6-18-2022

What constitutes a machine-learning-driven business model? A taxonomy of B2B start-ups with machine learning at their core

Oliver A. Vetter
Technical University of Darmstadt, oliver.vetter@tu-darmstadt.de

Felix Sebastian Hoffmann
Technical University of Darmstadt, f.hoffmann@ptw.tu-darmstadt.de

Luisa Pumplun
Technische Universität Darmstadt, pumplun@is.tu-darmstadt.de

Peter Buxmann
Technische Universität Darmstadt, buxmann@is.tu-darmstadt.de

WHAT CONSTITUTES A MACHINE-LEARNING-DRIVEN BUSINESS MODEL? A TAXONOMY OF B2B START-UPS WITH MACHINE LEARNING AT THEIR CORE

Research Paper

Oliver A. Vetter[a], oliver.vetter@tu-darmstadt.de

Felix Hoffmann[a], f.hoffmann@ptw.tu-darmstadt.de

Luisa Pumplun[a], luisa.pumplun@tu-darmstadt.de

Peter Buxmann[a], peter.buxmann@tu-darmstadt.de

[a]Technical University of Darmstadt, Darmstadt, Germany

Abstract

Artificial intelligence, specifically machine learning (ML), technologies are powerfully driving business model innovation in organizations against the backdrop of increasing digitalization. The resulting novel business models are profoundly shaped by ML, a technology that brings about unique opportunities and challenges. However, to date, little research examines what exactly constitutes these business models that use ML at their core and how they can be distinguished. Therefore, this study aims to contribute to an increased understanding of the anatomy of ML-driven business models in the business-to-business segment. To this end, we develop a taxonomy that allows researchers and practitioners to differentiate these ML-driven business models according to their characteristics along ten dimensions. Additionally, we derive archetypes of ML-driven business models through a cluster analysis based on the characteristics of 102 start-ups from the database Crunchbase. Our results are cross-industry, providing fertile soil for expansion through future investigations.

Keywords: Business Models, Machine Learning, Artificial Intelligence, Taxonomy.

1 Introduction

The rapidly advancing digitalization is leading to more and more data being collected by organizations. In 2020, 64.2 zettabytes of data were generated or replicated worldwide – an amount ten times larger than in 2012, with no end to the growth in sight (IDC, 2021). The high availability of data has also fuelled another technological trend, the growing use of machine learning (ML) to support or automate organizational processes (Jordan and Mitchell, 2015). ML is a technology that can be used to implement instances of artificial intelligence (AI) by learning patterns based on data that can be applied to make predictions (Brynjolfsson and Mitchell, 2017; Mitchell, 1997; Russell and Norvig, 2021). This learning process (i.e., training) is largely independent of human influence and is thus highly experimental in character (Amershi et al., 2019; Choudhury et al., 2021). ML fundamentally offers the potential to significantly change organizational processes and enable business models (BMs) in the business-to-business (B2B) segment that were previously inconceivable. For example, Salesforce utilizes ML in their *Einstein* solution to provide sales and marketing departments with insights and predictions to better understand their customers, drawn from past customer interactions (Salesforce, 2022). At the same time, the characteristics of ML hinder the creation of genuine business value for the organization from this technology (e.g., Burström et al., 2021). Harnessing the power of ML for an organization is therefore difficult to achieve and differs greatly from building BMs based on conventional technologies. Business model innovation (BMI) has always been a demanding and multi-faceted process, but ML exacerbates this challenge by adding another experimental component (Choudhury et al., 2021; Schneider and Spieth, 2013). Current research that might guide the development of ML-driven B2B-focused BMs, however, is in its infancy and is primarily focusing on specific use cases in dedicated domains, such as manufacturing (e.g., Burström et al., 2021).

In order to alleviate some of the complexity of ML-driven BMs and to establish a concise structure to guide researchers and practitioners in BMI, we aim to create an overarching taxonomy of B2B BMs enabled by ML technologies. Since established organizations often pursue multiple BMs whose boundaries become blurred, we focus particularly on start-ups whose core BM is still clearly discernible (Hartmann et al., 2016). More specifically, we focus on B2B start-ups due to the B2B segment's high potential to benefit from ML technologies (MIT Technology Review Insights, 2018). Therefore, in the interest of a meaningful taxonomy on B2B BMs, we exclude BMs in business-to-consumer (B2C) markets due to a variety of differences to B2B BMs (e.g., in customer approach (Iankova et al., 2019), value creation for customers (Grewal et al., 2021), or influence of innovative services (Dotzel and Shankar, 2019)). To reveal how B2B start-ups operate their organization, we ask:

Research question 1 (RQ1): *What are the characteristics of start-up B2B BMs that use ML at their core along different dimensions, and how can they be combined into an overarching taxonomy?*

Research question 2 (RQ2): *What are the ideal-typical archetypes of ML-driven B2B BMs based on recurring characteristics of start-ups that use ML at their core?*

According to Nickerson et al. (2013), taxonomies are artifacts that organize a set of objects according to their characteristics to help researchers and practitioners better comprehend complex domains. In relation to BMs, taxonomies serve to create a high-level abstraction of the BMs' essences. By creating the taxonomy and deriving corresponding archetypes through a cluster analysis, we contribute to structuring the research field of ML-driven B2B BMs. In particular, we highlight relevant dimensions and characteristics by which ML-driven BMs can be distinguished and thus provide other scholars a starting point to better define the object of organizational research and frame future studies. In addition, we provide practitioners with a clear overview and insights into archetypical BMs that they can use to develop new B2B BMs of their own in a more systematic way. In doing so, we support organizations in general, and start-ups in particular, to better recognize trends in the market, innovate their own business, and differentiate themselves from competitors.

The remainder of the paper is structured as follows: We start by describing research in the area of ML-driven BMs as well as existing taxonomies from the related field of data-driven BMs. We then discuss the methodology we adopted to develop the taxonomy and determine the archetypes of ML-driven B2B BMs. Finally, we present the resulting taxonomy and archetypes, discuss their value for theory and practice, and point to avenues for future research that can address limitations of this work.

2 Theoretical Background

In the following sections, we will present a brief overview of BM theory in the context of ML technologies and then report the current state of research on taxonomies related to ML-driven BMs.

2.1 Machine-learning-driven business models

The concept of BMs has received much attention by scholars in various literature streams such as e-commerce, strategy, or innovation management (Zott et al., 2011) and is to date considered a useful perspective for novel insights and further theory building in management literature (e.g., Lanzolla and Markides, 2021; Prescott and Filatotchev, 2021). In essence, a BM is a concept that illustrates the business logic of an organization and depicts how the organization creates and delivers value to customers as well as the associated architecture of revenue, costs, and profits (Teece, 2010). Various definitions of the term have been introduced and discussed in literature (Zott et al., 2011; Al-Debei and Avison, 2010; Casadesus-Masanell and Ricart, 2010). For the purpose of this research, we adopt the definition by Osterwalder and Pigneur (2010, p. 14), which states that "a business model describes the rationale of how an organization creates, delivers, and captures value." BMs are often conceptualized through the components that constitute them, e.g., the value proposition or the revenue stream (e.g., Al-Debei and Avison, 2010; Teece, 2010; Zott and Amit, 2010; Remane et al., 2016). Many of the BM components described in literature can be categorized in four types of components present in most BM conceptualizations (Burkhart et al., 2011): *Offering factors*, describing how the organization creates value for stakeholders; *Market factors*, detailing for whom value is created; *Internal capability factors*, describing activities and competences of the organization, and *economic factors*, including all economic-related aspects of the organization.

