

Clausen, Johan Bjerre Bach; Møller, Charles; Asmussen, Claus Boye; Jørgensen, Steffen Lundgaard

## Article

# Self-service business intelligence systems in production domains: Introducing data democratization and mass customizable mashups

Digital Business

## Provided in Cooperation with:

International Telecommunications Society (ITS)

*Suggested Citation:* Clausen, Johan Bjerre Bach; Møller, Charles; Asmussen, Claus Boye; Jørgensen, Steffen Lundgaard (2025) : Self-service business intelligence systems in production domains: Introducing data democratization and mass customizable mashups, Digital Business, ISSN 2666-9544, Elsevier, Amsterdam, Vol. 5, Iss. 2, pp. 1-19, <https://doi.org/10.1016/j.digbus.2025.100156>

This Version is available at:

<https://hdl.handle.net/10419/338027>

### Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

### Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by-nc-nd/4.0/>



# Self-service business intelligence systems in production domains: Introducing data democratization and mass customizable mashups<sup>☆</sup>

Johan Bjerre Bach Clausen<sup>a</sup>, Charles Møller<sup>b</sup>, Claus Boye Asmussen<sup>c</sup>,  
Steffen Lundgaard Jørgensen<sup>c</sup>

<sup>a</sup> Department of Business Development and Technology, Aarhus University, Birkcenter park 15, 7400, Herning, Denmark

<sup>b</sup> Department of Mechanical and Production Engineering, Aarhus University, Katrinebjergvej 89, 8200, Aarhus N, Denmark

<sup>c</sup> ArLA Foods, Sønderhøj 14, 8260, Viby J, Denmark

## ARTICLE INFO

### Keywords:

Business intelligence  
Internet-of-Things  
Self-service business intelligence  
Data democratization  
Digital twin  
Mass customization

## ABSTRACT

This paper explores the observation that making business intelligence available through self-service business intelligence (SSBI) systems has only to a lesser degree been successful in organizations. In this paper, we specifically explore SSBI in the production domain, where Internet-of-Things and production data are core data sources in production companies. This paper takes a design science approach, where a structured literature review is used to establish the current research frontier. From our analysis of the problem situation, and close empirical collaboration with a company experiencing problems with implementing SSBI systems, we propose and design two digital artifacts. The two artifacts are partly evaluated based on empirical data from the case company. In this paper, we find that data should not only be made available in SSBI systems, but must also be presented in accordance with domain knowledge of the users. This finding implies that the SSBI literature should to a larger degree draw upon insights from relevant IS literature on contextualized data. Practitioners should design IS systems with contextualized data at their core, as the digital twin discussed in this paper. The second main finding of the paper is that analytics in SSBI systems should be prescriptive analytics, which directly supports the users' decision-making. A novel contribution of this paper is integrating the concept of prescriptive analytics SSBI systems, and in this paper we discuss and partly validate how integrating prescriptive analytics and SSBI systems can improve decision-making in a production context.

## 1. Introduction

The quest for competitive advantage in the market is an ongoing challenge for organizations seeking to survive and thrive. In a production and supply chain context, that manifests in the desire of organizations to continually improve production and supply chain decision-making and efficiency. Decision support systems are a research area within information systems that study how information systems and data can be used to improve decision-making in organizations (Arnott & Pervan, 2008; Eom & Kim, 2006; Li, Dai, Issakhov, Almojil, & Souri, 2021). With ever-increasing availability of both structured and unstructured big data from diverse sources such as internet-of-Things (IoT) devices, webpages, social media, transactional data, etc., and the broader adoption of advanced analytics (Bose, 2009), new concepts and knowledge have been added to the area of decision support systems. Business Intelligence (BI) has emerged as a common name for the area of research and practice exploring decision-making

based on big data and advanced analytics (Cebotarean & Titu, 2011; Phillips-Wren, Daly, & Burstein, 2021). Using BI to improve decision-making in different organizational domains is currently of great interest to organizations (Analytics, 2015; Srivastava, Venkataraman, V, & N, 2022) but making BI systems broadly available in organizations has proven less successful (Analytics, 2015; Eckerson, 2012; Lennerholt, Van Laere, & Söderström, 2018, 2021).

BI's lack of success in decision support is discussed in Phillips-Wren et al. (2021) by explicitly linking BI and decision support systems. Here, it is argued that BI should be viewed as a subfield within decision support systems, which have evolved into focusing on decision-making based on big data. Furthermore, the authors argue that a lack of awareness of the connection between BI and the greater information system (IS) literature means that BI research often does not exploit the insights that have been generated in the greater IS and decision support systems (DSS) literature. We agree that BI has lacked a deeper connection with

<sup>☆</sup> This article is part of a Special issue entitled: 'Unleashing the Power of IoT' published in Digital Business.

\* Corresponding author.

E-mail address: [joce@btech.au.dk](mailto:joce@btech.au.dk) (J.B.B. Clausen).

decision support, but one central emerging topic in BI is not discussed in the paper: Self-Service Business Intelligence (SSBI). SSBI has rapidly grown in organizational and research interest (Alpar & Schulz, 2016; Michalczyk, Nadj, Azarfar, Maedche, & Gröger, 2020). Moreover, it is stated in the SSBI literature that companies often face substantial unique challenges when implementing SSBI systems (Lennerholt et al., 2018, 2021), and currently, unsuccessful applications of SSBI systems are not uncommon (Alpar & Schulz, 2016; Analytics, 2015; Eckerson, 2012; Lennerholt et al., 2018, 2021). Moreover, by putting the specific aims of SSBI systems into a perspective focused explicitly on decision support, we believe we can contribute new and valuable insights to the IS literature.

In this paper, we take a design approach (Hevner, March, Park, & Ram, 2004; Rai, 2017) to investigate the reasons behind the lack of success SSBI systems. More specifically, our methodological foundation is drawn from intervention-based research (IBR) (Chandrasekaran, de Treville, & Browning, 2020; Oliva, 2019). By drawing on insights from the wider IS literature and insights from our empirical problem situation, we design digital artifacts and the interventions in which they are employed. The goal of our IBR study is to arrive at principles for how researchers and practitioners can design and implement successful SSBI systems. Since we are seeking to link research in SSBI systems to other academic research areas, a structured literature review is required to identify if other researchers have explored the theoretical linkages. In this paper, we present the results of this literature review, and how we can use insights from the wider IS literature to create two digital artifacts aimed at improving SSBI systems in production contexts. We will also present feedback received from users and collaborators in the case company, and present the results of the initial empirical evaluation of one of the artifacts. In the IBR methodology, it is expected that researchers iterate between the artifact and empirical problem situation in order to ensure that an improved problem situation is actually achieved. While that iteration is ongoing, we believe that the design of the artifacts based on structured and unstructured literature reviews, and initial evidence for the effectiveness of the artifacts have already generated insights for researchers and practitioners.

The empirical problem situation is from a large multinational company in the food production field, which has identified decision-making based on big data and advanced analytics as a key source of competitive advantage. The case company has also experienced a historical lack of success with SSBI systems. The company generates and stores a vast amount of data from its business domains, and it has created hundreds of reports to aggregate and communicate the data, but the majority of these reports are rarely used and maintained. Moreover, the case company has identified that the reports are of limited value in enabling decision-makers across the company to make data driven decisions. At the start of the research project, the company had identified that it generates the data that domain experts need for decision-making, but it cannot deliver the data to these domain experts. Moreover, the case company has identified the need for advanced analytics to improve decision-making in business domains such as its production processes. In this paper, we limit the scope of our research to the case company's production process, since the company has an extensive integrated production process that starts with raw food materials and ends with making processed food products ready for distribution. Moreover, the efficiency of the production processes is central to the company's overall profitability. To improve the efficiency of its production processes, the company has already adopted SSBI systems across its production domains, which seek to link IoT data from the production process with business and product data, but the company has also experienced less than satisfactory results with SSBI systems in the production domain. The research is based on the IBR methodology, which means that the research is developed in close collaboration with a number of employees from the case company, whose positions in the company include management positions, IT and data engineering positions, domain experts, and BI positions. Moreover, during the research project,

we have continuously involved the stakeholders of the case company associated with the project, and we have returned to relevant elements of the problem situation as new insights were uncovered as prescribed by the IBR methodology (Chandrasekaran et al., 2020; Oliva, 2019).

Potential reasons for the lack of success of SSBI systems is discussed in the literature (Eckerson, 2008, 2012; Lennerholt et al., 2018). Lennerholt et al. (2018) performs a literature review of the SSBI literature in order to identify why SSBI systems lack success in organizations. Based on our review of the literature and our analysis of the case company's problem situation, we conclude that the main reasons for the lack of success of SSBI systems are: (1) They do not enable decision-makers to access the data they need for their decision-making in a form that is in accordance with their domain knowledge. (2) SSBI systems do not readily support the multitude of specific problems decision-makers face in their organizational domains, and further analytics is needed to arrive at optimal decisions. We believe that our work in an empirical context has already enhanced our understanding of why SSBI systems lack success in the production domains of organizations. In this paper, we therefore seek to contribute to the scientific literature on SSBI systems with a deeper understanding of the challenges facing SSBI systems and propose principles for improving the success of SSBI systems.

Inspired by Phillips-Wren et al. (2021) we draw upon recent concepts from the IS and DSS literature when designing interventions seeking to make SSBI systems more successful. However, these concepts must be augmented to fit into an SSBI system problem setting. This is in accordance with the methodological literature, which states that it is that design research applies concepts from the forefront of academic research (Hevner et al., 2004). The two digital artifacts we propose are (1) Data democratization (Awasthi & George, 2020; Zeng & Glaister, 2018) through the implementation of a digital twin of the company's production domains (Liu, Fang, Dong, & Xu, 2021). This artifact enables decision-makers in the production process to access the IoT and production data they need in a form that is in accordance with their domain knowledge. (2) The second digital artifact is mass customizable analytics mashups which seeks to use insights from the IS literature on mass customization (Tiihonen & Felfernig, 2017) to efficiently deliver personalized analytics trained on IoT and production data, denoted as mashups in the SSBI literature, to the decision-makers in the production process of the case company. We expand the theoretical base of the SSBI literature by introducing these artifacts, which are based on insights from the IS literature. Our empirical problem situation is production-oriented, but we believe that we arrive at insights that are relevant to the broader SSBI literature.

In the next section (Section 2) we will introduce and discuss the concepts of BI, SSBI and the challenges facing the success of SSBI systems. In Section 2 we also use the introduced concepts to introduce and contextualize relevant elements of the empirical problem situation. Section 3 contains our methodical considerations and presents the literature reviews we have performed to support our research. Section 4 contains an introduction and discussion of the first proposed digital artifact regarding data democratization. Section 5 contains an introduction and discussion of the second proposed digital artifact: Mass customizable mashups. We end the paper with a conclusion and future research topics in Section 6.

## 2. Business intelligence, self-service business intelligence and their current challenges

In this section, we will briefly introduce key concepts from the BI and SSBI literature to establish the challenges facing SSBI systems. The key concepts we introduce will be put into the context of the empirical problem situation of the case company, and we will also discuss how the generic challenges facing SSBI systems are present in the case company. An early definition of BI systems is given by H. Davenport (2014) where BI systems are defined as "Tools to support data-driven

decisions, with emphasis on reporting” and an updated definition is presented in [Arnott, Gao, Lizama, Meredith, and Song \(2019\)](#) where BI systems are defined as “BI systems are large-scale systems that combine information technologies, data reporting, and analytic processes in order to support decision-making in an organization.” The core goal of BI systems is supporting decision-making, and [Chen, Chiang, and Storey \(2012\)](#) describe how the use of BI in organizations has evolved through 3 stages. At each stage, BI systems expand the range of data sources they can draw on become capable of handling more complex analyses. The fact, that predictive and to a lesser extent prescriptive analytics have become a core part of BI systems is also seen in the definition given in [Arnott et al. \(2019\)](#). However, it is important to highlight that at the time of writing, descriptive reporting is still the central task of BI systems in many organizations. This is also observed in the case company, where SSBI systems adopted in the company are primarily used for reporting measures and KPI’s for different departments of the company. An observation that a senior data scientist and an information system infrastructure senior manager of the company also expressed (see the methodology section for an overview of the main individuals from the case company who contributed to the research).

In the literature, BI systems and their reports are often discussed as aiding strategic decision-making ([Alpar & Schulz, 2016](#); [Davenport, 2006](#); [Michalczyk et al., 2020](#)). Moreover, in the literature it is described how BI systems should be developed by teams of data analysts, data scientists, and IT developers with continuous feedback from business decision-makers ([Bara, Botha, Diaconita, Lungu, Velicanu et al., 2009](#); [Muntean & Surcel, 2013](#)). Data analysts and data scientists are also often required to assist decision-makers with interpreting the outputs of the BI systems ([Bara et al., 2009](#); [Muntean & Surcel, 2013](#)). BI systems have proven successful for strategic decision-making ([Alpar & Schulz, 2016](#); [Davenport, 2006](#); [Michalczyk et al., 2020](#)), but the resources of data scientist, data analysts, data engineers and IT developers are bottlenecks when organizations seek to expand their use of BI systems and especially when seeking to expand their data driven decision-making to tactical and operational decisions ([Alpar & Schulz, 2016](#); [Bani-Hani, Pareigis, Tona, & Carlsson, 2018](#)). The central challenge companies face when trying to expand their use of BI systems is that they often have a limited number of data scientists and data analysts to develop and use BI systems, and their resources can easily be overwhelmed by requests from decision-makers in the organization ([Yu, Lapouchnian, & Deng, 2013](#)). This situation is also observed in the case company, which has many decision-makers on the tactical and operational level with high subject matter knowledge of their domain in the production process, such as engineers monitoring and controlling the food production process, but who have limited capabilities to extract and analyze the data generated in the different production domains in the company. Data scientists and data analysts are a comparatively small group in the company, which is easily overwhelmed by requests for descriptive reports, predictive analyses, and prescriptive decision-making support.

