between the cost of the analysis and the value or business impact that it can provide. Knowing and understanding this tradeoff can make it easier for managers or decision makers to weigh up the decision on whether to allocate more time and resources for this topic.

We were not able to extract any solutions out of the survey that could help in overcoming the *inability to cross cultural gap* and the *inability to prove value*.

7.4. Categorization and validation of solutions

In order to structure and validate our findings, we assign the identified solutions to the key drivers in the vicious circle, then group the solutions into categories and finally discuss the categorized solutions with a group of experts who also rate the solutions' usefulness on a scale from one to five.

7.4.1. Solution categorization
Low data quality. For the first key driver we were able to identify six solutions that can either directly (e.g. through validation) or indirectly (e.g. via more effective usage or processing of data) help to improve data quality (see Table 8).

The indirect solutions include a centralized storage of data and the corresponding definition, a unified data model for the same granularity of all tables as well as the agreement on a subset of available tools for data storage and processing. These solutions are not directly related to the quality of data but they can indirectly contribute to it by making the access, usage, and processing of data more effective which leads to more time that can be spent on dealing with data quality. Since the solutions all cover aspects that need to be considered at the beginning of collecting data, they are assigned to the category *data storage & model definition*.

The more direct methods to improve data quality are mostly related to *data validation*. This includes a triangulation of multiple data sources, an enrichment of data with additional relevant data sources, and the use of qualitative channels for validation.

Inability to cross cultural gap. The solutions for this key driver are mainly related to the *collaboration* between the providers of AI4SABI solutions and stakeholders who are going to use it (see Table 9).

In order to cross the identified cultural gap, a continuous collaboration with stakeholders is crucial. This also includes an iterative definition and refinement of questions and regular feedback sessions to identify problems and support the stakeholders in understanding an analysis. Moreover, a sufficient amount of time should be taken to introduce stakeholders to the topic of AI4SABI as single introduction sessions will not be enough. On the other side, providers of AI4SABI need to become familiar with the goals and business decisions that are to be supported by their analyses. Showing the stakeholders how the data and models are linked to the business perspective will help them in building up trust in an analysis.

Ineffective prototypical analysis. In order to overcome the ineffectiveness of prototypical analyses, practitioners highlight the importance of planning and taking time apart from implementing and running an analysis even if that leads to fewer or more simple analyses (which are less time-consuming). Taking the time for both operationalization and acting on the results, is going to make analyses more effective by being more actionable and, therefore, also valuable for stakeholders. Moreover, the usefulness of applying AI to an use case should be carefully evaluated prior to implementation. Only use cases where it really makes sense to apply AI should be taken up and these use cases should be prioritized accordingly. The proposed solutions are also presented in Table 10 and can all be assigned to the category *planning*.

Inability to prove value. The value of an analysis can be clearly conveyed to stakeholders by providing *explanations* and *translations* of an analysis' result (see Table 11).

On the one side, regular sessions should be conducted to present and explain the results while the explanations should be kept simple so stakeholders feel like they can use it by themselves. It is important to discuss and explain what kind of actions can be taken based on the results and to think about how to present and visualize the results early on as part of the implementation.

On the other side, results should be translated to the stakeholder's domain. So instead of using data science measures other

Table 11
Identified solutions for key driver: *Inability to prove value.*

Solution	Category
Conduct sessions with stakeholders to present and explain results & collect feedback on explanations	Explanation
Keep explanations simple so stakeholders feel like they can use it by themselves	
Discuss, define and explain what actions can be taken based on results	
Think about how to present and visualize the results in the beginning	
Don't use data science measures, but translate results to the stakeholder's domain (e.g. what is the business value?)	Translation
Training should not be high level but on a level that non-data-scientist can understand, maybe even role-specific	

Table 12
Identified solutions for key driver: *Lack of priority, time and resources.*

Solution	Category
Hire students to support the implementation of use cases Clearly show cost/value tradeoff	Costs

domain-specific indicators should be communicated (e.g. "what is the business value?" or "how much time can be saved?"). This also applies to trainings that are conducted with stakeholders. These should not be on a high level but on a level that non-data-scientists can understand, maybe even role-specific.

