EMPREENDEDORISMO, STARTUPS E INOVAÇÃO

Big Data Analytics in the Business Networks: An Analyses under the Industry 4.0 Perspectives

Abstract. This research study intends in an exploratory way to understand the operational management of big data analytics technology in a company centered on Organizational Theory. This approach seeks to fill the gap in the literature on the subject and present management mechanisms that can guide the use of big data analytics in a company. The use of the industry 4.0 segment intends to enhance the study developed since in this manufacturing model there is a large use of digital technologies which produces huge amounts of data that are the base raw material of big data analytics technology. The conclusion of the study showed that the management of big data analytics from an organizational perspective leads to two complementary types of employment. As the main contribution to organizational theory, the study allowed to systematize the operational management of the big data analytics technology in manufacturing companies of industry 4.0.

Keywords: big data analytics, business network, industry 4.0, operational management.

1 Introduction

Economic and financial globalization and recent scientific and technological advances have produced a new business order with increased competitiveness between nations and companies on a world scale. The existence of new business logic, in addition to reducing the presence of companies competing in isolation, is imposing the creation and development of alliances between partners forming business networks (Gulati, Nohria & Zaheer, 2000) in this highly unstable and transformative context.

In this scenario, the growing use of digital technologies by companies, organizations, and people has intensified the current traffic of data, information, and communication, whether in the World Wide Web (WWW) - Internet, or even in corporate networks, with important consequences in inter-company relationships (Gligor, Pillai & Golgeci, 2021). The existence of an immense, varied and complex database available in these networks gives meaning to the term big data (Wang *et al.,* 2016; Papadopoulos *et al.,* 2017), but the Big Data Analytics (BDA) technology goes further and its architecture allows value extraction from extraordinarily large volumes of data (Mikalef *et al.*, 2019) in almost instantaneous time (Papadopoulos *et al.,* 2017).

In this sense, the strategic relevance of companies established in business networks, particularly in competitive environments, lies in their capacity for synergy of efforts between them (Balestrin & Verschoore, 2020). This condition points to the possibility of the existence of conformity, whether in the planning or in the use of a digital technology with a holistic concept (Ghasemaghaei & Calic, 2020) and with the characteristics of the BDA. Thus, the choice of this study for an approach in Industry 4.0 that has a multidisciplinary nature (Piccarozzi, Aquilani & Gatti, 2018) seeks to fully meet the two main characteristics of this industrial model: the existence of intense data traffic produced by the various digital technologies and machines of the manufacturing, and digitally integrated production and business environment in real time (Wang *et al.*, 2016).

Academic studies, as those cited in this research, have been developed to explore the uses of the BDA technology in cybersecurity (Leenen & Meyer, 2021), in data management and storage (Wamba *et al.*, 2017) and several authors have noticed an important gap in conducting research involving a perspective centered on organizational theory (Sheng, Amankwah-Amoah & Wang, 2017; Wamba *et al.*, 2017). They support the need to develop a BDA employment doctrine (Camargo Fiorini *et al.,*2018), including in the manufacturing segment (Dubey *et al.,*2019), as they understand this digital technology as an organizational challenge (Mikalef *et al.,* 2019). The main contribution of this study is the proposal to systematize, based on organizational theory, the operational management of the BDA technology in manufacturing companies. The business networks perspective was chosen because this field of study is deeply rooted in organizational theory (Humphrey *et al.,* 2020).

The objectives of this research study are to characterize the main elements that form business networks, present an overview of the data networks used in their relationships and propose mechanisms for operational management of the BDA technology by manufacturing companies in Industry 4.0. All of this with the purpose of contributing to the establishment of an adequate management of this modern technology in companies.

This work develops, in an exploratory way, on the development of an operational management model that can be applied to the digital BDA technology employed by manufacturing companies in Industry 4.0 and is organized as follows:

introduction; theoretical foundation with the presentation of organizational theory and existing content in the BDA literature from the perspective of business networks; discussions and methodology. Section 5 presents the results found in this research in the form of a BDA employment model for companies in the Industry 4.0 business network. The work ends with a conclusion.

2 Theoretical Foundation

To fully achieve the objectives of this research, the theoretical basis needs to reveal the fundamental elements of organizational theory and BDA literature. In this line, the perspective of business networks plays the role of link of convergence between the two approaches.

2.1 Organizational theory - perspective of business networks

In organizational theory, some authors consider that the network is the way to organize economic activities through coordination and cooperation between companies (Grandori & Soda, 1995). Business networks can be understood as a set of two or more organizations that develop business relationships with each other and that have market relationships with suppliers, customers, and competitors (Lu *et al.,* 2020). Other authors consider business networks as the participation of companies in multiple relationships (Leick & Gretzinger, 2020) and also the complex interaction with suppliers, buyers, and other actors encompassing many companies and business relationships (Hedvall, 2020).

Industry 4.0 must be understood as a new paradigm of production with a very broad domain (Piccarozzi, Aquilani & Gatti, 2018). Business networks formed by Industry 4.0 organizations comprise an integration of suppliers, manufacturers and customers in a Cyber-Physical System (CPS) composed of robots, intelligent machines, mechatronics and mobile devices, among other elements (Saniuk, Saniuk & Cagáňová, 2021) in environments of digital trust (Mubarak & Petraite, 2020). The concept of network production, mass, and customized provides an abundant amount of information about the manufacturing process itself to all companies in the network (Preuveneers, Joosen & Ilie-Zudor, 2018) and allows the exchange of information in real time about the resources used (Saniuk, Saniuk & Cagáňová, 2021). This condition becomes attractive for the use of Big Data technology, seeking to improve the efficiency and profitability of companies (Saniuk, Saniuk & Cagáňová, 2021), notably suppliers and customers.