In information systems (IS) research, the BM concept is seen as the missing link connecting business strategy, processes, and information technology (IT) (Veit et al., 2014). In turn, the recent trends of increasing availability of relevant data and technological advances in data analysis carry the potential to profoundly change existing BMs of organizations in the future (Veit et al., 2014) by forcing them to adapt their BMs to survive against globalized competition (Hanelt et al., 2015). While most organizations will certainly benefit from data and data analytics, some BMs go one step further and utilize data as their *key* resource, eventually becoming data-driven BMs (Schüritz and Satzger, 2016). For the term *data-driven BM*, we adhere to the definition of Hartmann et al. (2016, p. 1385) as "a business model relying on data as a key resource." AI, as "the science and engineering of making intelligent machines" (McCarthy, 2007, p. 2), offers the opportunity to leverage additional potential in the context of digitalization (Dingli et al., 2021). The most popular technology to realize AI systems is ML, which uses learning algorithms to derive patterns from observed data and saves them in ML models, which in turn can be used on new data to solve problems (Russel and Norvig, 2021). As ML is the foundation of most modern AI systems (Jordan and Mitchell, 2015; Brynjolfsson and Mitchell, 2017), we use the term ML to refer to ML-based instances of AI for terminological clarity throughout this paper. According to Hahn et al. (2020), ML-driven BMs are a subgroup of the previously introduced data-driven BMs due to their reliance on data, yet can be differentiated from the latter as they rely on ML as self-improving technology to draw applicable patterns from the data. The authors thus conceptualize a BM as ML-driven if it utilizes ML technologies in at least one of its BM components. Modern ML technologies urge organizations to reshape or develop entirely new BMs (Lee et al., 2019; Wamba-Taguimdje et al., 2020), as this technology significantly differs from other digital technologies and poses new challenges for organizations alongside diverse opportunities: First, ML technologies can complement, constrain, or substitute for humans at work (Murray et al., 2021). Second, being capable of human feats such as conversation, they blur the boundary between human and machine capabilities (Schuetz and Venkatesh, 2020). Third, the data-based learning approach not only renders ML technologies more complex and thus inscrutable but can also result in unexpected outcomes (Benbya et al., 2020). Given these striking differences from other digital technologies (Benbya et al., 2021) and considering that BMs will undergo ML-induced transformation (Burström et al., 2021), we argue that there is a need to study these new BMs driven by ML technologies.

2.2 Taxonomies of data-driven business models

Taxonomies are a widely used tool to analyze and represent complex systems and their interrelationships in a structured way. Through the holistic disclosure of the components of the system and its properties, different manifestations can be classified and compared with each other (Nickerson et al., 2013). As literature on ML-driven BM taxonomies is scarce, we draw on the broader field of data-driven BMs to identify transferable aspects. Table 1 summarises the general and industry-specific taxonomies that exist to date. The illustration is based on the categorization by Dehnert et al. (2021), which we expanded to include new publications on data-driven BM taxonomies and the only taxonomy found on the subject of ML-driven BMs (marked in cursive). Among the added taxonomy papers are Woroch and Strobel (2021), who develop a data-driven BM in the context of the Internet of Things and Weking et al. (2020), who establish a taxonomy of industry 4.0 BMs enabled by the Internet of Things and smart factories among others. The taxonomy of Baecker et al. (2021) examines data-driven value creation within organizations and focuses on the required underlying data, the gained business value, and the approach to create it. Lastly, Anton et al. (2021) present the first available taxonomy of ML-driven BMs with a focus on start-ups in the energy sector. By examining how ML technologies shape BMs in this sector, they contribute to a better understanding of organizations that are implementing ML-driven BMs as the energy sector continues to transform.

	General	**Industry-specific**
Data-driven business models	*Baecker et al. (2021); Dehnert et al. (2021); Woroch and Strobel (2021);* Passlick et al. (2021); Bock and Wiener (2017); Naous et al. (2017); Hartmann et al. (2016); Engelbrecht et al. (2016); Schroeder (2016); Schüritz and Satzger (2016)	*Manufacturing: Weking et al. (2020);* Logistics: Möller et al. (2020); Manufacturing: Müller and Buliga (2019); Urban: McLoughlin et al. (2019); E- Commerce: Dorfer (2016)
ML-driven business models		*Electric Power Industry: Anton et al. (2021)*

Table 1. Classification of existing taxonomies adapted from Dehnert et al. (2021).

The preceding discussion shows that existing taxonomies cover primarily data-driven BMs. However, as mentioned in the previous section, these can not be immediately applied to ML-driven BMs, as those utilize self-improving ML technologies (Hahn et al., 2020) that affect organizations differently (see Benbya et al., 2021). Nevertheless, only Anton et al. (2021) have studied ML-driven BMs to date, albeit specifically for the electric power industry. To the best of our knowledge, there is no cross-industry and thus universally applicable taxonomy for ML-driven BMs yet – a gap we aim to bridge.

3 Methodology

To develop the taxonomy of ML-driven BMs, we utilize the development approach proposed by Nickerson et al. (2013). The method is well-accepted in IS research, having been used by several researchers for taxonomies in related fields (e.g., Anton et al., 2021; Dehnert et al., 2021; Möller et al., 2019; Remane et al., 2016). Taxonomies comprise a set of dimensions that, in turn, contain characteristics that can describe the objects under study (Nickerson et al., 2013). As a first step in the taxonomy development process, Nickerson et al. (2013, p. 343) suggest specifying a meta-characteristic as the "most comprehensive characteristic" that should reflect the purpose of the taxonomy, from which all other characteristics can be derived logically. Next, Nickerson et al. (2013) propose an iterative process to add, change, or subtract dimensions and characteristics during each iteration. The iterations can either be carried out as empirical-to-conceptual or as conceptual-to-empirical approaches. In the former, researchers analyze a subset of objects – such as real-world start-ups – to obtain their characteristics and group them into dimensions. Researchers following the conceptual-to-empirical approach conceptualize the dimensions and characteristics based on the researchers' knowledge and existing literature and then examine real-world objects to revise the taxonomy. The iterative development process ends when all predefined ending conditions are met after an iteration. After four iterations of taxonomy development, we further conduct a cluster analysis to derive archetypal BMs from the studied start-ups based on their identified characteristics.

3.1 Taxonomy development

As previously stated, the meta-characteristic defines the purpose of the taxonomy, and for this research, we determined it as *distinguishing elements of B2B-focused, ML-driven BMs*. This wording best specifies our goal to identify the different essential components to reveal the core business logic behind ML-driven BMs while at the same time discerning distinctions between different instances found in reality. We decided to adhere to the eight objective (e.g., no new dimensions and characteristics added in the last iteration, at least one object is classified under every characteristic, all objects or a representative sample of objects has been examined) and five subjective (e.g., conciseness, robustness, extendibility) ending conditions proposed by Nickerson et al. (2013).

We chose the conceptual-to-empirical approach for our first iteration. Because little research on ML-driven BMs was available, we turned to literature on general BMs as a starting point. A large volume of published works describes possible configurations of BMs (e.g., Al-Debei and Avison, 2010; Hedman and Kalling, 2003; Osterwalder and Pigneur, 2010). As the Business Model Canvas by Osterwalder and Pigneur (2010) contains the majority of BM components discussed in literature (Passlick et al., 2021) and is additionally well-regarded in practice, we chose it as a starting point for the taxonomy. We drew from literature on data-driven and ML-driven BMs (see Table 1) to select the dimensions from the Business Model Canvas promising the highest discriminatory power for ML-driven BMs as starting dimensions, namely *value proposition* (which we split into *value promise* and *key offering* in the third iteration; see Möller et al., 2019), *customer segment*, *channel*, *key resources*, *key activities*, *revenue stream*. We supplemented these dimensions with aspects from data-driven BMs that are transferable to ML-driven BMs: We specified the dimension *key resources* into the more ML-relevant dimensions *data source* and *data type* (Engelbrecht et al., 2016; Möller et al., 2020; Azkan et al., 2020; Hartmann et al., 2016). Similarly, we specified the dimension *channel* into the most relevant aspect for distinguishing ML-driven BMs: The *deployment channel* (Passlick et al., 2021).