### 2.1. Self-service business intelligence

The concept of SSBI has been developed to alleviate the continuous strain decision-makers put on the limited resources of data scientists and data analysts, and thereby overcoming the data and analytical bottleneck organizations face when trying to scale BI supported decision-making ([Alpar & Schulz, 2016](#); [Bani-Hani et al., 2018](#)). [Michalczyk et al. \(2020\)](#) defines SSBI in the following way: “[SSBI] can empower casual users to create and share reports on their own from high volumes of diverse data without getting stuck into long-running business requests with IT departments”. SSBI systems should give decision-makers the capabilities they need to make data driven decisions independently. In the definition, decision-makers are described as casual users. Casual users are one of three types of users defined in the SSBI literature, the other two being power users and IT users, and user categories are a

central concept in the SSBI literature ([Alpar & Schulz, 2016](#); [De Mauro, Greco, Grimaldi, & Ritala, 2018](#)). Casual users generate value for the organization in their current role and SSBI systems should enhance their ability to generate value ([Alpar & Schulz, 2016](#)). When advanced analytics ([Bose, 2009](#)) are needed to realize improved decision-making further education of casual users can be necessary. The concept of the citizen data scientist has been proposed to describe casual users with limited amounts of extra analytical education, which enable them to perform and interpret some forms of advanced analytics using SSBI systems ([Alpar & Schulz, 2022](#)). However, since casual users generate value for their organization in their current role, it is neither feasible nor prudent to seek to turn all casual users into citizen data scientists.

In the SSBI literature, casual users are those whom SSBI systems must empower to combine their domain knowledge with data and analytics when making decisions within their field ([Spahn, Kleb, Grimm, & Scheidl, 2008](#)). [Lennerholt et al. \(2018\)](#) states that the success of SSBI implementations is dependent on achieving self-reliant casual users. At the same time, if casual users do not stop interacting with the SSBI systems, they can gain BI knowledge and become more self-reliant over time ([Michalczyk et al., 2020](#)). According to [Alpar and Schulz \(2016\)](#) casual users primarily interact with SSBI systems in the form of dashboards, reports, and reports with drill anywhere functionality, which the authors define as the lowest of three levels of BI self-service (usage of information). The case company also employs a large number of reports, some of which have drill anywhere functionality with which casual users can interact. Reports are used in the company to gain insights into how their production process performs on a large amount of KPIs and how this performance change over time. From interviewing the senior data scientist and information system infrastructure senior manager, and also indicated by documents from the company, it is clear that the majority of reports are rarely used or have been created for a specific use case, which might not be relevant after some time, or the use case itself is rare, but it has been observed in the company that for tactical or operational decision-making, reports can often only be of limited use. That is because casual users in the production process of the company need data and analytics to aid in making specific decisions in response to changes in the production process. From talks with an IT portfolio manager in the case company, information from company documents, and examining production data, a number of problem areas in which casual users of the company seek to use data and analytics to support have been identified. The problem areas are, but not limited to, determining the most energy efficient settings of their production machines given the current state of the machines and the products that are to be produced. Determining the optimal time to clean or perform other maintenance on the machine given the state of the machine and what has previously been produced. Determining if the products that are being produced meet quality measures and the best actions to take if quality measures are not met. Revising the production plan given the current state of the production process. These decisions are already made by casual users using existing decision processes, with lesser and greater use of descriptive reports and established analyses. However, the case company sees great potential in providing casual users with SSBI systems, enabling them to use data and analytics to optimally respond to changes in the production process.

Power users are users with the capabilities to extract data and create analyses that are either used in their own decision-making or are disseminated to casual users ([Alpar & Schulz, 2016](#); [De Mauro et al., 2018](#)). Data scientists and data analysts are typically classified as power users ([De Mauro et al., 2018](#)), but the concept of the citizen data scientist has also been introduced ([Alpar & Schulz, 2016](#)) and is between casual and power users. While data scientists generally are more self-reliant than data analysts ([De Mauro et al., 2018](#)), we believe it is prudent to view the self-reliance of power users as being on a spectrum. The higher the capability to extract data from internal and external data sources, and the higher the capability to perform advanced analytics the power user has, the more self-reliant the power user is. In [Alpar](#)

and Schulz (2016) the medium and the highest level of SSBI are the domains of power users. The medium level of SSBI is “creation of information“, where power users can use SSBI systems such as Power BI, Alteryx, KNIME, Tableau, etc., to extract data made available to them and create reports and analyses. The highest level of SSBI is “creation of information resources“, where power users can harness new data sources and create analytical mashups (Alpar & Schulz, 2016). The central characteristics of mashups are that they are “personalized, context-rich, role-tailored” (Kobielus, Karel, Evelson, & Coit, 2009) and are therefore designed to meet the specific decision-making needs of users to a higher degree than reports. In Kobielus et al. (2009), mashups are discussed mostly as role-tailored data visualizations, but the authors envision mashups that merges domain knowledge of casual users and advanced analytics. In a more recent paper by Michalczyk et al. (2020) examples of mashups that integrate advanced analytics are presented.

In the case company, there are power users across the spectrum of self-reliance. Some work exclusively with commercially available SSBI systems such as Power BI and Alteryx, while other power users work in general programming languages, which offer greater flexibility. This is a pattern also observed in the literature (Chard et al., 2019; Schuff, Corral, St. Louis, & Schymik, 2018). As stated above, the case company already uses SSBI reports but is dissatisfied with the decision support they offer, particularly in operational and tactical decision-making. The company has recognized that they need role-tailored mashups for operational and tactical decision-making, which casual users can interact with through their SSBI systems. Mashups are not a common term used in the company, but the decision support needs identified in the company are in accordance with the mashup definition from the SSBI literature. This need for role-tailored analytics and visualizations that present directly actionable insights and prescribe actions was also noted in interviews with a senior production engineer, BI specialist, and senior BI specialist in the company. As described above, casual users in the production process are faced with a multitude of tactical and operational decisions, where they need decision support in determining specific actions in response to changes in the production process. This form of decision support is not being met by (drill anywhere) reports and we have identified that role-tailored and personalized mashups are needed in the production domains of the company. However, as the company has started developing role-tailored SSBI mashups, they find that they consume more of the company’s limited analytical resources and have longer development times than report construction. Later in this paper, we present our proposed concept of mass customizable mashups which seeks to significantly lower resource consumption and development times of SSBI mashups.

IT users are responsible for the organizational IT infrastructure and are therefore responsible for the IT systems from which SSBI systems draw their data (De Mauro et al., 2018). However, IT users and power users must be able to cooperate, since IT users typically control the investment and perform updates and maintenance of IT systems on which power users rely (De Mauro et al., 2018). The case company has a large IT infrastructure supporting food production sites across several countries. Moreover, a core source of growth is through the absorption of production sites from other companies, and the IT users of the company are therefore experienced with integrating separate IT infrastructures. However, it is challenging for the company to develop the IT infrastructure and capabilities to support SSBI systems and advanced analytics.

## 2.2. The challenges facing self-service business intelligence

Even the proponents of SSBI note that SSBI systems have only had limited success in achieving the aim of empowering casual users to become self-reliant and make independent data driven decision-making. The “State Of Self-Service Report” from LOGI Analytics from 2015 is an older but comprehensive report based on 600 respondents (Analytics, 2015). A more recent but equally comprehensive report would be a

boon for the research area, but in the published report, 22% of surveyed business users state that they have access to useful SSBI systems, and the overall satisfaction percentage is around 17%. Without being able to present specific numbers, a similar picture is painted in Lennerholt et al. (2018, 2021). In Eckerson (2012), 64% of the surveyed business intelligence professionals report average or lower success with SSBI systems. Eckerson (2008) explores a case where a company implements SSBI systems, but a lack of success makes the company reverse course away from self-service business intelligence. Moreover, organizations are in the short term interested in investing more in SSBI systems (Analytics, 2015; Lennerholt et al., 2018), which only highlights the need for more research into how organizations can successfully implement self-service business intelligence systems.

In Lennerholt et al. (2018) a literature review is performed in order to identify challenges facing SSBI systems. From our own literature search and by working with the case company, we agree with the findings of the authors, but we can expand their research by drawing on insights from our empirical problem situation. The authors identify two main challenges when implementing and using SSBI systems, and using insights from our work with the empirical problem situation, we can add to these findings into designing successful SSBI systems. We propose that successful SSBI systems should:

**1: Enable casual and power users to access the data they need in their decision-making and have the data presented in a form that is in accordance with their domain knowledge.**

A necessary condition for SSBI systems empowering casual and power users to make data driven decision-making is that casual and power users can use the SSBI systems to access the data they need in their decision-making (Lennerholt et al., 2018). This might seem an obvious insight, but what is observed both in the literature and in the case company is that data is still often siloed in different departments or otherwise not accessible for users across an organization (Alpar & Schulz, 2016; Lennerholt et al., 2018). However, making data accessible through IS systems, such as cloud storage, is in itself also not sufficient for successful SSBI systems. If querying data requires the knowledge of different querying languages, and if the data storage design is complex to query into, then all but the most self-reliant power users will not be able to meaningfully access data (Lennerholt et al., 2018). The same insight holds if the data is not contextualized in accordance with the domain knowledge of users. Contextualization means that data is transformed into a form that is usable and actionable for the users with their domain knowledge. An example thereof is given in Section 4 of the paper. The naming conventions of the data must also be in accordance with the domain knowledge of the users.

From this follows our first design hypothesis: *Successful SSBI systems give domain experts access to data from their domain in a form that is in accordance with their domain knowledge.*

The design hypothesis informs the design of our first digital artifact, which is based upon insights from the data democratization literature (Awasthi & George, 2020; Lefebvre, Legner, & Fadler, 2021; Zeng & Glaister, 2018). This part of the literature is chosen since it discusses how organizations can make data accessible and give the users the capabilities to access and use it. Specifically, we will be discussing the design of a digital twin Werner et al. (2021) of the production process, which is being implemented in the production domains of the case company. We will describe how casual and power users will interact with the digital twin through SSBI systems, which will give casual and power users access to the data they need for decision-making, and it will present the data in a form that is in accordance with their domain knowledge.

**2: Give casual and power users analytics that supports solving the specific decision-making problems they are facing in their domain.**

Data access through SSBI systems can give casual and power users new insights into their domain, but users need to be able to easily arrive at information that improves their decision-making (Lennerholt

et al., 2018). Even when data is available and visualized in (drill-down) reports, there are many domains of the production process where analytics is needed when aiding decision-making. As described previously, in the production process of the case company, casual or power users, need to determine optimal machine settings, maintenance actions, and production planning in response to changes in the production process. Therefore, users need personalized analytics (mashups) that easily and precisely support the decision-making in their domain. While data access through SSBI systems can reduce the strain on an organization's limited data and analytical resources, it is outside the capabilities of casual users in different production domains to create the personalized analytics they require to improve their decision-making. In the case company the demand for personalized analytics far exceeds the resources of power and IT users in the company, and we believe that the low satisfaction with SSBI systems reported in the literature (Lennerholt et al., 2018) shows that the company is not unique in this regard.

From this follows our second design hypothesis: *Successful SSBI systems enable organizations to give domain experts access to personalized and actionable analytics.*

The design hypothesis informs the design of our second digital artifact, which seeks to efficiently deliver personalized analytics to the users in different production domains. To achieve this, we draw inspiration from the IS literature on mass customization (Tiihonen & Felfernig, 2017) and we name our proposed design "mass customizable mashups". In the mass customization literature, it is discussed how to design IS systems that support flexible manufacturing with which users interact in order to efficiently get personalized products (Tiihonen & Felfernig, 2017). With mass customizable mashups the personalized analytics is created by an automated machine learning analytics system (Elshawi & Sakr, 2020; Yang & Shami, 2022; Zöllner & Huber, 2021) that draws data from the digital twin, and with which the users interact through a self-service configuration system created based on insights from the mass customization literature (Tiihonen & Felfernig, 2017). The aim is that mass customizable mashups will enable casual users and power users with lower self-reliance to receive the personalized analytics they require for their domain-specific decision-making. Moreover, we propose that mass customizable mashups can lessen the strain on analytical resources in the case company, since power users with high self-reliance and IT users spend their resources building and updating the automated machine learning system and the configuration systems.