These cyber industry networks (Saniuk, Saniuk & Cagáňová, 2021) facilitate business by bringing together companies connected via the Internet and through intelligent resources (Grabowska, 2020), providing benefits as information sharing, increased transparency and visibility of operations, optimization of manufacturing systems, in terms of energy and resource consumption, and reduction of operational and production costs (Bhatia & Kumar, 2020). The wide use of digital technologies enables superior integration between firms, and the combination of Big Data with cloud computing allows for real-time collaboration between companies, improving productivity and security (Maskuriy *et al.,* 2019).

The characterization of an Industry 4.0 business network is conditioned to an unlimited communication of companies with their suppliers and customers through modern communication networks, block chain or IoT (Internet of Things) (Saniuk, Saniuk & Cagáňová, 2021). This enables secure and real-time data exchange (Saniuk,

Saniuk & Cagáňová, 2021). In this sense, Wamba (2017) considers the existence of a digital infrastructure that includes aspects like connectivity, compatibility and modularity. In the case of competitors, although they may be considered passive actors in business networks, a growing participation of this group in the formation of alliances and organizational networks with focal companies has recently been observed (Tauhata & Macedo-Soares, 2004).

2.2 Big Data Analytics literature in business networks

Big Data should be understood as an immense amount of data continuously generated by multiple sources (Sheng, Amankwah-Amoah & Wang, 2017) and its analysis comprises modern analytical resources of Information Technologies (Leenen & Meyer, 2021). The BDA has become a necessary technology in business operations (Sheng, Amankwah-Amoah & Wang, 2017) that aims to detect patterns, correlations, trends and useful information (Leenen & Meyer, 2021). It allows companies to see things they would otherwise not be able to see (Ranjan & Foropon, 2021) and facilitates decision making (Mikalef *et al.*, 2019) particularly in operations management (Bertsimas, Kallus & Hussain, 2016).

The flexibility provided by the BDA technology (Grover & Kar, 2017) enables the transformation of business operations through its insights (Mikalef *et al.*, 2019) and the achievement of competitive advantage (Ghasemaghaei & Calic, 2020; Wamba *et al*., 2017) sustainable in improving operational performance (Matthias *et al.,* 2017) allowing an increase in companies' revenue around 8% (Liu, 2014). The technological innovation produced by the use of the BDA and its ability to learn (Ghasemaghaei & Calic, 2020) promote changes in the company's management and operations, in addition to fostering new businesses (Sheng, Amankwah-Amoah & Wang, 2017) and improving the company's performance (Ghasemaghaei, 2018) and its operational efficiency (Dubey *et al.,* 2019; Mikalef *et al.*, 2019; Srinivasan & Swink, 2018)).

The use of the BDA at the company level proposes governance schemes with definition of responsibilities and cooperation between departments to achieve the analytical maturity desired by the organization (Mikalef *et al.,* 2019). Certain authors consider the need for a solid analytical infrastructure to facilitate departmental operations to optimize manufacturing processes in terms of cost and resources (Grover & Kar, 2017). In that regard, should avoid focusing exclusively on collecting a large amount of data but give preference to the possibility of integration between different types of data (Ghasemaghaei & Calic, 2020). A basic use of the BDA technology in industry would be related to forecasting electricity demand (Wang *et al.,* 2016) and manufacturing activities enabling the reduction of operating costs, product quality, and availability (Dubey *et al.,* 2019).

Regarding the Industry 4.0 business network, the existence of cooperation in the sharing of customer and supplier databases, with standardization of formats and exchange protocols, allows the full use of the BDA, favoring cyber security (Leenen & Meyer, 2021) and the development of a data analytic culture (Dubey *et al.,* 2019). Some authors also suggest that the BDA confers greater bargaining power in negotiating with customers and suppliers (Zhang *et al.,* 2017).

Customers, suppliers and even consumers provide data in different formats, whether structured, those coming from the company's databases, or semi-structured and unstructured coming from meta search engines and social networks, among others (Leenen & Meyer, 2021; Papadopoulos *et al.,* 2017; Talón-Ballestero *et al.,* 2018). Unstructured data, which represent 80% of existing data (Tan *et al.,*2015), usually refer to text, web and multimedia, social media (Ghasemaghaei & Calic, 2020; Sheng, Amankwah-Amoah & Wang, 2017) and sensors being useful for better understanding of customers and cost reduction (Sheng, Amankwah-Amoah & Wang, 2017).

The BDA facilitates decision-making processes in near real time (Papadopoulos *et al.,* 2017), based on large amounts of data, improving the relationship between companies and their customers (Fox & Do, 2013) particularly by increasing the volume of business and online customer transactions (Wright *et al.,* 2019) and decreasing customer acquisition costs by about 47% (Liu, 2014). In this interaction, it was found that text analyzes produce new knowledge and those performed on social media show human behavior (Sheng, Amankwah-Amoah & Wang, 2017) and can be useful in the development of new products (Chan *et al.,* 2016).

This technological tool allows a response to the identification of customer requirements (Sheng, Amankwah-Amoah & Wang, 2017) which generates benefits in marketing, new product development and business strategy realignment (Mishra *et al.,* 2017). In general, the use of the BDA enables a better understanding of customers' preferences (Tan *et al.,*2015) and their needs (Zhang *et al.,* 2017), on the other hand, increasing concerns about ethical issues and trust between companies (Mikalef *et al.,* 2019).

Industry 4.0 suppliers participate in an integrated production of their customers and through the application of the BDA can develop innovation activities. These activities aim to add value to end consumers (Wright *et al.,* 2019) and support supply chain operations (Tan *et al.,*2015). It should also be noted that the speed of data collection and analysis plays a key role in improving a company's innovation performance (Ghasemaghaei & Calic, 2020) and that obtaining digital data on consumer behavior habits in e-commerce improve the company's relationships with suppliers (Sheng, Amankwah-Amoah & Wang, 2017).

The BDA can provide substantial value in Supply Chain Management (Mikalef *et al.,* 2019) with social media analytics, text mining, and machine learning applications (Grover & Kar, 2017), enabling superior process performance (Tan *et al.,*2015) and competitive advantage for companies in the supply chain (Grover & Kar, 2017).