Since the taxonomy should not only be academically motivated but also consider emerging ML-driven BMs in practice, we conducted the following three iterations in compliance with the empirical-to-conceptual approach. We employed the Crunchbase database for our data collection and searched it for suitable start-ups (Crunchbase, 2021). We focused exclusively on start-ups because the available population is larger compared to established organizations, and start-ups presumably have purer BMs that are not hampered by old legacy systems (Hartmann et al., 2016). In particular, start-ups possess only a single or small number of BMs, facilitating their analysis (Sabatier et al., 2010). Our search terms to browse Crunchbase were *machine learning* and *artificial intelligence*, which we used to search the start-ups' tags as well as their short descriptions. We included the latter search term, since start-ups often operate under the more general buzzword *artificial intelligence* when referring to ML-based AI. We wanted to find start-ups that struck a balance between being young enough to pursue a singular BM still while being old enough that we could omit organizations that went bankrupt quickly after launch. Therefore, we focused on start-ups founded in 2018 and 2019. Our search yielded a total of 2,057 start-ups. Due to the large dataset size, we followed the recommendations by Nickerson et al. (2013) and randomly selected subsamples for each iteration (see Möller et al., 2019; Möller et al., 2020). We removed and replaced all start-ups that went bankrupt, did not realize an ML-driven, B2B-focused BM, or did not have sufficient information in German or English on their homepages (see Remane et al., 2016; Täuscher and Laudien, 2018). The subsample sizes are 22 for the second and 40 for the third and fourth iteration, resulting in a data set of 102 start-ups, an excerpt of which can be found in Table 4 (see Appendix). Consistent with Hunke et al. (2019), we started with a smaller number of start-ups in the second iteration and subsequently included more entities, as we wanted to roughly identify dimensions and characteristics of ML-driven BMs first, while relying on more information to refine and elaborate the taxonomy in more detail in the later iterations. In each iteration, we analyzed the ML-driven BMs of the subsample for their characteristics. In particular, we checked whether these characteristics were consistent with the previously found taxonomy or whether additions, revisions, or deletions of characteristics or dimensions would improve the usefulness of the taxonomy. We gathered the required information about the BMs in our dataset from publicly available sources such as the start-up's website, articles, blog entries, or other online presences. Because "gross elements of business models are often quite transparent" (Teece, 2010, p. 179), a start-up's BM can be inferred using such reliable public sources (see Hartmann et al., 2016; Möller et al., 2019). We employed multi-researcher triangulation to ensure a high degree of objectivity (e.g., Hsieh and Shannon, 2005; Flick, 2004). As few start-ups disclosed data on their revenue streams or utilized ML form, we derived the information for the corresponding dimensions with the help of pertinent literature and validated them empirically with the start-ups on which data was available. The taxonomy was finalized after the fourth iteration of the development process, as all ending conditions were met.

3.2 Cluster analysis

We performed a cluster analysis on our dataset to derive information on which archetypes of ML-driven BMs commonly appear in practice. Cluster analysis seeks to form groups of objects based on their similarities (Bailey, 1994), thus in our case, assembles the start-ups into archetypal BMs based on their similarity along the dimensions of the taxonomy. Regarding the design of our cluster analysis, we followed preceding research (Remane et al., 2016; Möller et al., 2019; Anton et al., 2021) and carried out the two-step procedure of Punj and Stewart (1983). The first step consists of the agglomerative hierarchical clustering algorithm of Ward's minimum variance method (Ward, 1963). The procedure starts with every object being a separate cluster and then iteratively merging the two closest clusters based on the calculated distance between them (Eszergár-Kiss and Caesar, 2017). We used Euclidean distances as distance metric, as it is suitable for binary variables and Ward's method is well defined for Euclidian distances (see Rencher, 2002). The results of Ward's method show that either a 6 or 7 cluster solution is optimal for our dataset. In the second step, we used the k-means partitional clustering method. The k-means algorithm finds a partition that minimizes the sum of squared distances between the empirical mean of each cluster and the objects in the respective cluster for an a priori defined number of clusters (Jain, 2010). We chose the 7 cluster solution for our final archetypes because it outperformed the 6 cluster solution both on the elbow curve and the Davies-Bouldin index (Davies and Bouldin, 1979). We implemented the data preparation and the k-means clustering in RapidMiner Studio, and the Ward's method in Python using the library SciPy.

4 Results

In this section, we present our final taxonomy and the derived BM archetypes.

4.1 Final taxonomy

The final taxonomy is shown in Table 2 and consists of ten dimensions, which in turn contain three to six different characteristics. Each of the 102 start-ups in our dataset is described by at least one of the characteristics in each dimension. Following Nickerson et al. (2013), only dimensions in which ML-driven BMs differ are included in the taxonomy, as characteristics that are identical among all ML-driven BMs are of little use to a taxonomy due to their lack of discriminatory power (Anderberg, 1973). To be able to represent the large variety of BMs from our industry-overarching dataset, our dimensions allow BMs to exhibit more than one characteristic (see Hunke et al., 2019; Möller et al., 2020). The following paragraphs describe the identified dimensions and characteristics in depth.

The dimension **value promise** describes what type of value the BM creates for its clients and can take four different characteristics. Hereafter, the term *client* denotes an organization utilizing the services of an ML-driven BM. BMs characterized by the first characteristic *cost and time reduction* either replace human labor for menial tasks or assist humans in their work, thus allowing them to complete workflows more quickly or cost-efficiently. The *quality increase* characteristic denotes BMs that provide effectivity improvements for their client's products, services, or processes, usually by supplementing them with some form of intelligent behavior. These BMs aim to modify their clients' services to deliver better results or to offer additional, previously impossible or infeasible features, e.g., enriching video material with ML-generated metadata. Clients of BMs with the characteristic *insight increase* are supplied with ML-generated knowledge derived from data and designed to improve the client's decision-making process either through faster or more informed decisions. The type of use case thereby determines the content of the provided information: Performance metrics calculated through ML methods might support clients in management decisions, while ML-created risk estimations might aid real estate investors in finding trustworthy debtors in their day-to-day operations. Lastly, *innovation increase* describes BMs with ML systems aiding the client in exploring previously uncharted territory. Providers aim to improve their clients' search for innovation or novel inventions. An example from our dataset is ML supporting pharmaceutical companies in their drug discovery process by suggesting possible solutions or identifying gaps in pre-existing knowledge.

Dimensions	**Characteristics**					
Value promise	Cost and time reduction	Quality increase	Insight increase	Innovation increase		
Key offering	Aggregation & filtering	Information enrichment	Detection	Optimization	Forecasting	Generation
Client influence on ML system	No influence	Selection of settings	Feedback loop	Development of model	Ownership of model	
Customer segment	Primary sector	Secondary sector	Tertiary sector	Quaternary sector		
Key activity	Consulting	Data science	Data sourcing & engineering	Software engineering	Hardware development	
ML form	Supervised learning	Unsupervised learning	Reinforcement learning			
Deployment channel	Edge	On-premise software	Hosted software	Plug-in		
Data source	Client data	Provider data	Publicly available data			
Data type	Structured	Semi-structured	Unstructured			
Revenue model	Pay-with-data	Subscription	Pay-per-X	Gain sharing	One-time fee	

Table 2. Final taxonomy, visualized as a morphological box. To improve readability, we have removed the characteristic "unspecified."