### 3. Methodology

The methodology of this paper is based on design science (Hevner et al., 2004; Rai, 2017), and more specifically, intervention based research (IBR) within information systems (Chandrasekaran et al., 2020; Oliva, 2019). IBR takes its starting point in an empirical problem situation (S) which is informed by a theoretical base (T), and researchers take an active role in reaching an improved state (S\*) through an intervention (Chandrasekaran et al., 2020; Oliva, 2019). In the context of IS research, a digital artifact is typically at the core of the intervention (Hevner et al., 2004). The active role researchers take, is not a traditional research role, but a core methodological motivation of IBR is that through designing and evaluating interventions, researchers can gain a deeper understanding of complex problems (Chandrasekaran et al., 2020; Oliva, 2019).

In this research, the problem situation is the lack of success of SSBI systems (S), and we seek to design interventions that can make SSBI systems more mature (S\*). The core of our theoretical base (T) is the BI and SSBI literature, and since our research is within IS, we seek to develop digital artifacts to reach S\*. In IBR researchers work from S, informed by a theoretical base (T), and the goal of IBR research is not only reaching S\*, but also contributing new knowledge to the theoretical base T. An IBR research cycle starts with an initial S and an initial T that inform the design of an intervention to move to S\*, which can be evaluated in order to identify if S\* has been reached.

**Table 1**

Roles of the main people collaborated with in the company.

Role	Research relationship
Senior data scientist (power user)	Working relationship
Information system infrastructure senior manager	Working relationship
Director level information system manager	Working relationship
IT portfolio manager	Working relationship
Senior IoT infrastructure manager	Working relationship
Senior production engineer	Multiple interviews
Senior production manager (power user)	Multiple interviews
Senior IoT data engineer	Multiple interviews
Senior BI specialist (power user)	Multiple interviews
BI specialist (power user)	Multiple interviews
Analytics manager	Single interview
IT manager (data infrastructure)	Single interview

However, IBR is an abductive method where the researchers actively iterate and update their understanding of S, T, and the interventions as insights are gained. An IBR research cycle can therefore be a protracted process (Chandrasekaran et al., 2020; Oliva, 2019), and our research is a multi-year research project. However, we believe that we have gained valuable insights into the challenges facing SSBI systems and that we propose innovative interventions, which can further mature our understanding of SSBI systems, and by which other researchers and practitioners can be inspired.

The reasons for choosing a methodology based on IBR are two-fold. (1) SSBI papers that also identify the problem situation, such as (Lennerholt et al., 2018), have not explored an IBR approach. We believe that our research significantly enriches the understanding of the problem situation by working with an empirical problem situation from a design science perspective. (2) The development, implementation, and evaluation of digital artifacts is a core element of IBR in IS, and developing successful SSBI systems is of interest to both researchers and practitioners.

Our initial understanding of the problem situation (S) is informed by recent systematic literature reviews of the BI and SSBI areas, our own comprehensive literature search, and by studying the SSBI offerings available in the market. We have then added further depth and contextualization to our findings from the literature by drawing upon insights arrived at in our close working relationship with the case company. In Table 1, we present an overview of the case company employees with whom we have collaborated in order to get a deeper understanding of the problem situation.

In the following, we will introduce the roles of the main collaboration partners and data collection procedures of the research to ensure methodological clarity and rigor of the research. We will keep the exposition concise, since the research is the outcome of a multiyear research process where a significant amount of time was spent following the IBR methodology and iterating between the problem situation and theoretical base to ensure that a full understanding of the problem situation could be achieved, and such that the artifacts could plausibly enable us to arrive at an improved state. The degree of involvement and the nature of the involvement of the individuals in Table 1 have therefore varied over the course of the research, and a full account of their involvement would be unnecessarily long. Below, we therefore outline the core elements of the involvement of the people in Table 1.

Author one is the researcher primarily responsible for the academic work presented in this paper, and in the company collaboration, he was responsible for identifying the research gap and direction, and managing the research process. He was also responsible for designing and developing the algorithms for creating predictions and prescriptions as well as the IS artifact for mass customizable mashups. He also evaluated the tests of the statistical and optimization algorithms. The senior

data scientist, a co-author on the paper, has worked for or with the company for over 5 years across a multitude of projects. The senior data scientist contributed through bi-weekly meetings of multiple hours and ad hoc discussions at the case company, which generated information and insights on all topics discussed in this paper. In particular, he contributed with subject-matter expert knowledge in the engineering area of study, decision-making needs of the end-users, feedback on statistical analysis and algorithmic design, being an active liaison to other collaboration partners. He was also responsible for the creation of the case company's production version of developed analytics and visualizations. The information system infrastructure senior manager, also a co-author on the paper, was a core senior manager overseeing the design and development of the digital twin in the collaboration company and his insights are core to the discussion of that artifact presented in this paper. He also gave feedback on the development of the mass customizable mashups artifact to ensure the design fit with the digital twin data infrastructure.

The senior IoT data engineer worked on the development and implementation of the digital twin. The multiple interviews with him are a key data source for technical details about the digital twin and its user requirements. The senior IoT infrastructure manager joined the project while it was ongoing and left the project again after some months for reasons not related to the project. While he was connected to the project, he took part in monthly meetings assessing the user needs for the mass customizable mashups artifact and how it integrated into the collaboration company's IoT data infrastructure. The director level information systems manager was the formal product owner of the project at the collaboration company, and through (on average) bi-monthly meetings feedback was given on the project and its progression.

As will be expanded upon later, the collaboration company produces at multiple sites, and the senior production engineer and the senior production manager (power user) were employed at the site where the data used for developing and validating the statistical models were initially collected. The senior BI specialist and the BI specialist, both having relevant engineering backgrounds but employed in production-facing BI roles, were employed at the site where the data used for retraining the statistical models on new machines was collected. It was also in the production process of that site that the experiments of the automated machine learning system were performed. The above-mentioned people contributed with subject-matter expert knowledge of the production process and decision-making needs of users. Since they would also be users of the mass customizable mashups, they have also provided feedback on the artifacts. In addition, they facilitated contact and meetings with users with lower education levels. The IT portfolio manager was not formally involved in the project, but his broad knowledge of users, stakeholders, and other IT projects in the company enabled him, on a primarily ad hoc basis, to provide insights into the design of the artifacts and whether they met the decision-making needs of users. On average, he contributed once or twice on a monthly basis with ad hoc insights. The analytics manager and the IT Manager (data infrastructure) were interviewed in a semi-structured interview at the beginning of the project, and they are included in this list since the interviews were very informative for our initial understanding of the problem situation.

The empirical data and information presented and used in this paper are based on two primary sources. The first is interviews, primarily semi-structured conducted during meetings with the aforementioned people, potential users, and to a lesser extent other stakeholders. The number of questions prepared for each interviewee varied and also depended on the stage of the research process at which the interview was conducted. After the development of the statistical analysis and optimization algorithms for the mass customizable mashups were started, a large amount of the interviews revolved around verifying data, identifying data sources, getting feedback on intermediate statistical results, ensuring that the optimization problem corresponded to the

**Table 2**

Search terms for the structured literature reviews.

Subject	Query
Data democratization	“democratization*” OR “democratisation*” AND bi OR “business intelligence” OR “data analytics” OR analytics
Mass customization	“mass customization” OR “mass customisation” OR “mass-customization” OR “mass-customisation” AND bi OR “business intelligence” OR “data analytics” OR analytics

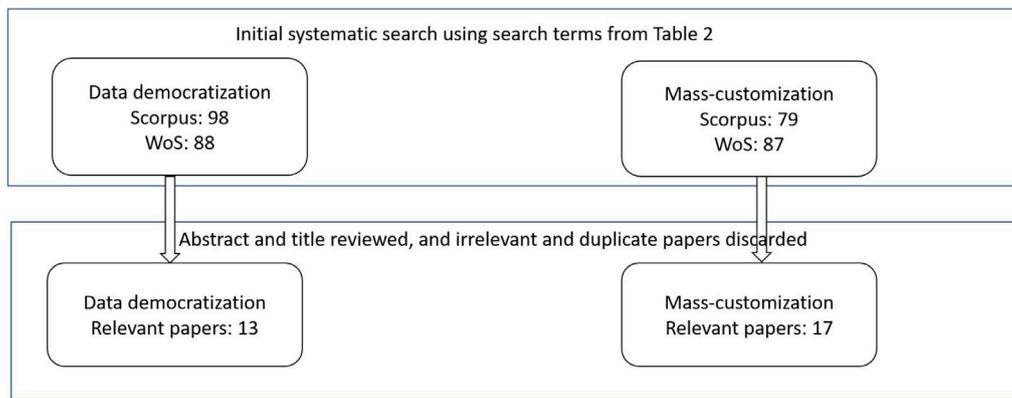
organizational problem, and getting feedback on proposed artifact designs or prototypes. Ad hoc unstructured interviews or correspondence also created a significant amount of data for this paper. Recordings of meetings were not possible, but research notes from the meetings were made and saved, and over 100 pages of ad hoc correspondence have also been saved. The second data source is internal documents from the case company, to which the researchers have been granted access to, and information and insights from these documents are presented anonymously, when relevant, in the paper. Since this research is based on the IBR methodology, where the researcher is an active participant, observations and insights gained through the active participation of the researchers in the development and evaluation of the artifacts are also a core part of the empirical foundation of this research paper.

#### 4. Structured literature reviews

We are not the only researchers who have identified the challenges that SSBI currently faces, and since other researchers might have proposed similar solutions, we therefore performed structured literature reviews. The structured literature reviews involved search terms capturing data and analytics democratization and mass customization. This is done in order to get a systematic overview of publications that potentially explore the specific concepts we discuss.

We have searched Scopus and Web-of-Science (WoS) literature databases using the search queries shown in Table 2. Since SSBI is still an emerging research area, and the concepts we explore are not typically related to the SSBI literature, we did not limit our search to that part of the literature. The results of our two systematic literature reviews are summarized in Fig. 1. In the systematic literature reviews, we evaluated titles and abstracts to identify relevant papers. For the papers deemed to be relevant, we evaluated the main text of the papers to identify insights into the concepts of data and analytics democratization and mass customization, and to determine the extent to which they discuss SSBI systems.

In Tables A.5 and A.6 in Appendix A, we report the findings of the two literature reviews for the reviewed papers. In our literature of data and analytics democratization, we find that in the SSBI literature, the concept is understood as giving people across an organization access to data and analytics, and it is claimed to be happening within the field of Business Intelligence (Chergui & Kechadi, 2022; Vuppapapati, Ilapakurti, Kedari, Vuppapapati, Vuppapapati et al., 2020). However, we have only been able to identify the papers by Alpar and Schulz (2022) and Harland, Hocken, Schröer, and Stich (2022) which substantially explore the concept of data and analytics democratization. The two papers explore how organizations can give their members access to data and analytics, and they both mention SSBI systems as part of the democratization process. In this paper, we seek to expand on this previous work by using further insights from the data democratization literature (Awasthi & George, 2020; Lefebvre et al., 2021; Zeng & Glaister, 2018). This literature discusses how data and analytics democratization must be understood as an organizational change process where data and analytics access, through SSBI systems, are only part of the democratization process, which must include data and analytics education, and the fostering of an organizational culture that value data and analytics driven decision-making (Awasthi & George, 2020;



**Fig. 1.** Overview of the structured literature reviews.  
Source: Created by the authors.

Lefebvre et al., 2021; Zeng & Glaister, 2018). Our conclusion based on the literature review is that the concept of data and analytics democratization is not completely foreign to the SSBI literature. However, current treatment of this concept in the literature is relatively superficial and the new insights from the literature on data and analytics democratization have yet to be fully integrated.

In the literature review of mass customization and its relation to analytics as reported in Table A.6 in Appendix A, the concept of mass customization is defined as the ability to produce personalized products with an efficiency approaching that of mass production of products (Coito et al., 2020; Dudek, Patalas-Maliszewska, & Kowalczywska, 2023), and it is stated to be a core part of Industry 4.0 (Coito et al., 2020; Saldivar, Goh, Chen, & Li, 2016; Zhong, Xu, Klotz, & Newman, 2017). This definition aligns with the general definition of mass customization (Gong, Wang, Jiao, & Gebraeel, 2023). Our review reveals that analytics supporting mass customization is a small research area (Saldivar et al., 2016; Tien, 2013), with little consideration given to how users should interact with these analytics. Moreover, we do not find any other papers discussing the use of insights from the mass customization literature for the efficient creation of personalized analytics. Our conclusion is therefore that extending insights from mass customization to the area of analytics is a research topic that has not yet been explored.

From the structured literature reviews, we found that the concept of data and analytics democratization has been introduced to the SSBI literature, but the concept of using insights from mass customization in the context of analytics has not been explored. The literature reviews also informed our understanding of the theoretical base, and relevant insights therefrom will be expanded upon and referenced in the description of the proposed artifacts. Moreover, based on our literature reviews, we find that the artifacts proposed in this paper are novel artifacts that have not been explored in the existing literature.

## 5. Data democratization through a digital twin intervention

The case company is a traditional production company that has embarked on what it views as a data and analytics democratization process. Specifically, it sees great potential in giving organizational members across its extensive production process access to data and analytics. As discussed above, the company's vision is to enable its members across different production domains to use data to get real-time or near-real-time information about the state of their domain of the production process and enable them to use analytics to act optimally when new information is available.