Data collected on the Internet enable companies to identify ongoing competitions (Sheng, Amankwah-Amoah & Wang, 2017) and organize their competitive counterintelligence by gathering and analyzing information about competitors (Kamboj *et al.,* 2018). This analysis can allow market forecasting by mining textual information and competing companies' web sites (Nassirtoussi *et al.,* 2015) and allow companies to receive alerts about market fluctuations and the movements of these competitors (Ranjan & Foropon, 2021). The potential for using the BDA with competitors is still very low (Ranjan & Foropon, 2021).

The online marketing environment is dynamic (Wright *et al.,* 2019) and is being intensified by the growing competition from companies to better understand the BDA technology (Johnson, Friend & Lee, 2017; Ur Rehman *et al.,* 2016). The processes of searching for real-time information from competitors are not standardized (Ranjan & Foropon, 2021) due to the lack of technical capacity of the team and the limited understanding of the use of the BDA, among other aspects (Ranjan & Foropon, 2021).

Some authors suggest applications to automate the process of social media research about competitors like LinkedIn Analytics and the collection of data from products, customers and other sources (Ranjan & Foropon, 2021), and highlight that organizations are looking for computing methods that can be powered by Big Data like

IoT, intelligent robots, content analysis, AI and machine learning, among others (Ranjan & Foropon, 2021).

3 Organizational perspective for using BDA in Industry 4.0 business networks - discussion

Some authors argue that the value produced by the BDA is the result of organizational diffusion (Mikalef *et al.,* 2019; Wamba *et al.*, 2017) that can be leveraged in network operations. In business networks, several actors stand out that form groups of companies with their own characteristics and similar interests, like suppliers, customers, and competitors (Lu *et al.,* 2020). The presence of other organizations as public, financial and university (Hedvall, 2020) in these relationships also indicate excellent possibilities for the use of the BDA aiming at developing the operational management of the focal manufacturing company.

The perception that Industry 4.0 brings together companies in the business network connected by the Internet through intelligent resources (Grabowska, 2020) needs to be better understood. If, on the one hand, there is a predominantly private connection – including the intranet, with its customers, suppliers (Leenen & Meyer, 2021; Wright *et al.,* 2019) and even financial organizations, on the other hand, development of relationships with competitors (Ranjan & Foropon, 2021), and other actors, carried out exclusively by the world wide web, reinforcing in this case the public nature of this typology of data, even though it is subject to the laws of access to information in numerous countries. Thus, the application of the BDA by an Industry 4.0 company comprises two network perspectives: internal and external.

In an internal view, the potential for using the BDA is strictly related to the level of coordination and cooperation between companies (Grandori & Soda, 1995) existing in the inter-company relationship, especially trust (Mikalef *et al.,* 2019). Thus, the sharing of structured, semi and unstructured data (Leenen & Meyer, 2021; Papadopoulos *et al.,* 2017; Talón-Ballestero *et al.,* 2018), including those generated by sensors and machines (Grover & Kar, 2017), among companies, it can be intense, allowing fullness in the use of the BDA with advantages in cybersecurity (Leenen & Meyer, 2021) and in the development of a data analytic culture (Dubey *et al.,* 2019). The benefits in the operational management of companies with the use of the BDA can be reproduced, or even increased in Industry 4.0, with customers improving decision-making processes (Papadopoulos *et al.,* 2017), favoring the relationship (Fox & Do, 2013; Wright *et al.,* 2019) and the development of new products (Chan *et al.,* 2016), and generating benefits in marketing (Mishra *et al.,* 2017). Otherwise, the advantages of operational management with suppliers can be seen in innovation (Wright *et al.,* 2019) and in supporting supply chain operations (Mikalef *et al.,* 2019; Tan *et al.,*2015).

In its external use, the BDA in Industry 4.0 assumes legal limits that directly influence relations with competitors (Ranjan & Foropon, 2021), and other actors in the business network. In this environment, the data of interest are essentially semi- and unstructured (Leenen & Meyer, 2021; Papadopoulos *et al.,* 2017; Talón-Ballestero *et al.,* 2018) requiring the mining of textual information and websites of competing companies (Nassirtoussi *et al.,* 2015) and enabling the check of public information from government agencies, class associations and educational institutions, among other elements. In operational management, benefits are related to decision making (Kamboj *et al.,* 2018; Sheng, Amankwah-Amoah & Wang, 2017) and to the organization of competitive counterintelligence (Kamboj *et al.,* 2018).

On the one hand, the two perspectives for approaching the BDA in Industry 4.0 differ substantially from each other in their form of organization. The sharing of information between the manufacturing company and its customers and suppliers, arranged and integrated into CPS systems (Saniuk, Saniuk & Cagáňová, 2021), is processed through a simple data collection, based on an exchange of information in real time (Saniuk, Saniuk & Cagáňová, 2021) and organized in the degree of trust desired by the actors. On the other hand, in the external approach, the process is characterized by being exploratory and the search for data involves all sorts of difficulties, since legal, even the access to sources of interest and the reliability of the data. This dichotomy can result in the existence of a confused analytical culture in the organization, which would somehow justify the arguments of some authors indicating that the BDA has even hindered the operational decisions of the company (Merendino *et al.,* 2018).

In this way, organizing activities with the use of the BDA technology by the Industry 4.0 company becomes essential. The external perspective of the focal industry, of an aggressive BDA on the internet, embrace aspects related to competitive intelligence (Kamboj *et al.,* 2018), to competitiveness (Nassirtoussi *et al.,* 2015; Ranjan & Foropon, 2021; Sheng, Amankwah-Amoah & Wang, 2017) and online marketing (Wright *et al.,* 2019). This mode of operation, of fundamental importance for the company and covered with confidentiality and legal assistance, needs to be conducted by an analytical element/group directly directed by Industry 4.0 – focal company. Otherwise, in the business network environment, the focal company collects data from customers, suppliers and organizations of interest. This internal perspective may suggest the creation of the BDA units by the business network itself, taking advantage of their synergistic nature of efforts (Balestrin & Verschoore, 2020), diluting costs and sharing the analysis for all interested actors.