Another critical dimension for ML-driven BMs is the **key offering**, or in other words, the type of service they provide to their clients to create the previously described added value**.** *Aggregation & filtering* organizations provide their clients with an ML solution that analyzes large amounts of data, omits irrelevant data, and condenses the essential information into meaningful output values for the client organization. An example would be a system that screens job applications and highlights each candidate's most relevant experiences for the position. Solutions with the *information enrichment* characteristic also analyze data but aim to expand the given data(set) with supplementary data. They either extend unstructured data with structured data (e.g., analyzing clinical images and displaying additional diagnostic information) or integrate information from complementary sources into the system (e.g., an ML system that crawls the social media sites of job applicants and derives their trustworthiness). Furthermore, the characteristic *detection* describes systems that continuously monitor data streams and alert the client when certain patterns or suspicious activities are detected. Prominent examples include credit fraud detection systems that raise alarms when credit cards are used irregularly or ML-based visual inspection systems that call attention to defective products. BMs with *optimization* offerings apply their ML systems to solve specific, well-defined problems that have a clear desired output but are challenging to solve with conventional methods. They may involve traditional optimization problems like scheduling or vehicle routing problems. However, the ML systems might also attempt to find the ideal candidate for a given job position or optimize the bidding process for E-commerce advertisements by only bidding when individual clients are likely to buy (minimizing costs while maximizing the likelihood of sale). As the name suggests, *forecasting* organizations offer their clients glimpses into the future. They attempt to predict future states of given dependent variables by incorporating large amounts of data. Examples include predictive maintenance solutions which aim to predict when equipment will need maintenance or renewable electricity generation forecasting, which calculates future power production by analyzing weather data. Organizations represented by *generation* offer ML systems for tasks with high degrees of freedom that use input data to independently create complex, context-specific output that resembles the solution a human might have conceived. Chatbots are a prime example of a generation ML solution, answering user queries with solutions relevant to the user's interests. Another exemplary generation system performs automated legal document generation, creating contracts based on tabular input data.

The **client influence on ML system** dimension differentiates BMs based on the extent to which they individually adapt their ML systems to satisfy their clients' needs, which is a double-edged sword. BMs can allow their clients different degrees of contribution, but the more influence clients have, the more difficult it generally becomes to scale the BM. This is because higher degrees of influence usually necessitate that ML models are (re-)trained individually for each client, which increases the effort of each sale. Possible characteristics start with clients having *no influence* at all. These clients either possess no ML knowledge of their own, do not want to draw on their ML resources, or implement a use case that can be fulfilled with a one-size-fits-all solution. With *selection of settings*, the clients still have minimal influence on the ML system but can alter certain predefined settings to cause changes in the system (e.g., a formality setting for a chatbot to adapt the system to different use cases). *Feedback loop* means that clients can evaluate the output of the ML system and feed their evaluation back into the system, which in turn learns from the additional data and corrects itself over time. To have more control over the finished system, clients might opt for a *development of model* organization. These organizations include their clients in the development process either through a joint team, regular interactions, or platforms simplifying ML development. Lastly, organizations with the *ownership of model* characteristic hand the finished ML model over to their clients, who gain full access and can analyze the model, improve, or re-train it for other use cases.

The dimension **customer segment** records the economic sector in which the target clients of the BM are allocated. According to the definition of Kenessey (1987), which is widely used today, a distinction can be made between four sectors: The *primary sector* supplies raw materials for a product and includes, among others, the harvesting of wood in forestry, fishing, or the generation of hydropower. The *secondary sector* is responsible for processing raw materials from the *primary sector*. It thus includes the manufacturing industry, craft production, and the energy industry, among others. The *tertiary sector* is comprised of all services provided by private organizations or government institutions such as transportation services, utilities, and wholesale or retail trade. The concept of the *quaternary sector* has gained in importance in the context of the transformation to an information society and subsumes all industries that deal with creating, processing, and selling information (data or knowledge). These include IT services and communications technology. Which sector a BM's clients stem from greatly influences their available data and IT infrastructure.

The dimension **key activity** describes "the most important things a company must do to make its business model work" (Osterwalder and Pigneur, 2010, p. 36). Naturally, the vast majority of ML-driven BMs require *data science* as key activity. However, some organizations get by with minimal data science activities, for example, when the applied ML models are already very mature, like it is the case for computer vision solutions. *Consulting* indicates that conveying ML-related knowledge in close contact with the clients is essential for the BM. *Data sourcing & engineering* characterizes BMs that spend much time on gathering, curating, and supplying data in order to provide their services, as it might be the case for organizations offering insights on financial markets for which they need carefully curated data. The key activity *software engineering* is assigned to organizations whose ML solutions are embedded into highly complex software that must naturally be developed and maintained. Similarly, *hardware development* describes organizations that rely on and must develop complex physical devices to execute the output of their ML system, with computer-vision-powered robots being one example.

Organizations can primarily apply three **ML form**s in their BM (Russell and Norvig, 2021), each with distinct capabilities and uses, as well as unique requirements regarding expertise and development. *Supervised learning* systems are given sets of input-output pairs and then learn a function that predicts the appropriate output, or label, when given new inputs. In *unsupervised learning*, the machine learns to find patterns in the input data without being given any explicit feedback. Lastly, in *reinforcement learning,* the system performs certain actions and is then given either rewards or punishments as feedback, which it uses to learn which actions lead to more rewards and alter its actions accordingly.

How ML-driven BMs deliver added value through their products or services generally differs between four **deployment channel**s. The channel *edge* means, that the BM's ML system is run on physical devices that are often supplied as a product package. A Chatbot that is implemented on a special tablet for direct in-store customer interaction would be an example for deployment via edge. *On-premise software,* on the other hand, denotes ML systems that are run on the client's network hardware, often on servers or in their cloud. Conversely, *hosted software* is executed on the BM's hardware (thus increasing cloud provisioning costs), with their clients gaining access through a website or APIs. Lastly, *plug-in* ML solutions integrate seamlessly into pre-existing software or platforms, with an example being a human agent augmentation plug-in for a contact center platform.

The data required for running a BM's specific ML solution can stem from three different **data source**s. *Client data* means the client either has pre-existing datasets with the information required or records new data to be used in the system. The ML models are thus either fully trained on client data or come as pre-trained models and are re-trained with client data. In contrast, BMs denoted by *provider data* sell their ML system along with their own supply of data for training and running the model. Furthermore, models can also utilize *publicly available data*, which any interested party can acquire from data platforms such as Kaggle (2021), from data vendors, or from other public sources.

ML systems can require data in many different **data type**s, which can be subsumed under three major categories (Sint et al., 2009; Abiteboul et al., 2000). *Structured data* denotes any type of data with an underlying structure, such as tabular data in a database. *Unstructured data,* on the contrary, does not exhibit an identifiable structure and includes images, video, audio, and free text. *Semi-structured data* has no separate, explicit description of its structure, yet it does demonstrate some structure within the data (e.g., e-mails consisting of subject, sender, and text). The type of data carries many implications for a BM: Each type necessitates different kinds of expertise within the BM and requires different preprocessing efforts for value creation.

The **revenue model** dimension depicts how the BM generates revenue in order to cover costs and thrive as an organization. Since many start-ups withheld information on their revenue model until after a demo or sales talk, the characteristics are based on Schüritz et al. (2017), Osterwalder and Pigneur (2010), and Hartmann et al. (2016) and validated with cases from our dataset. In the case of *pay-with-data*, there is no cash flow from the client to the provider; instead, the client gives the provider access to their data in return for the service. Said data can then either be sold by the BM or be used to re-train existing or train future ML models. The characteristic *subscription* is assigned to organizations whose clients must pay a monthly fee to gain access to their services. The subscription rate may vary depending on the service level selected, with the possibility of offering a basic version of the service free of charge in a Freemium model. *Pay-per-X* denotes revenue models where the clients pay a dynamic fee based on performance measurements. These measures can range from the amount of input or output data requested to the occupied computational resources and can also include the number of utilized billable hours. Dynamic fees are also incorporated in *gain sharing* models; however, in this model, they are directly dependent on monetary success measures of the ML system. An exemplary fee, in this case, is a commission that depends on the value of a mediated contract (e.g., between employers and employees). Lastly, in *one-time fee* models, clients pay one time for the ML system and associated services (e.g., including maintenance services for the first few years).

4.2 Business model archetypes

Our cluster analysis grouped the 102 start-ups of our dataset into seven clusters that each contained 12 to 19 start-ups. Each cluster has a centroid for every characteristic, representing the distribution of BM characteristics in the respective archetype, depicted in Table 3. As characteristics are deliberately not mutually exclusive, the percentages in each dimension do not add up to 100%. Instead, they show how many start-ups in the archetype exhibit each characteristic. We omitted the characteristic *unspecified* from the analysis as it does not contribute to an archetype's distinctness. Due to the high amount of unspecified ML forms (83%) and revenue models (75%), we consequently decided to omit these dimensions as well (see Möller et al., 2019). Table 3 also shows the consistency between the Ward's method and the k-means clustering for each archetype (see Anton et al., 2021; Möller et al., 2019). We analyzed the cluster centroids and validated them with the BMs contained in each archetype. The resulting descriptions for each archetype are presented in the following paragraphs and illustrated with archetypical examples from our dataset.