Lack of priority, time and resources. Due to the pressure of their every day work, the management's priority of spending time and resources on analyses that do not directly go into the product is often quite low, especially if the value of that analysis is not clearly visible. In order to keep the costs in the beginning to a minimum, students can be hired as a cost-effective way to implement some first use cases. Moreover, it can help to clearly show and discuss the tradeoff between the costs that will occur and the value that an analysis could provide. Both solutions are assigned to the category *costs* (see Table 12).

7.4.2. Solution validation

In order to validate the impact and usefulness of the solutions, we conduct expert interviews and ask the participants to discuss and rate the usefulness of solutions on a scale from one to five. Four of the five experts also participated in the initial interview study. The fifth, who was not available then, was added due to his expertise in the area. An overview of the solution categories, their sources as well as their expert rating is provided in Fig. 5. The solutions are grouped by their respective categories which are assigned to the key drivers of the vicious circle. The boxes next to the categories indicate the sources of the solutions (literature, interviews, and/or survey) and the average rating of the experts who participated in our study.

Data storage & model definition. The first solution category comprises solutions proposed in the literature and the interviews. Especially the unified data model is perceived as very useful to make the data more "transparent and easier to understand". Independently of each other, all experts rated the usefulness of the solution as 4 (on the scale from 1 to 5).

Data validation. The solutions of the category *data validation* were derived from the literature as well as the survey and are discussed a bit more critically by the experts. The operations engineer of *Product A* explains that the triangulation of multiple data sources and the enrichment of additional data sources mostly increases the "confidence for the users, but not for the providers".

The software architects of *Product B* agree that it is important to involve the stakeholders and take their domain knowledge into account. Overall, two experts of *Product A* gave this category a rating of 2.0, while the remaining experts rated it a 3.0 resulting in an average rating of 2.7/5.

Collaboration. The third solution category *collaboration* consists of solutions extracted from the literature and interviews. It is perceived as very useful in theory but not really feasible in practice. The data scientist of *Product A* explains that "it's a good idea but there is simply no time. Who is going to realize this?". The software architect of *Product B* states that "it depends on where the push is coming from". The communication within a team is usually much easier as compared to across-teams. Depending on the team setup of users and providers of AI4SABI, a close collaboration can become difficult. Moreover, the stakeholders' willingness to invest time in such a close and time-consuming collaboration depends on the stakeholders' personal interests, priorities, and motivation. Due to the difficulty of realizing this in practice, four out of six interviewees rate this solution category a 2.0 and two other interviewees a 3.0 and 4.0 respectively. This results in an overall expert rating of 2.5/5.

Planning. The solutions of the *planning* category are based on all the considered sources (literature, interviews, and survey) and are well received by the group of experts. The operations manager of *Product A* confirms that "it is better to have something simple than something very complex that is not complete". The software architect of *Product B* also agrees that this is "very relevant". In total, two out of six interviewees rate this solution category a 5.0 and the other two a 4.0. Ultimately, this results in an overall expert rating of 4.3/5.

Explanation. The solution category *explanation* comprises solutions derived from the literature as well as the interviews. The group of expert agrees that this is indeed a very important topic to consider. The data scientist of *Product A* states that "it is always good when stakeholders see some first visualizations early on". He adds that the information or numbers that are being reported "must be easy to understand". The software architects of *Product B* also consider the solutions of this category very useful. One of them explains that it also depends on the stakeholder. Some stakeholders really want to understand why a ML-model predicts something in a certain why while others only care about whether the results are useful or not and what the next actions are. Therefore, the explanations need to be adapted to the stakeholders' personalities and needs. As a result, four out of six interviewees rate this category a 5.0 and two a 4.0 resulting in an overall rating of 4.7/5.

Translation. Very similar to the previous category, the solutions of the category *translation* were derived from the literature and the interviews and are perceived as very valuable by the group of experts. The data scientist of *Product A* confirms that these solutions are "very important". The software architects of *Product B* also state that they fully agree with the solutions and have "nothing to add". Two of the interviewees give a rating of 5.0 and the remaining four interviewees rate this solution a 4.0. Consequently, this results in an overall rating of 4.3/5.