The case company is not the only company with this vision, and data and analytics democratization is a concept relatively recently introduced to the IS literature. The fundamental idea is that data has become a vital asset for organizations and like any asset it must

be made available to the members of an organization (Awasthi & George, 2020; Lefebvre et al., 2021; Zeng & Glaister, 2018). Data democratization is meant to reach members broadly in an organization, regardless of their technical skills (Zeng & Glaister, 2018). Using BI terminology, data democratization is meant to reach both casual and power users in an organization. The aim is to enable members of the organization to integrate the use of data into their organizational roles and improve their everyday work, while ensuring data compliance and security (Lefebvre et al., 2021). In Lefebvre et al. (2021) four catalysts of data democratization are identified as: (1) Greater access to data and tools for non-specialists, (2) development of data and analytics skills, (3) collaboration and knowledge sharing, and (4) promotion of data value. The authors also discuss that data democratization is an organizational change process that seeks to change how organizations understand data and its use in decision-making. In the data democratization literature, self-service and SSBI systems are viewed as part of the greater data democratization process, specifically the first catalyst described above, and the design of SSBI systems should fit into the larger data democratization process of an organization in order to achieve success (Lefebvre et al., 2021; Zeng & Glaister, 2018). The explicit link to the data democratization literature has been made by the researchers, but in interviews with the information system infrastructure senior manager and the senior IoT data engineer it was clear that they had identified that data availability is not the same as users being able to meaningfully access and gain value from data. These interviews also motivated the researchers to explore the academic literature on that topic. The lack of success of SSBI systems reported in the SSBI literature and also observed by the employees in the case company seems almost expected when drawing upon insights from the data democratization literature. According to the data democratization literature, SSBI systems themselves are not sufficient to enable casual users to perform data driven decision-making, instead, this requires significant changes in how an organization values and uses data. A perspective only discussed in Harland et al. (2022) and Alpar and Schulz (2022) in the SSBI literature.

In accordance with the insights from the data democratization literature, the case company is moving away from site-specific data silos to a cloud-based infrastructure, to which both IoT and production data from across the domains of the production process are extracted and stored, and where it can be accessed by members across the domains of the production process. The company is also expanding its use of self-service systems, such as PowerBI, and has already established an independent section that educates employees in the use of data and data driven decision-making. Moreover, the case company is developing a guide and vision for how data will be used in everyday work across the organization and how insights from data can be shared across the organization. The case company had already started its data democratization process before the initiation of our current research

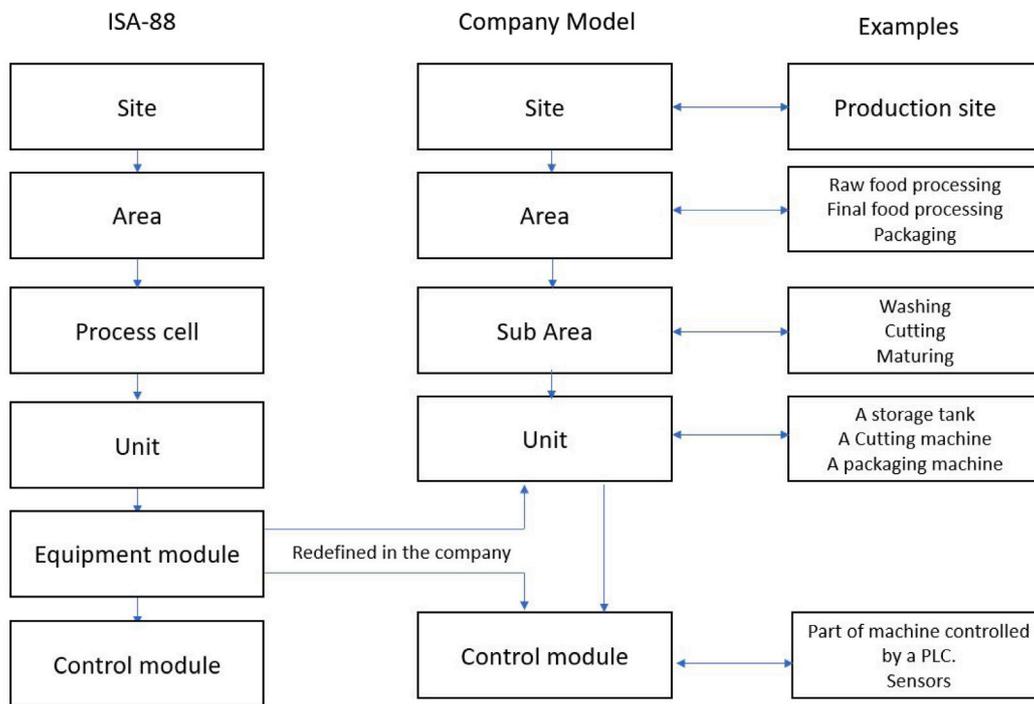


Fig. 2. ISA-88 anonymized mapping for the digital twin. Source: Created by authors based on information from the senior IoT data engineer.

collaboration, but the company has encountered a significant challenge in delivering the right data in the right form to casual and power users in its production domains.

Since the company is a traditional production company, the majority of its members in its production domains are domain specialists often with low data and analytics self-reliance (casual users). Part of the solution to this challenge is data education, as discussed in Zeng and Glaister (2018), Awasthi and George (2020), and Lefebvre et al. (2021), and the company has established data education initiatives. However, it is clear to the company that its data education efforts are a many-year project. To realize data democratization in the company, we and the company propose the creation of a digital artifact, which is a digital twin of the production process.

In essence, digital twins “link physical object[s] and digital object[s] in an accurate and real-time manner” (Liu et al., 2021), and in Liu et al. (2021) multiple slightly different definitions of digital twins are presented and discussed. The digital twin that is being developed seeks to create a digital object that links to the physical objects in the company’s production process, which consists of many different production machines located in different domains at many different production locations. In its production domains, the company creates and collects a large amount of IoT data from PLCs, sensors, and other data collection technologies. The implementation of digital twins in a production and supply chain context is discussed in detail in Werner et al. (2021) and Bao, Guo, Li, and Zhang (2019). For brevity, we will only discuss the specific digital twin being constructed for the case company and how it supports their data and analytic democratization process.

The case company is a batch producing company, and the digital twin is therefore being built based on the ISA-88 (ISA-88, 2023) and ISA-95 standards (ISA-95, 2023). When constructing digital twins, the use of the ISA-88 standard is also seen in the literature (Werner et al., 2021). Based on the ISA-88 standard, the physical structure of its production domains and its physical interconnections have been mapped, which in the ISA-88 standard is denoted as the asset model. The general and anonymized ISA-88 map for the case company is given in Fig. 2.

In Fig. 2 we show an anonymized version of the ISA-88 hierarchy constructed by the company and we present a few plausible anonymized examples. The ISA-standard also includes an enterprise level, which is the collection of the production sites of the company. At the control module level, the company collects streaming data from sensors and PLC’s in time series databases and from the equipment module and up the company collects data stored in relational databases. To connect the data from the control module with data from the equipment module and up, the company uses the ISA-95 standard. The ISA-95 standard contains an asset model, materials model, process model, and organizational model (ISA-95, 2023), which together allows the company to create a digital twin of the production process that connects the data available in its time series and relational databases. However, the company has found it necessary to use the more granular ISA-88 standard for its asset model, and that is why the digital twin of the production process of the company is being built using a combination of the ISA-88 and ISA-95 standards.

Constructing the digital twin is a costly project for the case company, but the company has identified that the majority of users in the production process can meet their data needs from the digital twin built on the ISA-88/95 standard. That is because the majority of users need to trace a product batch as it moves through the production process or track the behavior of machines in different production domains as different product batches move through the machines. The digital twin based on the ISA-88/95 standard is therefore suitable for mapping the company’s batch production (Werner et al., 2021).

Fig. 3 is an abstracted example of a possible construction of the production process within a site, where raw food products come into the site and finished food products leave the site. The green arrows represent how the products move through the site in batches, and the digital twin gives a digital representation of the batches as the product moves through the production process, and stores how the batches are connected. The yellow arrows represent products moving through the production machines (equipment modules) in batches, and the digital twin gives a digital representation of the production machine as products are processed by the production machines, and stores

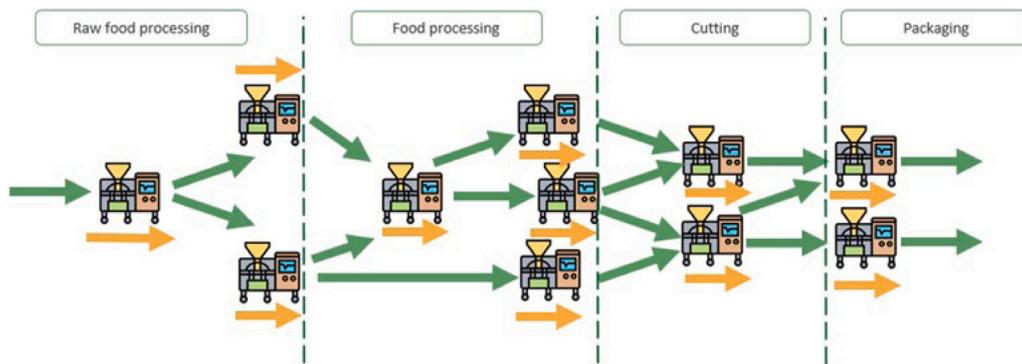


Fig. 3. Abstracted example of the production process within a site.  
Source: Created by authors based on information from the IT portfolio manager.

the characteristics of the batches having moved through a production machine.

The digital twin built on the ISA-88/95 standard is envisioned to be a core element of the data democratization process because it will give casual and power users a digital representation of the production process in which they are domain experts. Casual and power users will interact with the digital twin using SSBI systems, such as PowerBI. When constructing the digital twin, the company is also standardizing its data so that casual and power users can interact with the digital twin using production terminology. Casual and power users will therefore not need to understand or be able to query into the underlying IT systems of the digital twin. From interviewing and interacting with stakeholders in the case company, it was observed that implementing the digital twin on heterogeneous IT infrastructures at different sites complicates the process potentially requiring more resources than initially planned.

From talking with the IT portfolio manager, it was also noted that even though the digital twin is designed from a data democratization inspired theoretical base, ensuring that the visualized data presented in SSBI systems is actually in accordance with users' subject-matter expertise, this requires significant user feedback during the development of the visualizations. The user feedback is to ensure that the correct and relevant physical units are being visualized, that the data time-indexing is meaningful, and that the needed contextual production data, such as batch IDs and machine states, are presented in a way that the users can interpret. The digital twin also aims to enable power users with a high degree of self-reliance to query the underlying databases. When the digital twin was partially implemented, it was observed that power users, such as the senior data scientist BI specialists, were able to query the digital twin and access contextualized IoT and production data much more efficiently than by identifying, extracting, and transforming data from different previously unconnected databases.

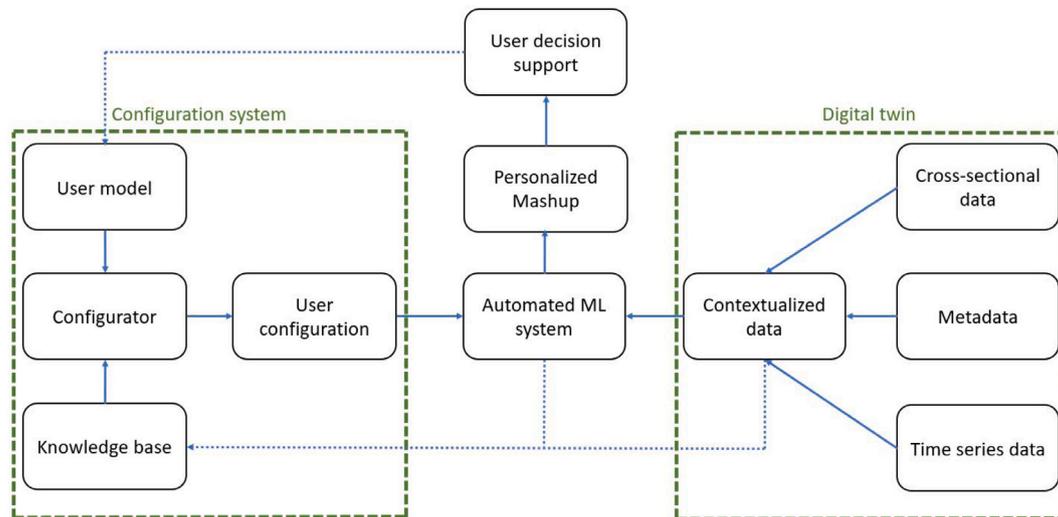
## 6. Mass customizable mashups intervention

The data democratization process of the case company aims to give casual and power users in the company the capabilities to access and visualize the data they need in their domain of the production process. While data access and data visualization can aid in decision-making (Alpar & Schulz, 2016), using analytics to support decision-making is also a key to the success of SSBI systems (Lennerholt et al., 2018). The company expects that its data democratization efforts and the implementation of the digital twin will lead to a greater demand for analytics, since users in its production process demand the use of data to improve their decision-making. A greater demand for analytics is also expected since the company is seeking to foster a culture of data driven decision-making through its data and analytics educational initiatives. As discussed above, the company's casual and power users

in the production process demand role-tailored analytics trained on IoT and production data, which support decision-making in areas such as machine optimization, maintenance planning, and production planning. The challenge the company is facing is that it does not have the analytics and IT resources to meet all the expected demand for personalized analytics.