Dimensions	Characteristics	Sample distribution	Intelligence for services	Automated sensing	Robotic process automation	ML development partner	Constructive assistant	Internal business diagnostics & prediction	Environmental diagnostics & prediction
	Consistency (algorithms):		73.68%	64.29%	53.85%	66.67%	68.75%	91.67%	68.75%
	n	102	19	14	13	12	16	12	16
Value promise	Cost and time reduction	47%	5%	100%	100%	33%	81%	17%	6%
	Quality increase	20%	89%	0%	15%	0%	0%	0%	6%
	Insight increase	39%	5%	0%	15%	67%	6%	100%	100%
	Innovation increase	5%	5%	0%	8%	8%	13%	0%	0%
	Unspecified	2%	0%	0%	0%	17%	0%	0%	0%
Key offering	Aggregation & filtering	38%	16%	0%	31%	42%	19%	83%	88%
	Information enrichment	23%	32%	71%	23%	8%	0%	17%	6%
	Detection	15%	11%	21%	38%	8%	0%	0%	25%
	Optimization	18%	32%	14%	31%	33%	6%	8%	0%
	Forecasting	15%	11%	0%	15%	17%	0%	50%	19%
	Generation	24%	21%	0%	23%	8%	100%	0%	0%
	Unspecified	3%	0%	0%	0%	25%	0%	0%	0%
Client influence on ML system	No influence	63%	79%	64%	77%	25%	38%	83%	69%
	Selection of settings	17%	11%	21%	8%	17%	25%	8%	25%
	Feedback loop	3%	0%	14%	0%	0%	6%	0%	0%
	Development of model	10%	5%	0%	0%	42%	19%	0%	6%
	Ownership of model	5%	0%	0%	15%	25%	0%	0%	0%
	Unspecified	5%	5%	0%	0%	8%	13%	8%	0%
Customer segment	Primary sector	6%	0%	29%	0%	0%	0%	8%	6%
	Secondary sector	14%	11%	0%	54%	0%	6%	8%	19%
	Tertiary sector	38%	53%	14%	46%	0%	50%	67%	31%
	Quaternary sector	13%	26%	21%	8%	0%	0%	8%	19%
	Unspecified	33%	21%	36%	0%	100%	44%	8%	31%
Key activity	Consulting	7%	5%	0%	0%	42%	6%	0%	0%
	Data science	84%	95%	79%	85%	75%	88%	58%	100%
	Data sourcing & engineering	26%	16%	7%	62%	0%	13%	0%	81%
	Software engineering	40%	21%	50%	38%	25%	69%	83%	6%
	Hardware development	3%	0%	21%	0%	0%	0%	0%	0%
Deployment channel	Edge	10%	5%	36%	15%	0%	6%	0%	6%
	On-premise software	22%	37%	29%	23%	25%	6%	25%	6%
	Hosted software	47%	26%	29%	46%	33%	69%	67%	63%
	Plug-in	15%	11%	7%	15%	0%	44%	0%	19%
	Unspecified	16%	21%	7%	15%	42%	6%	8%	13%
Data source	Client data	74%	53%	86%	77%	92%	75%	92%	56%
	Provider data	17%	26%	14%	15%	8%	25%	8%	13%
	Publicly available data	25%	21%	0%	23%	17%	6%	25%	75%
	Unspecified	4%	11%	0%	0%	8%	6%	0%	0%
Data type	Structured	26%	11%	0%	54%	33%	0%	92%	19%
	Semi-structured	4%	0%	0%	0%	8%	6%	8%	6%
	Unstructured	49%	79%	100%	8%	0%	75%	0%	50%
	Unspecified	23%	16%	0%	38%	58%	19%	0%	31%

0%

Distribution of characteristics

100%

Table 3. Distribution of the start-ups' characteristics within each archetype.

Intelligence for services: Start-ups in this archetype provide opportunities to integrate ML-enabled functionalities and intelligent behavior into their clients' services. Clients thus benefit through their own services achieving superior results. Due to the focus on ML-driven enhancement of services, clients of this archetype typically stem from the tertiary or quaternary sector. The most common key offerings, information enrichment and optimization, are each used by 32% of start-ups in this archetype to improve the services of clients – however, which key offering a BM chooses largely depends on its client's services, with cases of all key offerings existing in the sample. The majority of ML solutions in this archetype use unstructured data, with 53% of start-ups utilizing their client's data, 26% supplying their own data alongside their ML system, and 21% using publicly available data. An example for start-ups of this archetype is Bidnamic (2021), whose ML system supports retailers selling via search engine by calculating optimal prices for each individual product and search term.

Automated sensing: This archetype contains BMs with ML systems that can interpret unstructured data quicker or cheaper than humans can. Often, their solutions are computer vision systems that might be deployed on a physical visual sensor, recording their surroundings and analyzing the gathered data. The conducted analysis depends on the key offering, with most systems (71%) focusing on information enrichment – extracting additional information from the data, and passing it on for further use. If BMs deploy their ML systems on complex edge devices, they must perform the necessary hardware development as key activity (21%). Not all start-ups in this cluster are computer-vision-based, but they all process a type of unstructured data. EAIGLE (2021), as archetypical BM, provides a computer vision solution for automated visitor sign-in, including visitor health screening.

Robotic process automation: Robotic process automation uses software to imitate repetitive tasks that would otherwise be carried out by humans (Santos et al., 2020). Start-ups in this archetype automate routine workflows of their clients with ML systems to achieve cost and time reductions. The provided key offering depends primarily on the respective business process being automated, with detection (38%), optimization (31%), and aggregation & filtering (31%) being most common in our sample. In 54% of our cases, the data required for automating the workflows are structured data. Additionally, 62% of start-ups in this archetype perform extensive data sourcing & engineering tasks to reduce human labor as much as possible. In our dataset, start-ups in this archetype mainly serve the secondary and tertiary sectors. One example of such start-ups is Circuit Mind (2021), whose ML system automatically selects components and generates possible circuit schematics for electronics.

ML development partner: ML development partners work towards providing their clients with user-friendly access to the technology of ML. They can offer access to a variety of ML services and aid in developing ML systems that are specifically tailored to each client and their data. To achieve this high degree of individuality, BMs either provide a platform that does not require extensive ML-specific expertise for clients to develop their own ML models, or BMs conduct a collaborative development process. In the latter case, the start-up is often engaged in extensive consulting activities (42%) and can grant its clients different degrees of influence on the ML system as desired. Due to the focus on providing the technology without a specific business context, the target group of these start-ups is not limited to a specific sector. An archetypical example is AltaML (2021), with their ML experts working together with clients to realize ML solutions for various use cases.

Constructive assistant: All organizations in this BM archetype offer generation ML solutions (100%), most of which are designed to achieve cost and time reductions for client organizations (81%). These ML systems are usually given instructions through unstructured data (75%) and then aim to create output in an unstructured data format that emulates the way a human would have completed the task. Chatbots are a prominent example of such solutions. Many use cases for these ML systems exist in cross-sectional functions or functions with contact to end consumers. Consequently, they are primarily offered to clients in the tertiary sector (50%) or are not sector-specific. Lastly, users of constructive assistant solutions should not require experience with ML systems, so extensive integration efforts into comprehensive software systems through software engineering are often necessary (69%). These ML systems are delivered as plug-ins for software of other providers in 44% of cases. Scissero (2021) is an exemplary start-up, with ML software supporting legal departments by analyzing or suggesting drafts of legal documents.