4 Methodology

This research is exploratory in nature and originated from the perception of several authors who noted an important gap in the conduct of BDA research involving a perspective centered on organizational theory (Sheng, Amankwah-Amoah & Wang, 2017; Wamba *et al.*, 2017). The organizational theory chosen was the perspective of business networks due to the possibility that this model offers in saving resources for the creation and development of BDA teams. In this way, all companies belonging to the business network can contribute to the acquisition and maintenance costs of a BDA element to provide services to organizations.

On 06/24/21 a documental search was carried out in the Scopus (Elsevier) database, internationally known, using the search fields by title, abstract and keyword with the chain ("operations") AND ("Big Data" *"), from 2012 to 2021 and with the filters: (a) study area only Administration, Business and Accounting, (b) type of documents, only scientific papers, and (c) English language. The result was 56 documents that had their abstracts read and 14 papers dealing with BDA technology and its relationship with the operational level of organizations were used.

The second search in the Scopus database took place on 07/07/21 with the terms: (“Big Data*”) AND (“Business Network”) for paper title, abstract, and Keywords and under the same conditions as above. There were 27 documents left, which after analyzing their abstracts, left six papers of interest that were read and helped in the construction of the theoretical framework of this study.

The theme was expanded in order to absorb practical research that dealt with the application of BDA in manufacturing companies, thus carrying out a third search on 11/15/21. In the Scopus platform and in the conditions previously reported, changing only the search current ("Big Data*") AND ("Industry 4.0") OR ("digital manufacturing") the investigation resulted in seven scientific papers that were used in this work. Due to the exploratory characteristics of the research, a final search was needed on 11/25/21 to know the elements ("ERP") and ("digital Twins") in the context of Industry 4.0, finding three useful papers for the work.

5 BDA employment in Industry 4.0 business networks

This section proposes to present the BDA employment modeling based, to a large extent, on the organizational perspectives of the industry 4.0 business network. Two fundamental aspects of the use of BDA technology are presented below, which refer to the source and the preparation of data and algorithms for the development of an efficient analysis. At the end of the section some techniques that have been used in BDA are cited.

5.1 Data Sources for Big Data Analysis

It can be considered that the big data source plays a decisive role in the verification and analysis of BDA technology results. The sources available for this purpose are presented below.

5.1.1 Internal - *Intranet*

The corporate Enterprise Resource Planning (ERP)

Because it is a system based on Information Technology (Iris & Cebeci, 2014) with a high degree of integration of information systems for the processing of company information (Werner, Wiese & Maas, 2021) and also because it is considered a system source information with process recognition (Dumas *et al*., 2005) enterprise resource planning can provide data, and even information, about business processes (Iris &Cebeci, 2014) of companies, particularly those related to the organization's event registration like product sales and existing stocks.

There are several types of ERP systems: ERP SAP R/3, Oracle, BAAN, MFG Pro, JDEdwards, People Soft, etc., and there are even systems developed by the companies themselves (Tarigan *et al*., 2021) that contribute to the construction of a system single database that companies can share with suppliers (Tarigan *et al*., 2021). Some authors even consider the application of process mining in organizational purchases in an ERP system viable (Werner, Wiese & Maas, 2021).

Thus, it can be stated that the high degree of automation, particularly that existing in industry 4.0, allows the capture of data and quantitative information about transaction processing (Werner, Wiese & Maas, 2021). In this way, ERP systems in use by companies can integrate all partners involved in the supply chain (Tarigan *et al*., 2021), particularly suppliers and customers (Zhao *et al*., 2013), and better performance in the purchasing process can be achieved by integrating the company with its suppliers (He *et al*., 2014).

The digital twin

Digital twin comprises the existence of a physical entity and a related virtual entity that allows real-time simulation (Tao *et al*., 2019). This model is built by installing sensors (vibration, temperature, acoustics, force, velocity, position, and camera images) on the physical elements of the factory (Warke *et al*., 2021) generating huge amounts of data (Fattahi *et al*., 2021) in the digital simulation models that analyzed can favor the decision-making process and improve performance (Blomkvist *et al.*, 2020). This concept even enables digital representation applied to physical products (Grieves, 2015) and is also used in machine monitoring, predictive maintenance, process optimization, and economic production (Warke *et al*., 2021), allowing feedback from IoT manufacturing enablers (Zhang *et al.,* 2016).

Many authors understand that technological advances in data analysis and other enabling technologies present in Industry 4.0 facilitate the integration between digital twin and prominent fields (Fuller *et al.,* 2020). Some authors recognize that, in addition to the digital version of the documentation of operational activities, the datasets needed to build the digital twin must also be added to Big Data (Fattahi *et al*., 2021) since the creation of a twin phenomenon requires many steps and it becomes necessary to integrate it with the built-in IoT enablers (Fattahi *et al.,* 2021).

5.1.2 External **-** *Internet*

The Internet can be considered the main source for data capture for BDA. Companies are increasingly using the Internet in their organizational strategies, whether in assessing market demand, promoting the brand, sharing information and collecting feedback (Choo 2006). The data used on the websites of organizations allow to identify and measure the changes that occur in the business environment (Arora *et al.,* 2020). Some authors also consider that social media promote entrepreneurship processes due to the possibility of social interaction that develops (Fischer & Reuber, 2011).

Websites can demonstrate changes in market orientation, product portfolio and the composition of alliance networks (Li *et al*. 2016), even the company's ongoing research and development activities and the granting of patents (Gök *et al*., 2015). Despite these advantages, the use of data from these sites has limitations like sampling bias (Arora *et al.,* 2020), among others, which deserve special attention from BDA analysts.

With the development of the Semantic Web that proposes the exchange and use of information between large information silos, thus increasing the power of the Web (Berners-Lee *et al.,* 2006), Big Data sources multiply making it possible to organize information for computers and machines in RDF format (Resource Description Framework).