To overcome this challenge of delivering personalized analytics to casual and power users, we propose the concept of mass customizable mashups, which is designed based on insights from the IS literature on mass customization. Mass customization is the concept of efficiently creating individualized and personalized products and services (Baranauskas, Raišienė, & Korsakienė, 2020; Coito et al., 2020; Gong et al., 2023). Just as advances in manufacturing technologies and methods have enabled mass customization, we see the emergence of automated machine learning technologies and frameworks (Elshawi & Sakr, 2020; Yang & Shami, 2022; Yao et al., 2018; Zöllner & Huber, 2021) as enabling mass customizable mashups. In brief, automated machine learning seeks to automate analytical processes, such as data processing, feature engineering, analytics training, analytics validation, and prescription generation. When the automated machine learning system has been created for a given set of operational problems, users only need to specify what type of problem they want to solve, its parameters, and what IoT and production data to feed into the system.

For the initial design of the artifact, we have developed an automated machine learning system to be able to solve a machine setting optimization problem, where the aim is to determine the settings that minimize the energy consumption of the machine while ensuring certain product characteristics are achieved for a given batch and product. We have also applied it to a batch planning problem, where the cumulative output over time of a given batch is predicted, and this prediction is used to determine the optimal batch length when the cleaning times are also taken into account. To validate that we can actually call it an automated machine learning system, it was developed and applied to two different types of production machines, pasteurizers and ultrafiltration machines, and for each type of machine it was applied to a number of different machines across three different production sites at the case company (ultrafiltration machines were only from one site). In Fig. B.10 in Appendix B we show the design of the automated machine learning system. The predictive modules of the system (data cleaning and model training, and model validation) follow the literature on automated machine learning, and to achieve prescriptive models we added a module to the system where the specific type of optimization problem is defined and solved. That module is an automated machine learning version of the concept of data driven prescriptive analytics (Bertsimas & Kallus, 2020), where data is used not only to generate predictions, but the predictions are used to generate decisions based on an optimization problem. The prescriptive module takes problem information from the knowledge base and user



**Fig. 4.** Mass customizable mashups system design.  
 Source: Created by authors as an extension of the work of Felfernig et al. (2014).

configuration, described below, and data and predictions from the previous modules of the automated machine learning system in order to define and solve the optimization problem.

Although, the automated machine learning system is designed to provide actionable analytics, the user must be able to give relevant information regarding the problem objective, constraints, and the input data. The user then must be able to receive the data driven prescriptions with sufficient contextualization to act on them. This is where the insights from the mass customization literature become relevant, since a core part of the IS literature on that topic is on the construction of configuration systems that both take information from users to feed into production processes and how to visualize the effect of choices of the users on the product (Felfernig, Hotz, Bagley, & Tiihonen, 2014; Tiihonen & Felfernig, 2017). Mass customizable mashups is the name we give to the system we create by using insights from mass customization literature to create a configuration system for an automated machine learning system. In Fig. 4, the design of this system is presented.

Fig. 4 displays the core elements of the configuration systems and their connection to the automated machine learning system and the digital twin introduced earlier. The arrows illustrate how the different elements are connected. Moreover, as illustrated by the blue dotted lines in Fig. 4, the decision support needs of the user inform the user model which captures the user's. The user model contains the user's choice of type of operational problem (from a set of operational problems that are linked to the corresponding predictive and optimization problems in the automated machine learning system), choice of input data, problem objectives, problem constraints, and what form of contextualization of the optimal decisions the user wants. The knowledge base of the configuration system then implicitly defines the set of possible configurations (Felfernig et al., 2014; Tiihonen & Felfernig, 2017). In this case, the knowledge base contains knowledge about the possible problem objectives for the optimization problems, possible constraints, possible input data, and possible results visualization options. Moreover, the knowledge base holds rules for how objectives, constraints, and input data interact and are related. The output of the configurator, which is a valid configuration (Yao et al., 2018), is passed on to the automated machine learning system where data processing, predictive model training, prescription generation, and results visualization are automated for the specified operational problem. The automated machine learning system has been developed using programming languages such as R and Python. Since the user also chooses their preferred way of visualizing the optimal actions in the

configuration system, the results of the automated machine learning and the prescribed actions, and the consequences of said actions, are presented in a mashup to the user in a self-service system, such as a dedicated app or in SSBI software such as PowerBI.

#### 6.1. Implementation of artifact and initial evaluation of artifact

The digital artifact of mass customizable mashups was developed and refined together with individuals in the case company as part of the research process. The automated machine learning system, including the prescriptive module, was developed and validated on pasteurizers at multiple production sites. To test if the system was applicable to other types of machines, it was applied to a set of ultra filtration machines at a new site, where it was also possible to perform a test of the prescriptions of the system on the machine as part of the production process. It was not possible to do a full test of the configuration system, but feedback on the configuration system was collected. In order to complete the IBR research cycle, the configuration system must be implemented and tested. At the time of writing this remains a goal of the research, but it is not currently possible. However, this partial evaluation of the artifact still offers valuable insights, and in the following, we will therefore introduce the specific operational problem to which the automated system was applied, and the statistical test results of that application.

The operational problem that the automated machine learning system is designed to solve is to determine the machine settings (such as temperatures, flows, pressures, additions to the product, etc.) that minimize the total energy consumption (such as electricity, heating, and cooling) when producing a given product depending on the product characteristics (such as fat content, protein content, water content, etc.), and batch characteristics (such as batch length in time, machine age, and time since maintenance). This problem must be solved while ensuring that certain product characteristics or quality measures are met. This is an important problem for the case company because in parts of its production processes it faces an inherently stochastic production process, which is also influenced by the age and maintenance of a given machine. Moreover, for business and product reasons, the case company also changes its inputs and desired quality characteristics of its products over time. In this type of production environment, where millions of liters of materials can be processed every year at a single site, we observed that suboptimal machine settings could result in total energy consumption of a batch that is significantly higher than what

**Table 3**  
Predictive accuracy results for the automated machine learning system.

Accuracy measure	Electricity	Cooling	Product flow
MAPE	7.41%	11.09%	7.21%

are the realistic and optimal settings for a given batch. Determining optimal machine settings when input or product characteristics change is already something the case company spends resources on based on physical models and the knowledge of operators and engineers. This knowledge is used to determine the machine settings for a given batch, and we generally observed that settings were well-chosen, but the complexity of determining the settings that are optimal for multiple sources of energy and product and quality characteristics does leave the possibility of further optimization open. We are not the first researchers to explore the use of machine learning in related problems in the context of pasteurizers (Khadir & Ringwood, 2003). However, the decision engineers and operators ultimately need to make is what settings to use for a given batch, and predictive machine learning models can help determine that, but purely predictive machine learning models do not prescribe what settings to use for a given batch. That is why we have added the optimization module to the automated machine learning system, since the output of that module is the optimal settings for a given batch given user-specified constraints.

This type of operational problem is relevant for pasteurizers, for which the automated machine learning system was first developed. Pasteurizers generally consume electricity, heating, and cooling and the primary quality characteristic is to ensure that milk is pasteurized. Moreover, this type of operational problem is also relevant for ultrafiltration machines that generally consume electricity and cooling and where the primary quality characteristic is to ensure that the flow through the machine is not too low. To apply the automated machine learning system to this machine we connected the sensor and contextualized data, in the same format as it would be present in the digital twin, and created the relevant constraints and search space for the optimization problem, since the objective function is the same for both types of machines. Specifically, we started with data from one ultrafiltration machine. This enabled us to determine the optimal settings for minimizing the total electricity and cooling consumption in kWh, while also accounting for the fact that lower flows result in a higher kWh equivalent cost. For confidentiality reasons, we cannot describe the exact number of sensors and the contextualized data sources that were used when applying the automated machine learning system to the filtration machines. Following the system in Fig. B.10, the data cleaning module was applied, a machine learning model was trained to predict energy consumption of electricity and cooling for a given batch, and a predictor of the flow of the machine was also trained. The regression model chosen for this automated system was an XGBoost regression tree (Chen & Guestrin, 2016) with automated machine specific tuned hyperparameters. For confidentiality reasons, we can only report the standard statistical accuracy measures and mean average percentage error (MAPE) for the predictors of energy consumption and the flow. MAPE is calculated based on a standard 80%–20% train-test data split, and the results are reported in Table 3.

The MAPE values reported in Table 3 show that when predicting the energy consumption of the machine for a given batch, the machine learning model is on average 7.41% and 11.09% percent away from the true value for electricity and cooling energy consumption respectively. For the product flow prediction, which is used in the prescription generation module to ensure that the product flow of the machine does not become too low, the prediction is on average 7.21% percent away from the true value. Generally, the MAPE values indicate that the machine learning model can generate accurate predictions, and we observed similar accuracy for the pasteurizers. One could likely improve the accuracy of the predictions by using specialized machine

**Table 4**  
Changes in energy consumption after the change of machine settings.

Energy type	Observed change	Predicted change
Electricity	–19.50%	–17.20%
Cooling	–29.43%	–28.69%

learning models, but the XGboost regression tree is easily integrated into the automated machine learning system. Moreover, getting the highest accuracy possible on historical data does not in itself enable users to improve decision-making. What is important is that the machine learning model can give good enough results when applied to new unseen combinations of inputs and machine settings, and focusing too much on the historical data introduces significant risks of overfitting on the historical data.

As discussed above, we are not primarily interested in predicting the energy consumption of the machine, but we are interested in determining the optimal machine settings for a given product batch. This is done in the prescription generation module in Fig. B.10 where the predictions of the energy consumption and the product flow are put into an optimization problem which is solved using a genetic algorithm (Lambora, Gupta, & Chopra, 2019). The algorithm searches, within a user-defined search space, possible combinations of machine settings in order to determine the specific machine settings that minimize the energy consumption while keeping the product flow above a certain threshold for a given batch. The prescription generation problem is, in itself, a relatively simple optimization problem, but the inclusion of predictions in the optimization problem necessitates the use of a numerical solution algorithm like genetic algorithms. A thorough mathematical treatment of the automated machine learning system, including the optimization module, is beyond the scope of this paper and will be presented and discussed in forthcoming research. However, for the ultrafiltration machine, we were able to conduct an experiment, where the automated machine learning system was able to identify that an important setting was set too low for a set of future production batches. Exactly which setting was changed cannot be disclosed for confidentiality reasons. It was predicted that changing the setting would result in significant savings. We were permitted to instruct the operators to adjust the setting to a new value determined by the automated machine learning system. Subsequently, we collected three weeks' worth of data from the production line on which the modified machine setting was implemented. The results of the experiment are shown in Table 4.

In Table 4, the numbers reported in the *observed change* column are simply the observed average energy consumption of the batches produced over the three weeks after the setting was changed compared with the average observed energy consumption of the batches produced on the machine one month back. The observed reductions in energy consumption of 19.50% for electricity and 29.43% for cooling significant and result in substantial monetary and emission savings when producing on the ultrafiltration machine. For the batches produced after the setting was changed, we tried to keep product characteristics and batch characteristics the same as before the change, but we cannot with certainty fully attribute all of the observed reductions in energy consumption to the change in the machine settings, since the experiment was conducted on a machine used in the production process and there might be differences on which we do not have data. However, to give credence to our results, we also used our trained models to predict the effect of changing the setting. These are the numbers reported in the *predicted change* column in Table 4. The predictions were made for the energy consumption of the machine for the three test weeks using only the historical data of the machine up to the experiment start date. We do this by creating a prediction using the optimized machine settings, the known product characteristics data for the batches (since that is planned beforehand), and by projecting the

The image shows a configuration system interface. On the left is a vertical sidebar menu with the following items: Configuration system, Machine Type, Machine Number, Product Selection, **Analytics Selection** (highlighted), Dependent Variable Selection, Constraint Variable Selection, Independent Variable Selection, Prediction Analytics, Constraint Selection, Independent Variables Search Space, Run Time Selection, Weights Selection, Analytics Solving, and Decision Visualisation Mashup. The main content area is split into two panels. The left panel, titled 'Select Analytics Type', has the heading 'Choose the desired analytics' and two radio button options: 'Constrained Energy Optimization' (selected) and 'Stability optimization'. Below these are a 'Save Choice' button, the text 'Chosen analytics: Constrained Energy Optimization', and a 'Go to next page.' link. The right panel, titled 'Select Independent Variables', has the heading 'Choose one more input variables to include in the analytics' and a list of 13 variables from 'Independent\_1' to 'Independent\_13'. Variables 3 through 13 are checked with red boxes. Below the list is a 'Save Choices' button, the text 'Saving selected variables.', and a 'Go to next page.' link.

**Fig. 5.** Analytics and independent variables selection pages.  
*Source:* Examples from prototype of configuration system with anonymous data created by authors.

batch characteristics data forward in time to the three test weeks. This enables us to make predictions about the energy consumption of future batches using the optimal settings based on information available to the user prior to making the decision. The predicted changes are then the average of these predictions compared with the average observed energy consumption of the batches produced on the machine going back one month. The predicted reductions in energy consumption are 17.20% for electricity and 28.69% for cooling. These figures are within a couple of percentage points of the observed reductions, which supports the idea that the change in settings has primarily led to the reduction in energy consumption. Overall, we believe that the results of the experiment are strong evidence that researchers and practitioners can use an automated machine learning system as proposed in this paper to solve important problems in production contexts, such as the machine optimization problem we have explored in greater detail here.