Costs. For the last solution category *costs* we were only able to identify solutions from one of the sources, specifically the survey. Regarding the proposal of hiring students to support the implementation of use cases, the operations engineer of *Product A* confirms that they are also doing that and that "it is good for a quick start to get something done at a lower cost". The software architect of *Product B* adds that while hiring students is cost-effective, it sometimes comes at the cost of quality. As a result, five out of six interviewees give a rating of 4.0 and one rates it a 3.5. Finally, this results in an overall rating of 3.9/5.
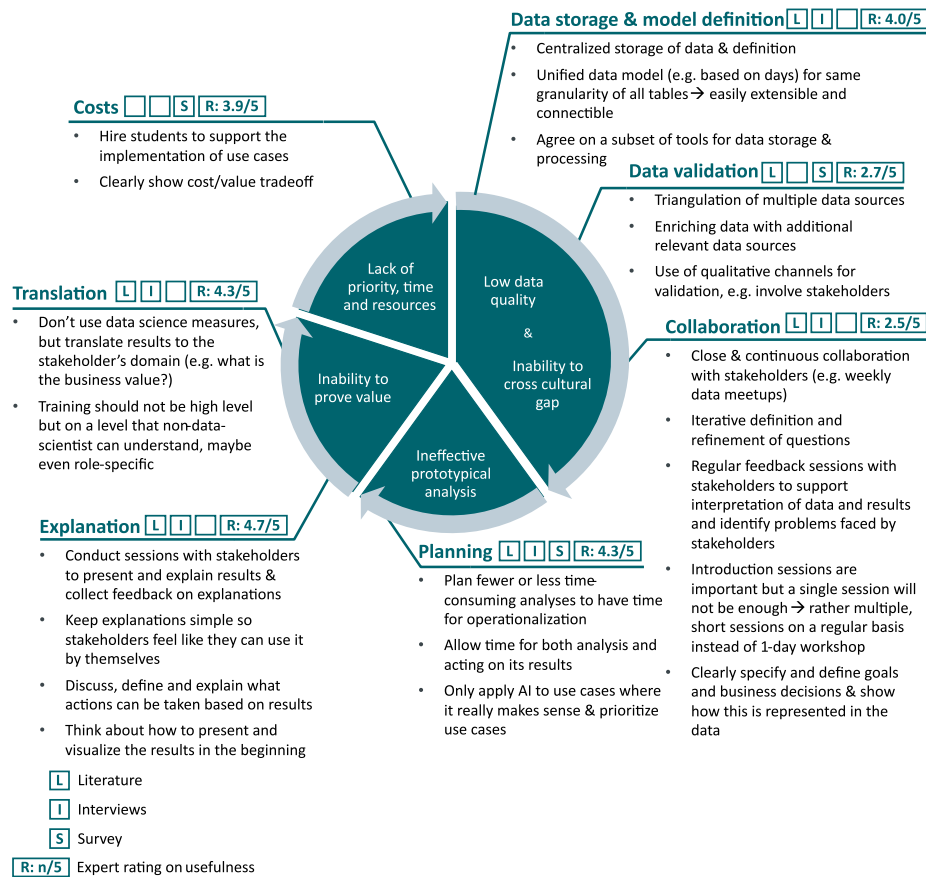
Fig. 5. Identified solutions for the key drivers in the vicious circle.

8. Threats to validity

8.1. Internal validity

One threat to the internal validity of our study is the limited number of products and the unequal distribution of interviewees among product teams. We compensated this by including the consultants who both work for *Product B* but also for a number of other products.

Moreover, it is likely that the respondents of our survey participated due to their interest in the topic of AI4SABI, thereby potentially causing a positive bias in the results. Due to the limitations of the survey tool, dependencies between questions and factors could not be taken into consideration.

Due to the lack of successfully deployed AI4SABI solutions, the group of experts interviewed for validating the solutions consists mostly of technical experts (providers) as compared to non-technical stakeholders (users). Therefore, we cannot rule out a slightly biased distortion of the results due to the over-represented providers of AI4SABI.

8.2. External validity

The limited number of interviewees in the interview study might have an impact on the generalizability of the results which constitutes an external threat to validity. We aimed at compensating this by using a mixed-methods approach and additionally conducting a survey to answer the same research questions. However, the same applies to the expert interviews conducted to validate the potential solutions.

Furthermore, the transferability of results to other contexts might be affected by a positive bias in the survey results.

While the overarching case study was conducted in multiple unrelated and independent product teams, they all originate within a similar industrial context which might impact the external validity of the study. However, with the many roles, organizational units and products represented in the study and by multi-source data collection as the basis for our results, we believe that the findings we present provide value also outside the specific context of the case study company.