The analysis of real-time data sought in traffic applications and companies' transport information would facilitate operations by indicating available and efficient routes (Mishra & Tripathi, 2021). With the application of BDA in the network companies and in this context, new opportunities for cost savings could be revealed.

LinkedIn or Twitter platforms should also be considered as Big Data sources for business analysis and innovation. (Arora *et al.,* 2020), among other social media. The analysis of existing Big Data sources on the Internet enables data intelligence that allows recognizing customer inclinations (Muthuveloo & Ping, 2013).

5.2 Data preparation and algorithms applied in Big Data analysis

As a general feature, Big Data must be readily accessible and readable by humans or machines (Fattahi *et al.,* 2021). In certain situations, using sensors for example, this dataset needs to be prepared using relevant, pre-processed documentation. (Fattahi *et al.,* 2021) so that the analysis process can be carried out in the manufacturing industry. A Big Data preparation model considering Digital Twins must use a set of surface roughness data for each material process developed in the industry (turning, milling, grinding, polishing, and machining by electrical discharge) (Fattahi *et al.,* 2021). In other situations where there is no possibility of data preparation, industries can create and develop specific algorithms to search for Big Data available on the Internet and on the companies' network Intranet or even in the organization itself.

Nowadays, general purpose, heuristic, fuzzy, and genetic mining algorithms are sophisticated and are available on the market even in programming languages (Werner, Wiese & Maas, 2021). The logic of a mining algorithm can be understood as the analysis of input data from an information system to create a desirable output process. (Werner, Wiese & Maas, 2021). Algorithm input data comes from the business activities of interest and is tracked in a defined sequence of desirable events (Werner, Wiese & Maas, 2021) enabling the analysis of business processes in a quantitative way (Werner, Wiese & Maas, 2021). The main limitation of process mining algorithms is restricted to the need to record transactions and events in the source information systems (Werner, Wiese & Maas, 2021).

Algorithms built for the site mining technique require a detailed study of the validity of variables based on keywords and their correlation (Arora *et al.,* 2020). To model topics on websites, the algorithm called Latent Dirichlet Allocation (LDA) can be used to reveal the probability distribution of words and topics (Arora *et al.,* 2020).

5.3 BDA employment techniques

As it is a sophisticated technological tool, the BDA already has some employment techniques that can be seen below. Again, the techniques are guided by the organizational perspectives of the industry 4.0 business network.

5.3.1 Internal - *Intranet*

The data analysis technique that uses process mining is performed by algorithms that have the ability to investigate computer operating systems by gathering information about an organization's business processes (Werner, Wiese & Maas, 2021). This technique makes it possible to create reliable business process models automatically as it uses data from recorded events (Werner, Wiese & Maas, 2021) helping to overcome difficulties related to using a large amount of data.

Real-time monitoring and control of manufacturing systems in the automotive industry uses unstructured datasets collected from sensors that are pre-processed by specific software (Apache Kafka, Apache Storm e MongoDB) and when analyzed allow the detection of failures (Syafrudin *et al.,* 2018). The sensors existing in the factory also produce a Big Data ecosystem that is handled by various technologies (data lake, banco de dados NoSQL, Apache Spark, Apache Drill, Apache Hive and OPC Collector) enable predictive maintenance (Yu *et al.,* 2019) of machinery and equipment. The use of BDA can also be developed to ensure low energy consumption of the cyberphysical industrial production system by matching the energy-related dataset with those relating to production (Zhang, *et al.,* 2020).

5.3.2 External - *Internet*

Searching the World Wide Web can preferably be performed by analytical software that captures text on standard HTML (Hyper Text Markup Language) website pages (Arora *et al.,* 2020). Greater sophistication can be achieved in investigating Adobe Flash and PDF content also used on the Internet. The site mining technique has advantages like accessibility and opportunity (Arora *et al.,* 2020) and limitations (Gök *et al.* 2015) requiring special preparation from the BDA teams particularly with regard to the reliability of data sources.

The Wayback machine tool despite its limitations (Arora *et al.,* 2020), can be used as a basis for starting website mining due to its ability to archive historical HTML pages. For web data extraction work can be used the software IBM Content Analytics (Arora *et al.,* 2020). In this context, the website pages of competing companies can be analyzed using the unstructured text mining technique and the strategic changes that have taken place over time can be verified, as the web content undergoes changes and you can even reveal the dynamic capabilities of each company (Arora *et al.,* 2020).

Ferreira (2016), in turn, it indicates the use of datasets to optimize product price decisions in retail companies, improving Big Data forecasting models, since speed in decision-making is essential for companies to preserve stability in a market in changes (Siagian *et al.,* 2020).

Process mining has been widely used as an analysis technique in social networks (Maita *et al.*, 2018) and with the possibility of application also in the manufacturing industry (Lee *et al*., 2013).

6 Conclusions

It is concluded, therefore, that the use of the BDA in the operational management of a manufacturing company in Industry 4.0, established in business networks, suggests the organization of two structures, each with its own characteristics and distinct analytical culture. The first, preferably under the hybrid governance of the business network and with passive action, with the objective of collecting data from suppliers and customers, and eventually from organizations on the internet. The second structure organized and directly controlled by the manufacturing company to actively search for data in an exploratory way on the World Wide Web.

As the main contribution to organizational theory, the study allowed to systematize the operational management of the BDA technology in manufacturing companies of Industry 4.0. It was evidenced the organization in the company of two BDA employment mechanisms, collection and search, which can be adapted for companies from other economic sectors.

Secondly, the research proposes that the Industry 4.0 business networks are oriented towards the use of BDA technology. This possibility shows advantages for the associated companies, since this technology brings benefits both to the focal companies, as well as to suppliers and customers, and allows cost savings for the installation of the necessary infrastructure.

The main limitation of this research study was the scarce literature on the use of big data analytics technology centered on Organizational Theory. For future research, it is recommended to better understand the existing analytical culture in companies, which use the BDA technology, in order to confirm the assumptions presented.