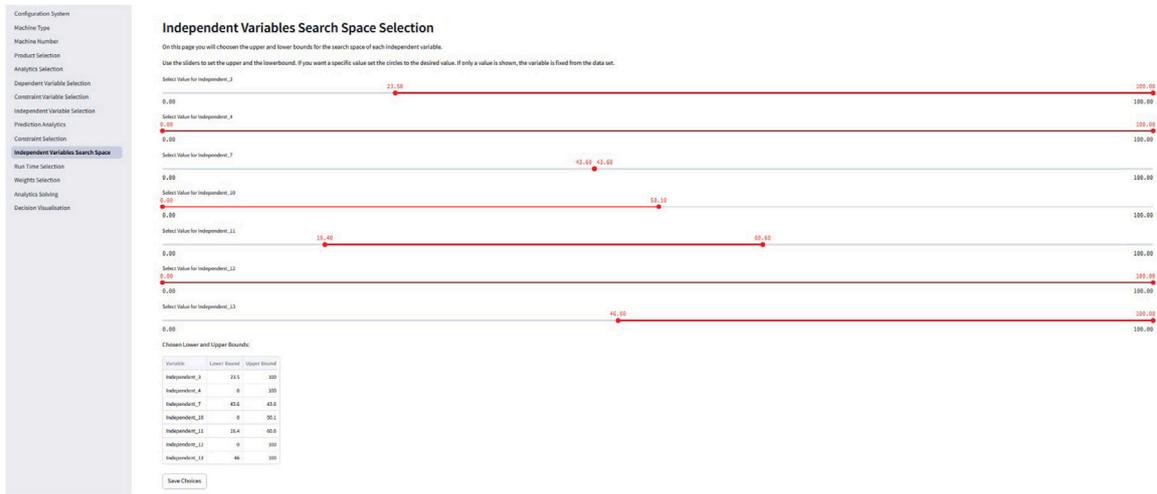
With the automated machine learning system introduced above, and evidence for the actionability of the outputs of the system in a machine setting optimization context, the next question is then how do users interact with this system to specify the batches they want to optimize, how do they receive the suggested settings, and how can the settings be visualized to add context to the user's decision-making. The answer to this question is the configuration system that has been developed for the mass customizable mashups digital artifact. As part of the research, the authors of this paper developed and programmed a prototype of this configuration system, which is illustrated in anonymized form in Figs. 5–8, and 9. These figures are discussed in more detail in the following paragraphs. The steps of the configurator are shown on the left of the figures, along with examples of what users will see. In Fig. 5, two steps are put in the same figure for brevity. All numbers in the figures are anonymized and are purely illustrative.

In the configurator, the user specifies which variables should be optimized, such as minimizing the usage of different types of energy (cooling, heating, and electricity), which variables control the process, such as which temperatures need to be reached to ensure food safety (which is one of the possible constraints of the problem), and which variables the user can use to configure the machine. The knowledge

base implicitly guides user's the choices, such that only variables associated with the chosen machine are presented. This ensures that the objective function, constraints, and selected input variables constitute a feasible energy optimization problem. Once the user has configured their personalized machine optimization problem, it is sent to the automated machine learning system, which determines the optimal configuration of the specified production machine for the specified product. The results are then presented to the user in a personalized mashup, where the optimal values of the user-chosen variables and their consequences are visualized.

In Fig. 5, we show examples of two pages of the configuration system that have been pasted together to save space. These illustrate the way the users are presented with the choices they need to make, which is also implicitly guided by the knowledge base of the configuration system. In these examples, we demonstrate how the user must select the type of problem they are facing, which can be selected after machine type, machine ID, and product type (see the left-hand column). In the empirical case discussed above, it is here the user would select the optimal machine setting problem for a given machine. We also illustrate how it looks like when the user has to choose what specific variables to include in the analysis for a given problem, machine and product combination (the set of possible variables implicitly guide by the knowledge base). In this configurator, these are chosen after the user has specified what energy consumption mix they want to model and what the product or quality characteristics are. In essence, it is here the user selects what machine settings they would like to optimize. The headline and description in these examples are quite general because names and descriptions have been anonymized with general terminology.

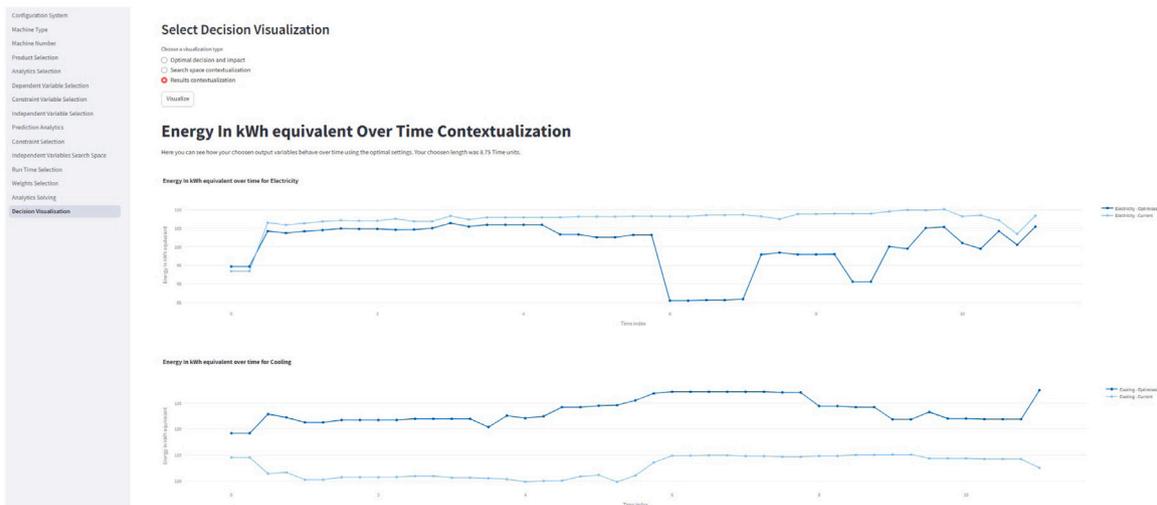
Fig. 6 shows one of the final steps in the configurator, where the search space of the optimization problem is defined. This defines the range of values of the independent variables within which the genetic algorithm can search for an optimal solution. Only the previously chosen independent variables can be searched, and the bounds of the search space are implicitly set from information in the knowledge base and digital twin. These bounds can also be modified explicitly by the



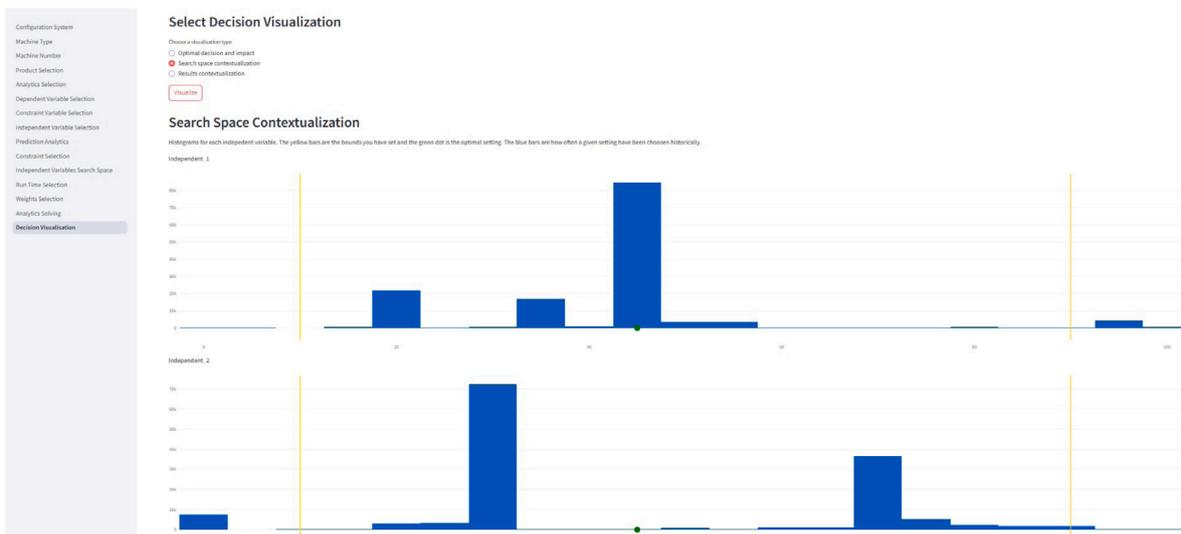
**Fig. 6.** Search space for independent variables configuration. *Source:* Example from prototype of configuration system created by authors. All numbers are anonymized.



**Fig. 7.** Personalized mashup that shows optimal machine settings and effect. *Source:* Examples from prototype of configuration system created by authors. All numbers are anonymized.



**Fig. 8.** Personalized mashup that shows the effect over time of the machine settings. *Source:* Examples from prototype of configuration system created by authors. All numbers are anonymized.



**Fig. 9.** Personalized mashup that show new machine setting compared with previously applied machine settings.  
*Source:* Examples from prototype of configuration system created by authors. All numbers are anonymized.

user. In the machine setting problem, the user can see the preset bounds for the machine settings here and make the bounds smaller if needed. Similar pages are presented to the user where they set values for the product or quality constraints and batch characteristics constraints if relevant.

The possible personalized mashups currently available in the prototype are shown in Figs. 7, 8, and 9. The user can choose between different options that contain different relevant information and visualizations, and these are personalized because the content is also dependent on the previous choices the user has made in the configuration system. In Fig. 7, we present the recommended actions, which, in the case of the machine setting optimization problem, are the optimal machine settings for a given batch on a given machine. Together with the estimated effects of the chosen settings on energy consumption, and with the product or quality values (stability in this case), the user has everything they need to set new settings for a given batch. If the user wants more context before making a decision, other mashups are available in the configuration system. Fig. 8 shows a mashup where the user can see the behavior of the dependent variables over time, which, in the machine setting optimization problem, is the energy consumption over time. This is relevant since the production time may be shorter or longer than expected, and the user can thereby check the robustness of the optimal settings, or use this information to investigate why the chosen machine behaves differently for different production lengths. Fig. 9 shows a mashup that allows users to see whether optimal machine settings are similar to previously chosen settings and where in the search space the optimal settings located. The green dot indicates the optimal value for a given machine setting, the blue bars show a histogram of the previously chosen values for the machine setting, and the yellow vertical lines illustrate the upper and lower limits of the search space. This mashup gives context to the user's choice of search space and allows the user to see if the optimal settings are close to previously chosen machine settings or not.

A full implementation and evaluation of the configuration system has, by the time of writing not been possible, but in interviews with both the information system infrastructure senior manager and senior IoT infrastructure manager, the potential of the configuration system approach was identified. It was highlighted that such a configuration system could be a core part of generating value from a digital twin of the production process, since it allows users to directly generate optimal actions and actionable insights from the data available in the digital twin. From an IT infrastructure perspective, it was discussed

whether such a configuration system should be a stand-alone application drawing data from the digital twin or whether it should be integrated, if possible, either completely or in modified form, into the SSBI software suite that connects to the digital twin. A central challenge in implementing the configuration system is that changes in the digital twin regarding what type of data is available, and in what underlying databases data is stored and named, require updates of both the knowledge base and the automated machine learning system. This makes it difficult to develop the digital twin and the configuration system in parallel. However, digital twins are also costly to develop and implement, which can lead to dissatisfied stakeholders if business value from the digital twin is not realized in a reasonable time. Therefore, extra effort is also needed to ensure that business value can be extracted from the digital twin through systems such as mass-customizable mashups in order to meet stakeholder demands.