9. Conclusion

While the use of artificial intelligence for software analytics and business intelligence can in theory be a valuable asset to retrieve and extract meaningful information for decision-making, it is very difficult to realize it in practice, specifically in a continuous and automated manner.

In our study we have identified five interdependent key drivers that impede a meaningful utilization of AI4SABI: A lack of priority, time and resources results in a low quality of data and the inability to cross the cultural gap between data scientists and other stakeholders. This leads to an inefficient prototypical analysis and the inability to prove its value, which again prevents an increase of the priority and consequently also time and resources. As a result, these key drivers form a vicious circle that is difficult to break.

To remedy that, we have extracted and categorized a set of solutions out of existing literature, the interviews and the survey to support practitioners in taking a targeted approach at breaking the circle. The solutions are grouped into seven categories: Data storage & model definition, data validation, collaboration, planning, explanation, translation, and costs.

Indeed, the selection of solutions may depend on project-specific characteristics. In fact, the mentality and priorities of the specific stakeholders involved could act as a lever to select the solutions that are most promising. For instance, if a technical stakeholder such as a data engineer aims at driving the topic forward, he or she will could start by improving the data quality and data scientists could additionally try to work on or improve explanations provided to other stakeholders. On the other side, potential users of AI4SABI (e.g. product managers) could proactively promote the collaboration with data scientists and maybe even try to include colleagues that are still skeptical.

Currently, the validation of our study is limited to the experience and perception of the interviewed group of experts. In future work, we intend to conduct a long-term evaluation of applying the proposed solutions in practice.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

Arpteg, A., Brinne, B., Crnkovic-Friis, L., Bosch, J., 2018. Software engineering challenges of deep learning. In: 2018 44th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, pp. 50–59.

Buse, R.P., Zimmermann, T., 2012. Information needs for software development analytics. In: 2012 34th International Conference on Software Engineering (ICSE). IEEE, pp. 987–996.

Butz, Jr., H.E., Goodstein, L.D., 1996. Measuring customer value: gaining the strategic advantage. Organ. Dyn. 24 (3), 63–77.

Chen, H., Chiang, R.H., Storey, V.C., 2012. Business intelligence and analytics: From big data to big impact. MIS Q. 1165–1188.

Ciolkowski, M., Laitenberger, O., Vegas, S., Biffl, S., 2003. Practical experiences in the design and conduct of surveys in empirical software engineering. In: Empirical Methods and Studies in Software Engineering. Springer, pp. 104–128.

Cito, J., 2016. Developer targeted analytics: supporting software development decisions with runtime information. In: 2016 31st IEEE/ACM International Conference on Automated Software Engineering (ASE). IEEE, pp. 892–895.

Copeland, J.B., 2020. Artificial intelligence. Encyclopedia Britannica, Retrieved May 9th 2021 from https://www.britannica.com/technology/artificial-intelligence.

Dam, H.K., Tran, T., Ghose, A., 2018. Explainable software analytics. In: Proceedings of the 40th International Conference on Software Engineering: New Ideas and Emerging Results, pp. 53–56.

Easterbrook, S., Singer, J., Storey, M.-A., Damian, D., 2008. Selecting empirical methods for software engineering research. In: Guide to Advanced Empirical Software Engineering. Springer, pp. 285–311.

Fabijan, A., Olsson, H.H., Bosch, J., 2016. The lack of sharing of customer data in large software organizations: challenges and implications. In: International Conference on Agile Software Development. Springer, pp. 39–52.

Figalist, I., Elsner, C., Bosch, J., Olsson, H.H., 2019a. Business as unusual: a model for continuous real-time business insights based on low level metrics. In: 45th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE, pp. 66–73.

Figalist, I., Elsner, C., Bosch, J., Olsson, H.H., 2019b. Customer churn prediction in B2B contexts. In: International Conference on Software Business. Springer, pp. 378–386.

Figalist, I., Elsner, C., Bosch, J., Olsson, H.H., 2020. Breaking the vicious circle: Why AI for software analytics and business intelligence does not take off in practice. In: 46th Euromicro Conference on Software Engineering and Advanced Applications (SEAA). IEEE.

Figalist, I., Elsner, C., Bosch, J., Olsson, H.H., 2021. Fast and curious: A model for building efficient monitoring-and decision-making frameworks based on quantitative data. Inf. Softw. Technol. 132, 106458.