References

Arora, S. K., Li, Y., Youtie, J., & Shapira, P. (2020). Measuring dynamic capabilities in new ventures: exploring strategic change in US green goods manufacturing using website data. *The Journal of Technology Transfer*, *45*(5), 1451-1480.

Balestrin A, Verschoore JR (2020) Relações interorganizacionais e complementaridade de conhecimentos: proposição de um esquema conceitual. RAM. Revista de Administração Mackenzie, 8, 153-177. doi.org/10.1590/167869712007/administracao.v8n4p153-177.

Berners-Lee, T., Hall, W., Hendler, J., Shadbolt, N., & Weitzner, D. J. (2006). Creating a Science of the Web. *Science*, *313*(5788), 769-771.

Bertsimas D, Kallus N, Hussain A (2016) Inventory management in the era of big data. Production and Operations Management, vol 25(12), pp 2006-2009. doi:10.1111/POMS.2_12637.

Bhatia M S, Kumar S (2020) Critical success factors of Industry 4.0 in automotive manufacturing industry. IEEE Transactions on Engineering Management. **doi:** 10.1109/TEM.2020.3017004.

Blomkvist, Y., & Ullemar Loenbom, L. E. O. (2020). Improving supply chain visibility within logistics by implementing a Digital Twin: A case study at Scania Logistics.

Camargo Fiorini P, Seles BMRP, Jabbour CJC, Mariano EB, de Sousa Jabbour, ABL (2018) Management theory and big data literature: From a review to a research agenda. International Journal of Information Management, vol 43, pp 112-129. doi.org/10.1016/j.ijinfomgt.2018.07.005.

Chan HK, Wang X, Lacka E, Zhang M (2016) A mixed-method approach to extracting the value of social media data. Production and Operations Management, vol 25 (3), pp 568-583. doi.org/10.1111/poms.12390.

Chen, C. T., & Chiu, Y. T. (2021). A study of dynamic fuzzy cognitive map model with group consensus based on linguistic variables. *Technological Forecasting and Social Change*, *171*, 120948.

Choo, C. W. (2006). The knowing organization. New York: Oxford University Press.

De Mauro, A., Greco, M., & Grimaldi, M. (2016). A formal definition of Big Data based on its essential features. *Library Review*.

Dubey R, Gunasekaran A, Childe SJ, Blome C, Papadopoulos T (2019) Big data and predictive analytics and manufacturing performance: integrating institutional theory, resource-based view and big data culture. British Journal of Management, vol 30(2), pp 341-361. doi.org/10.1111/1467-8551.12355.

Dumas, M., Van der Aalst, W. M., & Ter Hofstede, A. H. (2005). Process-aware information systems: bridging people and software through process technology. John Wiley & Sons.

Fattahi, S., Okamoto, T., & Ura, S. (2021). Preparing Datasets of Surface Roughness for Constructing Big Data from the Context of Smart Manufacturing and Cognitive Computing. *Big Data and Cognitive Computing*, *5*(4), 58.

Ferreira, K. J., Lee, B. H. A., & Simchi-Levi, D. (2016). Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management*, *18*(1), 69-88.

Fischer, E., & Reuber, A. R. (2011). Social interaction via new social media: (How) can interactions on Twitter afect efectual thinking and behavior? Journal of Business Venturing, 26(1), 1–18. https://doi.org/10.1016/j.jbusvent.2010.09.002.

Fox S, Do T (2013) Getting real about Big Data: applying critical realism to analyse Big Data hype. International Journal of Managing Projects in Business. doi:10.1108/IJMPB-08-2012-0049.

Fuller, A.; Fan, Z.; Day, C.; Barlow, C. Digital Twin: Enabling Technologies, Challenges and Open Research. IEEE Access 2020, 8, 108952–108971. [CrossRef]

Ghasemaghaei M (2018) Improving organizational performance through the use of big data. Journal of Computer Information Systems. doi. org/10.1080/08874417.2018.1496805.

Ghasemaghaei M, Calic G (2020) Assessing the impact of big data on firm innovation performance: Big data is not always better data. Journal of Business Research, vol108, pp147-162. doi: 10.1016/j.jbusres.2019.09.062.

Gligor DM, Pillai KG, Golgeci I (2021) Theorizing the dark side of business-to-business relationships in the era of AI, big data, and blockchain. Journal of Business Research, vol 133, pp 79-88. doi: 10.1016/j.jbusres.2021.04.043.

González-Serrano L, Talón-Ballestero P, Muñoz-Romero S, Soguero-Ruiz C, Rojo-Álvarez J L (2021) A Big Data Approach to Customer Relationship Management Strategy in Hospitality Using Multiple Correspondence Domain Description. Applied Sciences, vol 11(1), p 256. doi: 10.3390/app11010256.

Gök, A., Waterworth, A., & Shapira, P. (2015). Use of web mining in studying innovation. Scientometrics, 102(1), 653–671. https://doi.org/10.1007/s11192-014-1434-0.

Grabowska S (2020) Smart Factories in the Age of Industry 4.0. Management Systems in Production Engineering, vol 28(2), pp 90-96. doi: 10.2478/mspe-2020-0014.

Grandori A, Soda G. (1995) Inter-firm networks: antecedents, mechanisms and forms. Organization studies, vol 16(2), pp 183-214. doi.org/10.1177/017084069501600201.

Grover P, Kar AK (2017) Big data analytics: A review on theoretical contributions and tools used in literature. Global Journal of Flexible Systems Management, vol 18(3), pp 203-229. doi:10.1007/s40171-017-0159-3.

Gulati R, Nohria N, Zaheer A. (2000) Strategic networks. Strategic management journal, vol 21(3), pp 203-215.

He, Y., Lai, K. K., Sun, H., & Chen, Y. (2014). The impact of supplier integration on customer integration and new product performance: The mediating role of manufacturing flexibility under trust theory. *International Journal of Production Economics*, *147*, 260-270.

Hedvall K (2020) Conceptualising Solutions in Business Networks: The Case of Heavy Vehicle Maintenance. Doktorsavhandlingar vid Chalmers tekniska högskola. Ny serie, (4733).