The prototype of the configuration system was developed based on insights from Felfernig et al. (2014), and not surprisingly, potential problem areas in implementing the configuration system identified in that literature were also observed in the case company. As mentioned earlier, the case company has many casual users with low self-reliance and power users with different degrees of self-reliance. The design of the configuration systems must take this into account. Meetings with users were conducted by the senior data scientist, with some attended by the academic researchers. At these meetings, machine setting selection and visualizations based on those in the configuration system were discussed providing important insights. Depending on their level of self-sufficiency, different users wanted more decisions to be pre-selected and/or suggested by the configuration system. This is also a problem area that has been discussed in the literature. The senior data scientist observed that users generally found the new machine setting suggestions actionable and valuable, but that it was essential to build trust in the statistical results and suggested machine settings by visualizing the latter. This is because the visualizations could give context to the users on which they can apply their subject matter knowledge. However, it was discussed that depending on the self-sufficiency of the user, training should be developed to give users the necessary knowledge to interpret the visualizations. Moreover, it was observed that different users preferred different types of visualizations, as exemplified in Figs. 7, 8, and 9, some users preferred more visualizations with context, such as the degree of change from current machine settings or the impact of the changes over time, and others only wanted to know the new machine settings and their impact on energy consumption. While the configuration system was built to accommodate different types of

**Table A.5**  
Data and analytics democratization literature review.

Paper	Topic of paper	Understanding of democratization	Understanding of self-service or SSBI
Chergui and Kechadi (2022)	Explores analytics for crop management.	States that analytics have become democratized.	No mentions.
Harland et al. (2022)	Explores the state of data democratization in industry.	'At the core of the concept of "Data Democratization" is the idea to provide access to the company's data resources to all employees given reasonable limitations on legal confidentiality and security.'	Mentions self-service systems as tools for citizen analysts.
AbuHalimeh (2022)	Explores data quality in a clinical setting.	Mentions but do not elaborate on data democratization.	Mentions the importance of data quality in BI and SSBI systems.
Alpar and Schulz (2022)	Explores the concept of citizen data scientists.	'The term refers to organizations, esp. firms, where employees receive access to data without relying on IT people or DS.'	Self-service BI systems are tools to access data and enable democratization.
Ismail and Alias (2021)	Explores democratization from A TOE framework.	Democratization is the successful integration of technology, organization, environment, and stakeholders.	No mentions.
Saeed and Sidorova (2021)	Explores data democratization in an energy context.	'In brief, data democratization is the process of bringing digital information to the average, non-expert end user.'	No mentions.
Awasthi and George (2020)	Explores democratization of IT.	'Data Democratization is the act of opening organizational data to as many employees as possible, given reasonable limitations on legal confidentiality and security'	No mentions.
Vuppapapati et al. (2020)	Explores using analytics in context of small farms.	States that analytics have become democratized.	No mentions.
Terrazas, Ferry, and Ratchev (2019)	Explores shopfloor data management.	States that use of sensors have become democratized.	No mentions.
Hyun, Hosoya, and Kamioka (2019)	Explores the effect of democratized culture	'Hence, 'democratization culture' in this study is defined as an organizational culture that values the sharing of knowledge and willingness to accept diversity on the basis of new knowledge.'	No mentions
Alpar and Schulz (2016)	Explores self-service business intelligence.	The ability of users to perform data analysis using self-service business intelligence tools.	The use of self-service business intelligence tools are equated to democratization.
Kaminwar, Goschenhofer, Thomas, Thon, and Bischl (2023)	Explores validation techniques for machine learning models.	Democratization of artificial intelligence is giving scale and accurate machine learning models to non-experts.	No mentions

users, the users who did not need the flexibility of the configuration system, could find it unnecessarily cumbersome. In future iterations of the configuration system, the feedback from users and insights from the IS literature on configuration systems can be applied to refine the configuration system element of mass customizable mashups.

## 7. Conclusion and future research

Making SSBI systems successful in a production context is a complex challenge, and in this paper, we have analyzed the root causes of this challenge. Moreover, based on this analysis, two digital artifacts have been designed, and we have provided insights from the initial implementation and iterative development of the artifacts in the case company. The first digital artifact we proposed is based on the fundamental insight that data availability is in itself rarely sufficient for data driven decision-making. Instead, users need contextualized data, which presents data to users in accordance with the domain knowledge. In a production context, we propose a digital twin based on the ISA-88 and ISA-95 standards, which is being implemented in the case company, as the source of contextualized data for SSBI systems. Proposing to use the two ISA standards is not in itself a novel contribution, but we contribute to the academic IS literature by also identifying the need to draw upon insights from the data democratization literature when

developing and implementing a digital twin. By integrating insights from the data democratization literature, researchers and practitioners are in a better position when designing a digital twin that not only makes data available but also ensures that the data needs of users are met. This was also observed in the development of the digital twin at the case company: The core ISA-based approach could be used to guide the design of the IT infrastructure of the digital twin, but there were significant challenges and costs associated with identifying and contextualizing data across different types of production machines. However, when the digital twin was partly implemented, and the first contextualized data became available, power users could efficiently access data that directly aided in decision-making and could be used as the data source for prescriptive advanced analytics.

Another clear implication of this paper is it provides further evidence in support of the claim made by SSBI researchers that it is paramount for practitioners to identify not only what data is available and demanded by users, but also the context is required for that data to actually generate actionable and insightful information for the target users. This is especially the case when expanding SSBI systems to organizational areas where SSBI systems are not already widely in use. In such contexts, as in the case of the production processes of the case company, ensuring that data is actionable is important to achieve adoption and value generation by users, especially users with low self-reliance. Although this research was performed in a production context,

Table A.6

Mass customization literature review.

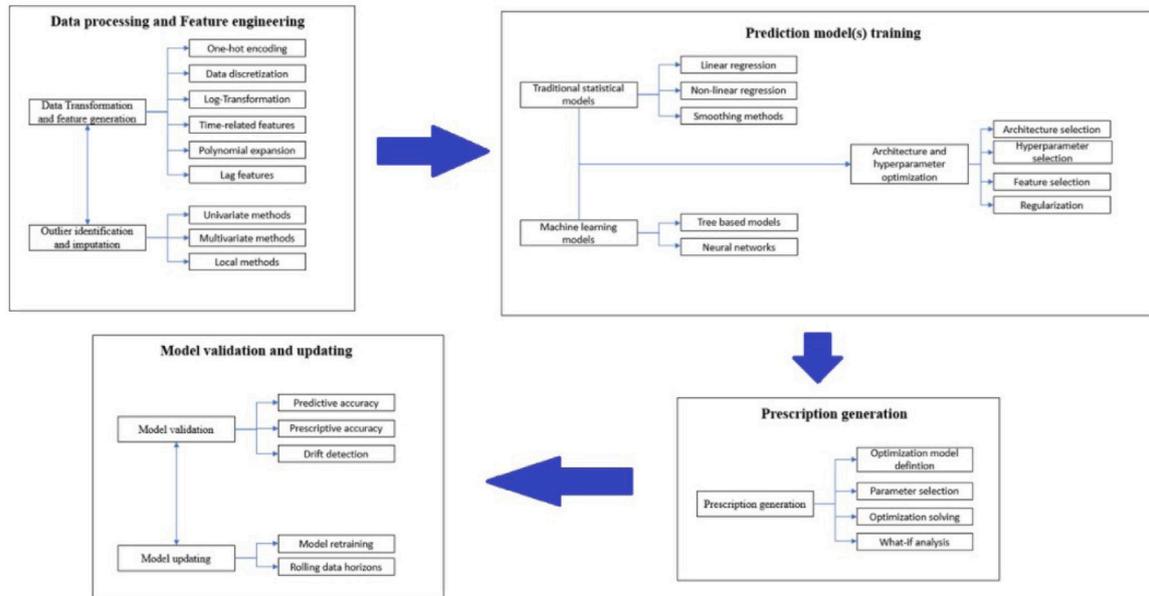
Paper	Topic of paper	Understanding of mass customization.	Understanding of self-service or SSBI
Gong et al. (2023)	Explores mass customization of brace and use of analytics to determine producers.	The capability to efficiently make products to personalized specs	Professionals interacts with a product spec IT system, which ensure that the personalized product is able to be produced.
Dudek et al. (2023)	Explores the integration of consumer preferences into ERP systems.	'Mass customization is an economic rebalancing strategy dictated by the current competitiveness of the market. It involves manufacturing different product models while reducing costs in terms of tools and equipment.'	The customers give their products specs in an online system which ensure that the personalized product is able to be produced.
Agrawal, Navgotri, and Nagesh (2023)	Explores new manufacturing technologies.	States that internet shopping has led to higher demand for mass customization.	No mentions.
Coito et al. (2020)	Explores the standardization of Industry 4.0 system architectures.	'Mass customization is the effort of combining both the low costs of mass production with the flexibility of custom production.'	No mentions.
Xia, Fang, Gebrael, Xi, and Pan (2019)	Explores using sensors for opportunistic maintenance.	'In mass customization, a manufacturing line normally processes variable product orders, and system disruptions need to be kept to a minimum.'	No mentions.
Detzner, Eigner et al. (2018)	Explores using a digital twin for root cause diagnosis.	States that mass customization requires new methods to ensure product quality.	No mentions.
Wache, Dinter, and Kollwitz (2019)	Explores customer analytics to support mass customization.	States that mass customization is a core part of current industry.	No mentions.
Zhong et al. (2017)	Explores new manufacturing technologies.	States that mass customization is a core part of industry 4.0.	Cloud systems in industry 4.0 should have self-service systems.
Saldivar et al. (2016)	Explores a predictive analytics framework supporting industry 4.0.	States that mass customization is a core part of industry 4.0.	Use of analytics to capture users design needs and presenting prototypes to users.
Tien (2013)	Explore customer analytics to support mass customization.	States that mass customization is a core part of the third industrial revolution.	No mentions,
Tien (2012)	Explore the third industrial revolution.	States that mass customization is core part of the third industrial revolution.	States that further development human machine interfaces are important.
Dou, Sun, Wu, Wang, Fan et al. (2021)	Explores the integration of mass customization and the social internet-of-things.	'Mass customization (MC) has reformed the service-oriented Behavior of manufacturing enterprises.'	Different users interact with mass customization system through different IS systems.
Baranauskas et al. (2020)	A literature review of the mass customization literature.	'Mass customization (further MC) is primarily recognized as a differentiation strategy in the manufacturing and retail spheres that provide a wide range of customized products or services as well as co-design or other value co-creation options.'	No mentions.
Martínez-Olvera (2020)	Explores energy what-if calculations in a mass customization production process	'A mass customization production system operating within a reconfigurable CPS context.'	No mentions.
Feng and Shanthikumar (2018)	Explores new areas of research into the use of big data.	States that mass customization is a core part of current industry.	States that the use of smart production and human interaction with smart production through digital interfaces will keep growing.
Jamwal, Agrawal, Sharma, and Giallanza (2021)	A literature review on the state of industry 4.0.	States that mass customization is a core part of industry 4.0.	No mentions.

contextualized data is likely to be important for actionable SSBI systems in other organizational contexts too.

The contextualized data enables the second and completely novel digital artifact, which we introduce to the literature in this paper, and which we denote as mass customizable mashups. At the conceptual core of the artifact is the insight that SSBI systems, which are planned to support operational decision-making, must be prescriptive by design and directly support the decision-making needs of users. This paper's focus on how SSBI systems directly support decision-making is its core contribution. Mass customizable mashups constitute a system designed for this purpose in the context of production processes. While statistical and mathematical optimization models can be used to model decision-making processes, users still need to ensure that their specific problem is solved and that the system's outputs are actionable. For this reason, we designed the configuration system as part of the artifact.

We believe that integrating automated machine learning based analytics and configuration systems could also be highly relevant in other organizational contexts, such as supply chain planning, and the work in this paper thereby has potential implications in other organizational contexts. Moreover, we believe that a core implication of our research is that when researchers and practitioners in the area of SSBI systems seek to integrate analytics into their systems, then it must be prescriptive analytics that give actionable insights. The SSBI systems must also be designed so that users are able to apply them to the specific problems they are facing. Regarding future research, we believe that this paper demonstrates both the opportunities and challenges of applying the IBR methodology. In this paper, we discuss and demonstrate why information system research into data driven decision-making should consider an IBR methodology, since it enables researchers to both analyze data driven decision-making in an empirical problem situation and for researchers to test their insights by improving data driven

### Automated machine learning systems including prescription generation



**Fig. B.10.** Automated machine learning system including prescription generation module.  
Source: Created by the authors as an extension of the work in [Yang and Shami \(2022\)](#).

decision-making in the problem situation. However, future research should learn from the challenges discussed in this paper, especially regarding the empirical evaluation of developed artifacts. For future research, we can only recommend that researchers, together with their case companies, plan the empirical evaluation of the digital artifacts from the very start of the research process.

#### Founding sources

Johan Bjerre Bach Clausen and Charles Møller are researchers employed at their respective institutions, but some part of their salaries during the research process were financed by Manufacturing Academy of Denmark (MADE) an independent industrial and research organization of which the case company is a dues-paying member. MADE through the research platform MADE FAST (Funding nr.: 9090-0002B) is also sponsoring (but not fully funding) the research published in this paper. MADE is sponsored by public and private grants; a comprehensive list of all relevant grants can be given if needed.

#### CRedit authorship contribution statement

**Johan Bjerre Bach Clausen:** Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Charles Møller:** Writing – review & editing, Supervision, Methodology, Formal analysis. **Claus Boye Asmussen:** Writing – review & editing, Validation, Methodology, Investigation, Conceptualization. **Steffen Lundgaard Jørgensen:** Writing – review & editing, Validation, Investigation, Conceptualization.

#### Declaration of generative AI in scientific writing

During the preparation of this work the author(s) used the tool ‘DeepL Write - AI-powered edits’ has been used as part of the grammatical proofreading processes. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

#### Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Claus Boye Asmussen and Steffen Lundgaard Jørgensen were during the creation of this paper employed at ARLA foods which is the collaboration company. Steffen Lundgaard Jørgensen has since changed to another position at a company in another industry. Johan Clausen have full editorial control of the content of the paper, and the collaboration company can only request information to be anonymized or omitted if anonymization was not possible.

#### Appendix A. Literature review tables

See [Tables A.5](#) and [A.6](#).

#### Appendix B

See [Fig. B.10](#).

#### Data availability

The data in this paper is from an anonymous collaboration company and is confidential and cannot be shared.