Hu, H., Wen, Y., Chua, T.-S., Li, X., 2014. Toward scalable systems for big data analytics: A technology tutorial. IEEE Access 2, 652–687.

Jordan, M.I., Mitchell, T.M., 2015. Machine learning: Trends, perspectives, and prospects. Science 349 (6245), 255–260.

Kaner, C., Bond, W.P., 2004. Software engineering metrics: What do they measure and how do we know? In: METRICS 2004. IEEE CS. Citeseer.

Kim, M., Zimmermann, T., DeLine, R., Begel, A., 2016. The emerging role of data scientists on software development teams. In: 2016 IEEE/ACM 38th International Conference on Software Engineering (ICSE). IEEE, pp. 96–107.

Kim, M., Zimmermann, T., DeLine, R., Begel, A., 2017. Data scientists in software teams: State of the art and challenges. IEEE Trans. Softw. Eng. 44 (11), 1024–1038.

Kitchenham, B.A., Pfleeger, S.L., 2001-2003. Principles of survey research: parts 1-6. ACM SIGSOFT Softw. Eng. Notes.

2021. Machine Learning. Oxford University Press, Retrieved May 13th 2021 from https://www.oxfordlearnersdictionaries.com/definition/english/machine-learning.

Maguire, M., Delahunt, B., 2017. Doing a thematic analysis: A practical, step-by-step guide for learning and teaching scholars.. All Ireland J. Higher Educ. 9 (3).

Marciuska, S., Gencel, C., Abrahamsson, P., 2014. Feature usage as a value indicator for decision making. In: 2014 23rd Australian Software Engineering Conference. IEEE, pp. 124–131.

Menzies, T., Zimmermann, T., 2013. Software analytics: so what? IEEE Softw. 30 (4), 31–37.

Menzies, T., Zimmermann, T., 2018. Software analytics: What's next? IEEE Software 35 (5), 64–70.

Misirli, A.T., Caglayan, B., Bener, A., Turhan, B., 2013. A retrospective study of software analytics projects: In-depth interviews with practitioners. IEEE Softw. 30 (5), 54–61.

Mohri, M., Rostamizadeh, A., Talwalkar, A., 2018. Foundations of Machine Learning. MIT Press.

Negash, S., Gray, P., 2008. Business intelligence. In: Handbook on Decision Support Systems 2. Springer, pp. 175–193.

Olsson, H.H., Bosch, J., 2013. Towards data-driven product development: A multiple case study on post-deployment data usage in software-intensive embedded systems. In: Lean Enterprise Software and Systems. Springer, pp. 152–164.

Olsson, H.H., Bosch, J., 2015. Towards continuous customer validation: A conceptual model for combining qualitative customer feedback with quantitative customer observation. In: International Conference of Software Business. Springer, pp. 154–166.

Olsson, H.H., Bosch, J., 2019. Data driven development: Challenges in on-line, embedded and on-premise software. In: International Conference on Product-Focused Software Process Improvement. Springer, pp. 515–527.

Passi, S., Jackson, S.J., 2018. Trust in data science: collaboration, translation, and accountability in corporate data science projects. Proc. ACM Human-Comput. Interact. 2 (CSCW), 1–28.

Provost, F., Fawcett, T., 2013. Data science and its relationship to big data and data-driven decision making. Big Data 1 (1), 51–59.

Runeson, P., Höst, M., 2009. Guidelines for conducting and reporting case study research in software engineering. Empir. Softw. Eng. 14 (2), 131.

Saggi, M.K., Jain, S., 2018. A survey towards an integration of big data analytics to big insights for value-creation. Inf. Process. Manage. 54 (5), 758–790.

Sculley, D., Holt, G., Golovin, D., Davydov, E., Phillips, T., Ebner, D., Chaudhary, V., Young, M., Crespo, J.-F., Dennison, D., 2015. Hidden technical debt in machine learning systems. In: Advances in Neural Information Processing Systems. pp. 2503–2511.

Seaman, C.B., 1999. Qualitative methods in empirical studies of software engineering. IEEE Trans. Softw. Eng. 25 (4), 557–572.

Thierauf, R.J., 2001. Effective Business Intelligence Systems. Greenwood Publishing Group.