Humphrey J, Todeva E, Armando E, Giglio E (2020) Global value chains, business networks, strategy, and international business: Convergences. Revista Brasileira de Gestão de Negócios, vol 21, pp 607-627. doi.org/10.7819/rbgn.v21i4.4014.

Iris, C., & Cebeci, U. (2014). Analyzing relationship between ERP utilization and lean manufacturing maturity of Turkish SMEs. *Journal of Enterprise Information Management*.

Ivanov, D., Tang, C. S., Dolgui, A., Battini, D., & Das, A. (2021). Researchers' perspectives on Industry 4.0: multi-disciplinary analysis and opportunities for operations management. *International Journal of Production Research*, *59*(7), 2055-2078.

Johnson J S, Friend S B, Lee H S (2017) Big data facilitation, utilization, and monetization: Exploring the 3Vs in a new product development process. Journal of Product Innovation Management, vol 34(5), pp 640-658. doi: 10.1111/jpim.12397.

Kamboj S, Sarmah B, Gupta S, Dwivedi Y (2018) Examining branding co-creation in brand communities on social media: Applying the paradigm of Stimulus-Organism-Response. International Journal of Information Management, vol 39, pp 169-185. doi:10.1016/j.ijinfomgt.2017.12.001.

Lee, S. K., Kim, B., Huh, M., Cho, S., Park, S., & Lee, D. (2013). Mining transportation logs for understanding the after-assembly block manufacturing process in the shipbuilding industry. *Expert systems with applications*, *40*(1), 83-95.

Leenen L, Meyer T (2021) Artificial intelligence and big data analytics in support of cyber defense. In Research Anthology on Artificial Intelligence Applications in Security pp. 1738-1753. IGI Global. doi: 10.4018/978-1-5225-8304-2.CH002.

Leick B, Gretzinger S (2020) Business networking in organisationally thin regions: a case study on network brokers, SMEs and knowledge-sharing. Journal of Small Business and Enterprise Development, vol *27*(5), pp 839-861. doi.org/10.1108/JSBED-12-2019-0393.

Li, Y., Arora, S. K., Youtie, J., & Shapira, P. (2016). Using web mining to explore Triple Helix infuences on growth in small and mid-size frms. Technovation. https://doi.org/10.1016/j.technovation.2016.01.002.

Liu Y (2014) Big data and predictive business analytics. The Journal of Business Forecasting, vol 33(4), p 40.

Lu Y, Wei W, Li Y, Wu Z, Jin H (2020) The formation and evolution of interorganisational business networks in megaprojects: a case study of Chinese skyscrapers. Complexity, 2020. doi: 10.1155/2020/2727419.

Maita, A. R. C., Martins, L. C., Lopez Paz, C. R., Rafferty, L., Hung, P. C., Peres, S. M., & Fantinato, M. (2018). A systematic mapping study of process mining. *Enterprise Information Systems*, *12*(5), 505-549.

Maskuriy R, Selamat A, Ali KN, Maresova P, Krejcar O (2019) Industry 4.0 for the construction industry-how ready is the industry? Appl. Sci. 2019, vol 9, pp 2819. doi:10.3390/app9142819.

Matthias O, Fouweather I, Gregory I, Vernon A (2017) Making sense of big data–can it transform operations management? International Journal of Operations & Production Management. doi.org/10.1108/IJOPM-02-2015-0084.

Merendino A, Dibb S, Meadows M, Quinn L, Wilson D, Simkin L, Canhoto A (2018). Big data, big decisions: The impact of big data on board level decision-making. Journal of Business Research, vol 93, pp 67-78. doi:10.1016/j.jbusres.2018.08.029.

Mikalef P, Boura M, Lekakos G, Krogstie J (2019) Big data analytics and firm performance: Findings from a mixed-method approach. Journal of Business Research, vol 98, pp 261-276. doi:10.1016/j.jbusres.2019.01.044.

Mishra N, Singh A, Rana NP, Dwivedi YK (2017) Interpretive structural modelling and fuzzy MICMAC approaches for customer centric beef supply chain: application of a big data technique. Production Planning & Control, vol 28(11-12), pp 945-963. doi.org/10.1080/09537287.2017.1336789.

Mishra, S., & Tripathi, A. R. (2021). AI business model: an integrative business approach. *Journal of Innovation and Entrepreneurship*, *10*(1), 1-21.

Mubarak MF, Petraite M (2020) Industry 4.0 technologies, digital trust and technological orientation: What matters in open innovation? Technological Forecasting and Social Change, vol 161, p120332. doi:10.1016/j.techfore.2020.120332.

Muthuveloo, R., & Ping, T. A. (2013). Achieving business sustainability via I-Top Model. *American Journal of Economics and Business Administration*, *5*(1), 15.

Nassirtoussi AK, Aghabozorgi S, Wah TY, Ngo DCL (2015) Text mining of news-headlines for FOREX market prediction: A Multi-layer Dimension Reduction Algorithm with semantics and sentiment. Expert Systems with Applications, vol 42(1), pp 306-324. doi: 10.1016/j.eswa.2014.08.004.

Papadopoulos T, Gunasekaran A, Dubey R, Altay N, Childe SJ, Fosso-Wamba S (2017) The role of Big Data in explaining disaster resilience in supply chains for sustainability. Journal of Cleaner Production, vol142, pp1108-1118. doi:10.1016/j.jclepro.2016.03.059.

Piccarozzi M, Aquilani B, Gatti C (2018) Industry 4.0 in management studies: A systematic literature review. Sustainability, vol10(10), p 3821. doi.org/10.3390/su10103821.

Preuveneers D, Joosen W, Ilie-Zudor E (2018) Policy reconciliation for access control in dynamic cross-enterprise collaborations. Enterprise Information Systems, vol12(3), pp 279-299. doi.org/10.1080/17517575.2017.1355985.