Internal business diagnostics & prediction: All start-ups of this archetype aim to support their clients' decision-making processes by supplying relevant information about the respective client's internal business activities. A software package that combines an IS with ML capabilities analyzes the client's internal data and extracts essential facts for decision-makers. 50% of start-ups in the archetype additionally forecast measures to reduce uncertainty in decisions. The archetype focuses mainly on clients in the tertiary sector, and its ML systems primarily utilize structured (92%) client data available on internal processes (92%). Start-ups of this archetype integrate their ML technologies into ISs, so software engineering is often a key activity (82%). As an archetypical example, CognitOps (2021) utilizes client data to assist warehouse managers with operational decisions, e.g., in scheduling.

Environmental diagnostics & prediction: Start-ups in this category also assist their clients' decision-making processes through insight increases – usually through an aggregation & filtering key offering (88%) as well. However, unlike the archetype *internal business diagnostics & prediction*, they provide information on elements external to the client organization, such as financial markets or public opinion. This seemingly small distinction has further implications for BMs, distinguishing this archetype: Due to the high relevance of publicly available data (75%), sourcing data and setting up pipelines to provide curated data for the ML systems is essential for 82% of the BMs. In 50% of cases, the incoming data is also unstructured, requiring interpretation. Start-ups in this archetype slightly favor targeting the tertiary sector (31%), or all of them with an ML solution for a cross-sectional function. Nevertheless, there are some with a focus on the secondary or quaternary sector, making this a widely diffused archetype. The example for this archetype, Sanctify Financial Technologies (2021), aids in asset management decisions by analyzing non-financial news articles on potential investments.

5 Conclusion, Limitations & Outlook

To examine how ML technologies influence BMs of organizations, we use the methodical approach of Nickerson et al. (2013) to develop an industry-overarching taxonomy of ML-driven B2B BMs by analyzing a sample of 102 start-ups. The taxonomy describes possible characteristics of start-ups along ten dimensions and addresses **RQ 1**. Moreover, we aggregate the start-ups into seven BM archetypes that represent designs commonly found in practice and thus address **RQ 2**.

Our study provides several **theoretical contributions**. Literature has been calling for further taxonomic research related to data-driven BMs (e.g., Veit et al., 2014; Müller and Buliga, 2019; Omerovic et al., 2020). With ML giving rise to opportunities and challenges unique to the technology (Benbya et al., 2021) when transforming BMs (Björkdahl, 2020; Burström et al., 2021), we extend the discussion to ML-driven BMs. In this field, the presented taxonomy with the described dimensions and characteristics fosters a deeper understanding of the anatomy of ML-driven BMs. It allows researchers to specify ML-driven BMs in a unified manner and distinguish them from each other. By standardizing the vocabulary in this topic area, we facilitate the scientific exchange between researchers and future work in the context of ML-driven BMs. Using a common language, new ideas can be objectified, and considerations can be shared among scientists to build a more profound theoretical understanding of ML-driven BMs. This allows further systematization of research in this field. Furthermore, the presented artifacts provide a basis for future taxonomic research. Researchers can validate or extend them for narrower scopes like specific application domains (see Anton et al., 2021), specifying and extending the dimensions and characteristics. Another contribution of our paper is the method-based identification of BM clusters and the subsequent derivation of archetypes of ML-driven BMs. The seven derived archetypes provide deeper insight into the structural composition of commonly implemented ML-driven BMs. Additionally, our research also offers **practical contributions**, of which we focus on two: First, the taxonomy in combination with the archetypes of BMs offers a comprehensive overview of the market. Organizations and other stakeholders can benefit from this increased understanding by improving investment decisions or assessing their own as well as their competitors' BMs. Second, practitioners can further benefit from our artifacts by using them as supporting tools for the conception of novel ML-driven BMs through the structured recombination of BM components (Bouwman et al., 2020), therefore employing our research results to facilitate BMI.

As with every research, our study is subject to certain **limitations**. Despite the previously mentioned advantages, our focus on start-ups as a data basis excludes established organizations, leading to several consequences: In particular, the identified archetypes can only be transferred to established organizations to a limited extent, as these may already have higher resources and established structures in place and can offer different services accordingly. Further, the developed taxonomy may need to be adapted and expanded to reflect the specifics of established organizations, e.g., in terms of the key activities undertaken. Even though this would be a valuable line of further research, we are confident that the taxonomy stemming from the start-ups is broadly applicable to more mature organizations as well, as we have intentionally abstracted from specifics of start-ups within the taxonomy and kept dimensions and characteristics rather generic. Moreover, despite the methodical approach followed, many steps in the taxonomy development process required the researchers' own judgment (Nickerson et al., 2013). Therefore, another group of researchers might encounter different dimensions and characteristics from this study. Similarly, archetypes may vary based on the chosen number of clusters, which in turn depends on the employed algorithm (see Mojena, 1977). Our dataset includes a large variety of ML-driven BMs as per our purpose. However, aggregating them into a manageable number of distinct yet general archetypes is challenging, which can be seen in the consistency between algorithms of some clusters (see Table 3). Nevertheless, as taxonomies and archetypes are never perfect, they should instead be assessed based on whether they are useful (Nickerson et al., 2013; Remane et al., 2016); a quality revealed when researchers and practitioners start using them.

Acknowledgment

This research and development project is/was funded by the German Federal Ministry of Education and Research (BMBF) within the "Innovations for Tomorrow's Production, Services, and Work" Program (funding number 02L19C150) and implemented by the Project Management Agency Karlsruhe (PTKA). The authors are responsible for the content of this publication.

Appendix

Start-up	Website	Start-up	Website
Acquired Insights	www.aiinc.cloud	Alectio	www.alectio.com
Animatico	animati.co	Arva Intelligence	www.arvaintelligence.com
Arytic	arytic.com	Bevov	www.bevov.com
Blyng	blyng.io	Clinicgram	www.clinicgram.com
Cordian	www.cordian.com	CropSafe	www.cropsafe.io
DeepHow	www.deephow.com	DeepRisk.ai	deeprisk.ai
Donna	donna.legal	Edgematrix	edgematrix.com
edisn.ai	edisn.ai	Eiffo Analytics	www.eiffo-analytics.com
FACTIC	www.factic-sf.com	Frontier Medicines	frontiermeds.com
Gamyte	www.gamyte.com	Hasty	hasty.ai
Inlet Laboratories	inletlabs.com	Intuaition	www.intuaition.com
LabVoice	www.labvoice.ai	Logmind	logmind.com
Mapxus	www.mapxus.com	Myst AI	www.myst.ai
Nanochomp	www.nanochomp.com	Nucleus Cyber	nucleuscyber.com
Orbem	orbem.ai	Pixofarm	pixofarm.com
RailVision Analytics	www.railvision.ca	REIGO Investments	www.reigo-inv.com
Salesken	salesken.ai	Swiftlane	swiftlane.com
ThreatLandscape	threatlandscape.com	Traverse	www.traverse.ai
uman.ai	uman.ai	Uservision	www.user.vision
Virtual Facility	www.virtualfacility.ai	Xelera Technologies	xelera.io

Table 4. Excerpt of the examined sample set of B2B ML-driven start-ups.

References

Abiteboul, S., Buneman, P., and Suciu, D. (2000). *Data on the web. From relations to semistructured data and XML.* San Francisco, CA: Morgan Kaufmann.

Al-Debei, M. M. and Avison, D. (2010). "Developing a unified framework of the business model concept," *European Journal of Information Systems* 19 (3), 359-376.

AltaML (2021). *Elevating business through AI*. URL: https://www.altaml.com/ (visited on November 16, 2021).

Anderberg, M. R. (1973). *Cluster analysis for applications. Probability and mathematical statistics: A series of monographs and textbooks*. New York: Academic Press.

Amershi, S., Begel, A., Bird, C., DeLine, R., Gall, H., Kamar, E., Nagappan, N., Nushi, B., and Zimmermann, T. (2019). "Software engineering for machine learning: A case study," in: *Proceedings of the 41st IEEE/ACM International Conference on Software Engineering: Software engineering in practice (ICSE-SEIP 2019)*. Montreal, QC, Canada, 291-300.