#### References

- AbuHalimeh, A. (2022). Improving data quality in clinical research informatics tools. *Frontiers in Big Data*, 5, Article 871897.
- Agrawal, P., Navgotri, S., & Nagesh, P. (2023). Impact of emerging technologies on digital manufacturing: Insights from literature review. In *Materials today: proceedings*.
- Alpar, P., & Schulz, M. (2016). Self-service business intelligence. *Business & Information Systems Engineering*, 58, 151–155.
- Alpar, P., & Schulz, M. (2022). More data analysis with citizen data scientists? In *World conference on information systems and technologies* (pp. 122–130). Springer.
- Analytics, L. (2015). State of self-service BI report.
- Arnott, D., Gao, S., Lizama, F., Meredith, R., & Song, Y. (2019). Are business intelligence systems different to decision support systems and other business information systems? In *ACIS 2019 proceedings: vol. 66*.

- Arnott, D., & Pervan, G. (2008). Eight key issues for the decision support systems discipline. *Decision Support Systems*, 44(3), 657–672.
- Awasthi, P., & George, J. J. (2020). A case for data democratization. In *AMCIS 2020 proceedings: vol. 23*.
- Bani-Hani, I., Pareigis, J., Tona, O., & Carlsson, S. (2018). A holistic view of value generation process in a SSBI environment: a service dominant logic perspective. *Journal of Decision Systems*, 27(sup1), 46–55.
- Bao, J., Guo, D., Li, J., & Zhang, J. (2019). The modelling and operations for the digital twin in the context of manufacturing. *Enterprise Information Systems*, 13(4), 534–556.
- Bara, A., Botha, I., Diaconita, V., Lungu, I., Velicanu, A., & Velicanu, M. (2009). A model for business intelligence systems' development. *Informatica Economica*, 13(4), 99.
- Baranauskas, G., Raišienė, A. G., & Korsakienė, R. (2020). Mapping the scientific research on mass customization domain: a critical review and bibliometric analysis. *Journal of Risk and Financial Management*, 13(9), 220.
- Bertsimas, D., & Kallus, N. (2020). From predictive to prescriptive analytics. *Management Science*, 66(3), 1025–1044.
- Bose, R. (2009). Advanced analytics: opportunities and challenges. *Industrial Management & Data Systems*, 109(2), 155–172.
- Cebotorean, E., & Titu, M. (2011). Business intelligence. *Journal of Knowledge Management, Economics and Information Technology*, 1(2), 101–113.
- Chandrasekaran, A., de Treville, S., & Browning, T. (2020). Intervention-based research (IBR)—What, where, and how to use it in operations management.
- Chard, R., Li, Z., Chard, K., Ward, L., Babuji, Y., Woodard, A., Tuecke, S., Blaiszik, B., Franklin, M. J., & Foster, I. (2019). DLHub: Model and data serving for science. In *2019 IEEE international parallel and distributed processing symposium* (pp. 283–292). IEEE.
- Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: From big data to big impact. *MIS Quarterly*, 36(4), 1165–1188.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining* (pp. 785–794).
- Chergui, N., & Kechadi, M. T. (2022). Data analytics for crop management: a big data view. *Journal of Big Data*, 9(1), 1–37.
- Coito, T., Martins, M. S., Viegas, J. L., Firme, B., Figueiredo, J., Vieira, S. M., & Sousa, J. M. (2020). A middleware platform for intelligent automation: An industrial prototype implementation. *Computers in Industry*, 123, Article 103329.
- Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98.
- De Mauro, A., Greco, M., Grimaldi, M., & Ritala, P. (2018). Human resources for big data professions: A systematic classification of job roles and required skill sets. *Information Processing & Management*, 54(5), 807–817.
- Detzner, A., Eigner, M., et al. (2018). A digital twin for root cause analysis and product quality monitoring. In *DS 92: proceedings of the DESIGN 2018 15th international design conference* (pp. 1547–1558).
- Dou, Z., Sun, Y., Wu, Z., Wang, T., Fan, S., & Zhang, Y. (2021). The architecture of mass customization-social internet of things system: Current research profile. *ISPRS International Journal of Geo-Information*, 10(10), 653.
- Dudek, A., Patalas-Maliszewska, J., & Kowalczywska, K. (2023). Automatic configuration of an order as an integral part of a cyber-physical system in a manufacturing operating according to mass-customisation strategy. *Applied Sciences*, 13(4), 2499.
- Eckerson, W. W. (2008). Pervasive business intelligence techniques and technologies to deploy BI on an enterprise scale. TDWI Best Practices Report.
- Eckerson, W. (2012). Business-driven BI: Using new technologies to foster self-service access to insights. Tableau Software.
- Elshawi, R., & Sakr, S. (2020). Automated machine learning: Techniques and frameworks. In *Big data management and analytics: 9th European summer school, EBIS 2019, berlin, Germany, June 30–July 5, 2019, revised selected papers 9* (pp. 40–69). Springer.
- Eom, S., & Kim, E. (2006). A survey of decision support system applications (1995–2001). *Journal of the Operational Research Society*, 57, 1264–1278.
- Felfernig, A., Hotz, L., Bagley, C., & Tiihonen, J. (2014). *Knowledge-based configuration: From research to business cases*. Newnes.
- Feng, Q., & Shanthikumar, J. G. (2018). How research in production and operations management may evolve in the era of big data. *Production and Operations Management*, 27(9), 1670–1684.
- Gong, X., Wang, S., Jiao, R. J., & Gebrael, N. Z. (2023). Collaborative contracting for manufacturing-as-a-service (maas) by information content measurement and decision tree learning. *Advanced Engineering Informatics*, 56, Article 101911.
- H. Davenport, T. (2014). How strategists use “big data” to support internal business decisions, discovery and production. *Strategy & Leadership*, 42(4), 45–50.
- Harland, T., Hocken, C., Schröer, T., & Stich, V. (2022). Towards a democratization of data in the context of industry 4.0. *Sci*, 4(3), 29.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 75–105.
- Hyun, Y., Hosoya, R., & Kamioka, T. (2019). The moderating role of democratization culture: Improving agility through the use of big data analytics. In *PACIS 2019 proceedings: vol. 181*.
- ISA-88 (2023). ANSI/ISA-88.00.01-2010 batch control part 1: Models and terminology.. <https://www.isa.org/products/ansi-isa-88-00-01-2010-batch-control-part-1-models>. (Accessed 14 August 2023).
- ISA-95 (2023). ANSI/ISA-95.00.01-2010 (IEC 62264-1 mod) enterprise-control system integration - part 1: Models and terminology. <https://www.isa.org/products/ansi-isa-95-00-01-2010-iec-62264-1-mod-enterprise>. (Accessed 14 August 2023).
- Ismail, N. H., & Alias, R. A. (2021). A conceptual framework for democratization of big data analytics (BDA). In *International conference of reliable information and communication technology* (pp. 707–712). Springer.
- Jamwal, A., Agrawal, R., Sharma, M., & Giallanza, A. (2021). Industry 4.0 technologies for manufacturing sustainability: a systematic review and future research directions. *Applied Sciences*, 11(12), 5725.
- Kaminwar, S. R., Goschenhofer, J., Thomas, J., Thon, I., & Bischl, B. (2023). Structured verification of machine learning models in industrial settings. *Big Data*, 11(3), 181–198.
- Khadir, M. T., & Ringwood, J. (2003). Linear and nonlinear model predictive control design for a milk pasteurisation plant. *Control and Intelligent Systems*, 31(1), 1–8.
- Kobielius, J., Karel, R., Evelson, B., & Coit, C. (2009). Mighty mashups: Do-it-yourself business intelligence for the new economy. (pp. 1–18). Forrester Research.
- Lambora, A., Gupta, K., & Chopra, K. (2019). Genetic algorithm—a literature review. In *2019 international conference on machine learning, big data, cloud and parallel computing* (pp. 380–384). IEEE.
- Lefebvre, H., Legner, C., & Fadler, M. (2021). Data democratization: toward a deeper understanding. In *Proceedings of the international conference on information systems*.
- Lennerholt, C., Van Laere, J., & Söderström, E. (2018). Implementation challenges of self service business intelligence: A literature review. In *Proceedings of the 51st hawaii international conference on system sciences*.
- Lennerholt, C., Van Laere, J., & Söderström, E. (2021). User-related challenges of self-service business intelligence. *Information Systems Management*, 38(4), 309–323.
- Li, J., Dai, J., Issakhov, A., Almojil, S. F., & Souri, A. (2021). Towards decision support systems for energy management in the smart industry and internet of things. *Computers & Industrial Engineering*, 161, Article 107671.
- Liu, M., Fang, S., Dong, H., & Xu, C. (2021). Review of digital twin about concepts, technologies, and industrial applications. *Journal of Manufacturing Systems*, 58, 346–361.
- Martínez-Olvera, C. (2020). An entropy-based formulation for the support of sustainable mass customization 4.0. *Mathematical Problems in Engineering*, 2020, 1–21.
- Michalczyk, S., Nadj, M., Azarfar, D., Maedche, A., & Gröger, C. (2020). A state-of-the-art overview and future research avenues of self-service business intelligence and analytics. In *ECIS 2020 research papers: vol. 46*.
- Muntean, M., & Surcel, T. (2013). Agile BI—the future of bi.. *Informatica Economica*, 17(3), 114–124.
- Oliva, R. (2019). Intervention as a research strategy. *Journal of Operations Management*, 65(7), 710–724.
- Phillips-Wren, G., Daly, M., & Burstein, F. (2021). Reconciling business intelligence, analytics and decision support systems: More data, deeper insight. *Decision Support Systems*, 146, Article 113560.
- Rai, A. (2017). Editor's comments: Diversity of design science research. *MIS Quarterly*, 41(1), iii–xviii.
- Saeed, K., & Sidorova, A. (2021). Democratization of corporate IT using is architecture representation framework (ISARF). In *Proceedings of the 54th hawaii international conference on system sciences*.
- Saldívar, A. A. F., Goh, C., Chen, W. n., & Li, Y. (2016). Self-organizing tool for smart design with predictive customer needs and wants to realize industry 4.0. In *2016 IEEE congress on evolutionary computation* (pp. 5317–5324). IEEE.
- Schuff, D., Corral, K., St. Louis, R. D., & Schymik, G. (2018). Enabling self-service BI: A methodology and a case study for a model management warehouse. *Information Systems Frontiers*, 20, 275–288.
- Spahn, M., Kleb, J., Grimm, S., & Scheidl, S. (2008). Supporting business intelligence by providing ontology-based end-user information self-service. In *Proceedings of the first international workshop on ontology-supported business intelligence* (pp. 1–12).
- Srivastava, G., Venkataraman, R., V, K., & N, P. (2022). A review of the state of the art in business intelligence software. *Enterprise Information Systems*, 16(1), 1–28.
- Terrazas, G., Ferry, N., & Ratchev, S. (2019). A cloud-based framework for shop floor big data management and elastic computing analytics. *Computers in Industry*, 109, 204–214.
- Tien, J. M. (2012). The next industrial revolution: Integrated services and goods. *Journal of Systems Science and Systems Engineering*, 21, 257–296.
- Tien, J. M. (2013). Big data: Unleashing information. *Journal of Systems Science and Systems Engineering*, 22, 127–151.
- Tiihonen, J., & Felfernig, A. (2017). An introduction to personalization and mass customization. *Journal of Intelligent Information Systems*, 49(1), 1–7.
- Vuppapalapati, C., Ilapakurti, A., Kedari, S., Vuppapalapati, R., Vuppapalapati, J., & Kedari, S. (2020). Democratization of AI to small scale farmers, albeit food harvesting citizen data scientists, that are at the bottom of the economic pyramid. In *International conference on intelligent human systems integration* (pp. 348–358). Springer.
- Wache, H., Dinter, B., & Kollwitz, C. (2019). Analytics use cases for mass customization—a process-based approach for systematic discovery. In *Proceedings of the 52nd hawaii international conference on system sciences*.
- Werner, R., Takacs, R., Geier, D., Becker, T., Weissenberg, N., Haße, H., Sollacher, R., Thalhofer, M., Schumm, B., & Steinke, I. (2021). The challenge of implementing digital twins in operating value chains. *Digital Twins: Applications To the Design and Optimization of Bioprocesses*, 127–166.

- Xia, T., Fang, X., Gebraeel, N., Xi, L., & Pan, E. (2019). Online analytics framework of sensor-driven prognosis and opportunistic maintenance for mass customization. *Journal of Manufacturing Science and Engineering*, 141(5), Article 051011.
- Yang, L., & Shami, A. (2022). IoT data analytics in dynamic environments: From an automated machine learning perspective. *Engineering Applications of Artificial Intelligence*, 116, Article 105366.
- Yao, Q., Wang, M., Chen, Y., Dai, W., Li, Y. F., Tu, W. W., Yang, Q., & Yu, Y. (2018). Taking human out of learning applications: A survey on automated machine learning. arXiv preprint arXiv:1810.13306.
- Yu, E., Lapouchnian, A., & Deng, S. (2013). Adapting to uncertain and evolving enterprise requirements: The case of business-driven business intelligence. In *IEEE 7th international conference on research challenges in information science* (pp. 1–12). IEEE.
- Zeng, J., & Glaister, K. W. (2018). Value creation from big data: Looking inside the black box. *Strategic Organization*, 16(2), 105–140.
- Zhong, R. Y., Xu, X., Klotz, E., & Newman, S. T. (2017). Intelligent manufacturing in the context of industry 4.0: a review. *Engineering*, 3(5), 616–630.
- Zöllner, M. A., & Huber, M. F. (2021). Benchmark and survey of automated machine learning frameworks. *Journal of Artificial Intelligence Research*, 70, 409–472.