Ranjan J, Foropon C (2021) Big data analytics in building the competitive intelligence of organizations. International. Journal of Information Management, vol 56, p 102231. doi: 10.1016/j.ijinfomgt.2020.102231.

Saniuk S, Saniuk A, Cagáňová D (2021) Cyber Industry Networks as an environment of the Industry 4.0 implementation. Wireless Networks, vol 27 (3) pp 1649-1655. doi:10.1007/s11276-019-02079-3.

Sheng J, Amankwah-Amoah J, Wang X (2017) A multidisciplinary perspective of big data in management research. International Journal of Production Economics, vol 191, pp 97-112. doi:10.1016/j.ijpe.2017.06.006.

Siagian, H., Jade, K., & Tarigan, Z. (2020). The role of affective leadership in improving firm performance through the integrated internal system and external integration FMCG Industry. *International Journal of Data and Network Science*, *4*(4), 365-372.

Srinivasan R, Swink M (2018) An investigation of visibility and flexibility as complements to supply chain analytics: An organizational information processing theory perspective. Production and Operations Management, vol 27(10), pp1849-1867. doi.org/10.1111/poms.12746.

Syafrudin, M., Alfian, G., Fitriyani, N. L., & Rhee, J. (2018). Performance analysis of IoT-based sensor, big data processing, and machine learning model for real-time monitoring system in automotive manufacturing. *Sensors*, *18*(9), 2946.

Talón-Ballestero P, González-Serrano L, Soguero-Ruiz C, Munoz-Romero S, Rojo-Álvarez JL (2018) Using big data from customer relationship management information systems to determine the client profile in the hotel sector. Tourism Management, vol 68, pp187-197. doi: 10.1016/J.TOURMAN.2018.03.017.

Tan KH, Zhan Y, Ji G, Ye F, Chang C (2015) Harvesting big data to enhance supply chain innovation capabilities: An analytic infrastructure based on deduction graph. International Journal of Production Economics, vol 65, pp 223-233. doi: 10.1016/j.ijpe.2014.12.034.

Tao, F.; Liu, W.; Zhang, M.; Hu, T.; Qi, Q.; Zhang, H.; Sui, F.; Wang, T.; Xu, H.; Huang, Z.; et al. Five-dimension digital twin model and its ten applications. Jisuanji Jicheng Zhizao Xitong/Comput. Integr. Manuf. Syst. CIMS 2019, 25, 1–18.

Tarigan, Z. J. H., Siagian, H., & Jie, F. (2021). Impact of Enhanced Enterprise Resource Planning (ERP) on Firm Performance through Green Supply Chain Management. *Sustainability*, *13*(8), 4358.

Tauhata TL, Macedo-Soares TD (2004) Redes e alianças estratégicas no Brasil: caso CVRD. RAE eletrônica, vol 3. doi: 10.1590/s1676-56482004000100004.

Ur Rehman MH, Chang V, Batool A, Wah, TY (2016) Big data reduction framework for value creation in sustainable enterprises. International Journal of Information Management, vol 36(6), pp 917-928. doi: 10.1016/j.ijinfomgt.2016.05.013.

Wamba SF, Gunasekaran A, Akter S, Ren SJF, Dubey R, Childe SJ (2017) Big data analytics and firm performance: Effects of dynamic capabilities. Journal of Business Research, vol 70, pp 356-365. doi: 10.1016/j.jbusres.2016.08.009.

Wang S, Wan J, Li D, Zhang C (2016) Implementing smart factory of industrie 4.0: an outlook. International journal of distributed sensor networks, vol 12(1), p 3159805. doi.org/10.1155/2016/3159805.

Wang G, Gunasekaran A, Ngai E W, Papadopoulos T (2016) Big data analytics in logistics and supply chain management: Certain investigations for research and applications. International Journal of Production Economics, vol 176, pp 98-110. doi: 10.1016/j.ijpe.2016.03.014.

Warke, V., Kumar, S., Bongale, A., & Kotecha, K. (2021). Sustainable Development of Smart Manufacturing Driven by the Digital Twin Framework: A Statistical Analysis. *Sustainability*, *13*(18), 10139.

Werner, M., Wiese, M., & Maas, A. (2021). Embedding process mining into financial statement audits. *International Journal of Accounting Information Systems*, *41*, 100514.

Wright LT, Robin R, Stone M, Aravopoulou DE (2019) Adoption of big data technology for innovation in B2B marketing. Journal of Business-to-Business Marketing, vol 26(3-4), pp 281-293. doi.org/10.1080/1051712X.2019.1611082.

Yu, W., Dillon, T., Mostafa, F., Rahayu, W., & Liu, Y. (2019). A global manufacturing big data ecosystem for fault detection in predictive maintenance. *IEEE Transactions on Industrial Informatics*, *16*(1), 183-192.

Zhao, L., Huo, B., Sun, L., & Zhao, X. (2013). The impact of supply chain risk on supply chain integration and company performance: a global investigation. *Supply Chain Management: An International Journal*.

Zhang, Y., Qian, C., Lv, J., & Liu, Y. (2016). Agent and cyber-physical system based self-organizing and self-adaptive intelligent shopfloor. *IEEE Transactions on Industrial Informatics*, *13*(2), 737-747.

Zhang Y, Ren S, Liu Y, Si S (2017) A big data analytics architecture for cleaner manufacturing and maintenance processes of complex products, Journal of Cleaner Production, Vol. 142, pp. 626-641. doi.org/10.1016/j.jclepro.2016.07.123.

Zhang, C., Wang, Z., Ding, K., Chan, F. T., & Ji, W. (2020). An energy-aware cyber physical system for energy Big data analysis and recessive production anomalies detection in discrete manufacturing workshops. *c*, *58*(23), 7059-7077.

Web references

Grieves, M. Digital Twin: Manufacturing Excellence through Virtual Factory Replication This paper introduces the concept of a Whitepaper by Dr. Michael Grieves. White Pap. 2015. Available online: https://www.researchgate.net/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication (accessed on 10 December 2021).