Anton, E., Oesterreich, T. D., Schuir, J., Protz, L., and Teuteberg, F. (2021). "A business model taxonomy for start-ups in the electric power industry – the electrifying effect of artificial intelligence on business model innovation," *International Journal of Innovation and Technology Management* 18 (03), 2150004.

Azkan, C., Iggena, L., Gür, I., Möller, F., and Otto, B. (2020). "A taxonomy for data driven services in manufacturing industries," in: *Proceedings of the Pacific Asia Conference on Information Systems (PACIS 2020)*. Dubai, UAE.

Baecker, J., Böttcher, T. P., and Weking, J. (2021). "How companies create value from data – a taxonomy on data, approaches, and resulting business value," in: *Proceedings of the 29th European Conference on Information Systems (ECIS 2021)*. Marrakech, Morocco.

Bailey, K. D. (1994). *Typologies and taxonomies. An introduction to classification techniques.* Thousand Oaks, CA: Sage.

Benbya, H., Davenport, T. H., and Pachidi, S. (2020). "Artificial intelligence in organizations: Current state and future opportunities," *MIS Quarterly Executive* 19 (4).

Benbya, H., Pachidi, S., and Jarvenpaa, S. L. (2021). "Special issue editorial: Artificial intelligence in organizations: Implications for information systems research," *Journal of the Association for Information Systems* 22 (2), 281-303.

Bidnamic (2022). *Google shopping management software*. URL: https://www.bidnamic.com/ (visited on March 22, 2022).

Björkdahl, J. (2020). "Strategies for digitalization in manufacturing firms," *California Management Review* 62 (4), 17-36.

Bock, M. and Wiener, M. (2017). "Towards a taxonomy of digital business models – conceptual dimensions and empirical illustrations," in: *Proceedings of the 38th International Conference on Information Systems (ICIS 2017)*. Seoul, South Korea.

Bouwman, H., Reuver, M. de, Heikkilä, M., and Fielt, E. (2020). "Business model tooling: Where research and practice meet," *Electronic Markets* 30 (3), 413-419.

Brynjolfsson, E. and Mitchell, T. (2017). "What can machine learning do? Workforce implications," *Science* 358 (6370), 1530-1534.

Burkhart, T., Krumeich, J., Werth, D., and and Loos, P. (2011). "Analyzing the business model concept – a comprehensive classification of literature," in: *Proceedings of the 32nd International Conference on Information Systems (ICIS 2011)*. Shanghai, China.

Burström, T., Parida, V., Lahti, T., and Wincent, J. (2021). "AI-enabled business-model innovation and transformation in industrial ecosystems: A framework, model and outline for further research," *Journal of Business Research* 127, 85-95.

Casadesus-Masanell, R. and Ricart, J. E. (2010). "From strategy to business models and onto tactics," *Long Range Planning* 43 (2-3), 195-215.

Choudhury, P., Allen, R. T., and Endres, M. G. (2021). “Machine learning for pattern discovery in management research,” *Strategic Management Journal* 42 (1), 30-57.

Circuit Mind (2021). *AI electronics design software*. URL: https://www.circuitmind.io/ (visited on November 16, 2021).

CognitOps (2021). *Automate warehouse management*. URL: https://cognitops.com/ (visited on November 16, 2021).

Crunchbase (2021). Search less. Close more. URL: https://www.crunchbase.com/ (visited on March 13, 2021).

Davies, D. L. and Bouldin, D. W. (1979). “A cluster separation measure,” *IEEE Transactions on Pattern Analysis and Machine Intelligence* PAMI-1 (2), 224-227.

Dehnert, M., Gleiss, A., and Reiss, F. (2021). “What makes a data-driven business model? A consolidated taxonomy,” in: *Proceedings of the 29th European Conference on Information Systems (ECIS 2021)*. Marrakech, Morocco.

Dingli, A., F. Haddod and C. Klüver (eds.) (2021). *Artificial intelligence in industry 4.0. A collection of innovative research case-studies that are reworking the way we look at industry 4.0 thanks to artificial intelligence.* Cham: Springer.

Dorfer, L. (2016). “Datenzentrische Geschäftsmodelle als neuer Geschäftsmodelltypus in der Electronic-Business-Forschung: Konzeptionelle Bezugspunkte, Klassifikation und Geschäftsmodellarchitektur,” *Schmalenbachs Zeitschrift für betriebswirtschaftliche Forschung* 68 (3), 307-369.

Dotzel, T. and Shankar, V. (2019). “The relative effects of business-to-business (vs. business-to-consumer) service innovations on firm value and firm risk: An empirical analysis,” *Journal of Marketing* 83 (5), 133-152.

EAIGLE (2021). *Visitor management | Wellness screening*. URL: https://www.eaigle.com/ (visited on March 16, 2022).

Engelbrecht, A., Gerlach, J., and Widjaja, T. (2016). “Understanding the anatomy of data-driven business models-towards an empirical taxonomy,” in: *Proceedings of the 24th European Conference on Information Systems (ECIS 2016)*. Istanbul, Turkey.

Eszergár-Kiss, D. and Caesar, B. (2017). “Definition of user groups applying Ward’s method,” *Transportation Research Procedia* 22, 25-34.

Flick, U. (2004). *Triangulation. Eine Einführung.* 1st Edition. Wiesbaden: VS Verl. für Sozialwiss.

Grewal, D., Guha, A., Satornino, C. B., and Schweiger, E. B. (2021). “Artificial intelligence: The light and the darkness,” *Journal of Business Research* 136, 229-236.

Hahn, C., Traunecker, T., Niever, M., and Basedow, G. N. (2020). “Exploring AI-driven business models: Conceptualization and expectations in the machinery industry,” in: *2020 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM 2020)*, 567-570.

Hanelt, A., Hildebrandt, B., and Polier, J. (2015). “Uncovering the role of IS in business model innovation - a taxonomy-driven approach to structure the field,” in: *Proceedings of the 23rd European Conference on Information Systems (ECIS 2015)*. Münster, Germany.

Hartmann, P. M., Zaki, M., Feldmann, N., and Neely, A. (2016). “Capturing value from big data – a taxonomy of data-driven business models used by start-up firms,” *International Journal of Operations & Production Management* 36 (10), 1382-1406.

Hedman, J. and Kalling, T. (2003). “The business model concept: theoretical underpinnings and empirical illustrations,” *European Journal of Information Systems* 12 (1), 49-59.

Hsieh, H.-F. and Shannon, S. E. (2005). “Three approaches to qualitative content analysis,” *Qualitative Health Research* 15 (9), 1277-1288.

Hunke, F., Engel, C. T., Schüritz, R., and Ebel, P. (2019). “Understanding the anatomy of analytics-based services-a taxonomy to conceptualize the use of data and analytics in services,” in: *Proceedings of the 27th European Conference on Information Systems (ECIS 2019)*. Stockholm & Uppsala, Sweden.

Iankova, S., Davies, I., Archer-Brown, C., Marder, B., and Yau, A. (2019). "A comparison of social media marketing between B2B, B2C and mixed business models," *Industrial Marketing Management* 81, 169-179.
IDC (2021). *Data creation and replication will grow at a faster rate than installed storage capacity, according to the IDC global DataSphere and StorageSphere forecasts*. URL: https://www.idc.com/getdoc.jsp?containerId=prUS47560321 (visited on November 2, 2021).
Jain, A. K. (2010). "Data clustering: 50 years beyond k-means," *Pattern Recognition Letters* 31 (8), 651-666.
Jordan, M. I. and Mitchell, T. M. (2015). "Machine learning: Trends, perspectives, and prospects," *Science* 349 (6245), 255-260.
Kaggle (2021). *Your machine learning and data science community*. URL: https://www.kaggle.com/ (visited on November 16, 2021).
Kenessey, Z. (1987). "The primary, secondary, tertiary and quaternary sectors of the economy," *Review of Income and Wealth* 33 (4), 359-385.
Lanzolla, G. and Markides, C. (2021). "A business model view of strategy," *Journal of Management Studies* 58 (2), 540-553.
Lee, J., Suh, T., Roy, D., and Baucus, M. (2019). "Emerging technology and business model innovation: The case of artificial intelligence," *Journal of Open Innovation: Technology, Market, and Complexity* 5 (3), 44.
McCarthy, J. (2007). *What is artificial intelligence?* URL: http://jmc.stanford.edu/articles/whatisai.html (visited on March 21, 2022).
McLoughlin, S., Puvvala, A., Maccani, G., and Donnellan, B. (2019). "A framework for understanding & classifying urban data business models," in: *Proceedings of the 52nd Hawaii International Conference on System Sciences (HICSS 2019)*. Maui, Hawaii, USA.
MIT Technology Review Insights (2018). *Professional services firms see huge potential in machine learning*. URL: https://www.technologyreview.com/2018/11/02/139216/professional-services-firms-see-huge-potential-in-machine-learning/ (visited on November 14, 2021).
Mitchell, T. M. (1997). *Machine learning.* New York, NY: McGraw-Hill.
Mojena, R. (1977). "Hierarchical grouping methods and stopping rules: An evaluation," *The Computer Journal* 20 (4), 359-363.
Möller, F., Bauhaus, H., Hoffmann, C., Niess, C., and Otto, B. (2019). "Archetypes of digital business models in logistics start-ups," in: *Proceedings of the 27th European Conference on Information Systems (ECIS 2019)*. Stockholm & Uppsala, Sweden.
Möller, F., Stachon, M., Hoffmann, C., Bauhaus, H., and Otto, B. (2020). "Data-driven business models in logistics: A taxonomy of optimization and visibility services," in: *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020)*. Maui, Hawaii, USA.
Müller, J. and Buliga, O. (2019). "Archetypes for data-driven business models for manufacturing companies in industry 4.0," in: *Proceedings of the 40th International Conference on Information Systems (ICIS 2019)*. Munich, Germany.
Murray, A., Rhymer, J., and Sirmon, D. G. (2021). "Humans and technology: Forms of conjoined agency in organizations," *Academy of Management Review* 46 (3), 552-571.
Naous, D., Schwarz, J., and Legner, C. (2017). "Analytics as a service: Cloud computing and the transformation of business analytics business models and ecosystems," in: *Proceedings of the 25th European Conference on Information Systems (ECIS 2017)*. Guimarães, Portugal.
Nickerson, R. C., Varshney, U., and Muntermann, J. (2013). "A method for taxonomy development and its application in information systems," *European Journal of Information Systems* 22 (3), 336-359.
Omerovic, M., Islam, N., and Buxmann, P. (2020). "Unlashing the next wave of business models in the internet of things era: New directions for a research agenda based on a systematic literature review," in: *Proceedings of the 53rd Hawaii International Conference on System Sciences (HICSS 2020)*. Maui, Hawaii, USA.
Osterwalder, A. and Pigneur, Y. (2010). *Business model generation: A handbook for visionaries, game changers, and challengers.* Hoboken, NJ: John Wiley & Sons.

Passlick, J., Dreyer, S., Olivotti, D., Grützner, L., Eilers, D., and Breitner, M. H. (2021). "Predictive maintenance as an internet of things enabled business model: A taxonomy," *Electronic Markets* 31, 67-87.

Prescott, J. E. and Filatotchev, I. (2021). "The business model phenomenon: Towards theoretical relevance," *Journal of Management Studies* 58 (2), 517-527.

Punj, G. and Stewart, D. W. (1983). "Cluster analysis in marketing research: Review and suggestions for application," *Journal of Marketing Research* 20 (2), 134-148.

Remane, G., Nickerson, R. C., Hanelt, A., Tesch, J. F., and Kolbe, L. M. (2016). "A taxonomy of carsharing business models," in: *Proceedings of the 37th International Conference on Information Systems (ICIS 2016)*. Dublin, Ireland.

Rencher, A. C. (2002). *Methods of multivariate analysis.* 2nd Edition. New York: Wiley.

Russell, S. J. and Norvig, P. (2021). *Artificial intelligence. A modern approach.* 4th Edition, global Edition. Harlow: Pearson.

Sabatier, V., Mangematin, V., and Rousselle, T. (2010). "From recipe to dinner: Business model portfolios in the european biopharmaceutical industry," *Long Range Planning* 43 (2-3), 431-447.

Salesforce (2022). *Artificial intelligence technology and resources: Salesforce Einstein*. URL: https://www.salesforce.com/products/einstein/overview/?d=cta-body-promo-90 (visited on March 11, 2022).

Sanctify Financial Technologies (2021). *Taking financial forecasting into the future*. URL: https://sanctify.ai/ (visited on November 16, 2021).

Santos, F., Pereira, R., and Vasconcelos, J. B. (2020). "Toward robotic process automation implementation: An end-to-end perspective," *Business Process Management Journal* 26 (2), 405-420.

Schneider, S. and Spieth, P. (2013). "Business model innovation: Towards an integrated future research agenda," *International Journal of Innovation Management* 17 (01), 1340001.

Schroeder, R. (2016). "Big data business models: Challenges and opportunities," *Cogent Social Sciences* 2 (1), 1166924.

Schuetz, S. and Venkatesh, V. (2020). "Research perspectives: The rise of human machines: How cognitive computing systems challenge assumptions of user-system interaction," *Journal of the Association for Information Systems* 21 (2), 460-482.

Schüritz, R. and Satzger, G. (2016). "Patterns of data-infused business model innovation," in: *Proceedings of the 18th IEEE Conference on Business Informatics (CBI 2016)*. Paris, France, 133-142.

Schüritz, R., Seebacher, S., and Dorner, R. (2017). "Capturing value from data: Revenue models for data-driven services," in: *Proceedings of the 50th Hawaii International Conference on System Sciences (HICSS 2017)*. Waikoloa Village, Hawaii, USA.

Scissero (2021). *Unlock data, save time and cut legal costs with Scissero*. URL: https://www.scissero.com/home (visited on November 16, 2021).

Sint, R., Schaffert, S., Stroka, S., and Ferstl, R. (2009). "Combining unstructured, fully structured and semi-structured information in semantic wikis," in: *Proceedings of the 4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009)*. Hersonissos, Greece, 73-87.

Täuscher, K. and Laudien, S. M. (2018). "Understanding platform business models: A mixed methods study of marketplaces," *European Management Journal* 36 (3), 319-329.

Teece, D. J. (2010). "Business models, business strategy and innovation," *Long Range Planning* 43 (2-3), 172-194.

Veit, D., Clemons, E., Benlian, A., Buxmann, P., Hess, T., Kundisch, D., Leimeister, J. M., Loos, P., and Spann, M. (2014). "Business models," *Business & Information Systems Engineering* 6 (1), 45-53.

Wamba-Taguimdje, S.-L., Fosso Wamba, S., Kala Kamdjoug, J. R., and Tchatchouang Wanko, C. E. (2020). "Influence of artificial intelligence (AI) on firm performance: The business value of AI-based transformation projects," *Business Process Management Journal* 26 (7), 1893-1924.

Ward, J. H. (1963). "Hierarchical grouping to optimize an objective function," *Journal of the American Statistical Association* 58 (301), 236-244.

Weking, J., Stöcker, M., Kowalkiewicz, M., Böhm, M., and Krcmar, H. (2020). "Leveraging industry 4.0 – a business model pattern framework," *International Journal of Production Economics* 225, 107588.

Woroch, R. and Strobel, G. (2021). "Understanding value creation in digital companies – a taxonomy of IoT enabled business models," in: *Proceedings of the 29th European Conference on Information Systems (ECIS 2021)*. Marrakech, Morocco.

Zott, C. and Amit, R. (2010). "Business model design: An activity system perspective," *Long Range Planning* 43 (2-3), 216-226.

Zott, C., Amit, R., and Massa, L. (2011). "The business model: Recent developments and future research," *Journal of Management* 37 (4), 1019-1042.