

IS STRATEGIC MANAGEMENT READY FOR BIG DATA? A REVIEW OF THE BIG DATA ANALYTICS LITERATURE IN MANAGEMENT RESEARCH

M. BOGDAN¹, D.C. LUNGESCU²

¹ Babeş-Bolyai University, mihai.bogdan@econ.ubbcluj.ro

² Babeş-Bolyai University, dan.lungescu@econ.ubbcluj.ro

ABSTRACT

The purpose of this paper is to obtain an overview of the current understanding of big data analytics in the strategic management research. Big data's positive effects upon performance are strongly advocated by practitioners, yet it hasn't been decisively proven by researchers. It is acknowledged that academia has to cover the gap between itself and the practitioner's field in this domain.

We discuss the case of why big data analytics has strategic implications from multiple perspectives: big data analytics implies a change in organizational culture, towards a data-driven decision-making process; top level leadership support is required; companies need to align the right resources, skills and capabilities; it is a long and arduous process to gather sufficient data in order to obtain significant insights.

The methodology of our study is based on the review of the articles that approach the link between big data analytics and strategic management and are indexed on Web of Knowledge during period 2010 – 2018.

Based on the reviewed literature we identified the following: (1) the scope of big data analytics in organization's strategy, (2) the types of analytics used, (3) best practices for big data strategy, (4) challenges of big data analytics, (5) further research questions concerning the big data strategy.

Keywords: big data analytics, data science, business analytics, literature review, strategic management.

JEL classification: temp.

1. Introduction

Big data appears to be the new excitement of the business world. It belongs to a bigger field of activities, all having extravagant names such as: machine learning, artificial intelligence, robotic process automation, data mining etc. Part statistics, part programming, it definitely creates a lot of uncertainty within the economic environment as it has the premises to be the next major industrial revolution.

As any environment change, it impacts the way organizations work and hence, the management of these organizations will be changed with it. Managers are both excited about opportunities it brings, yet a lot are worried as this phenomenon is not well understood and it might obsolete a lot of today's activities, jobs and organizations.

This paper aims to review the state of the knowledge of big

data analytics amongst strategic management researchers. First, we propose to present the field of big data analytics with its characteristics. Next, we identify the link between the big data analytics and strategic management, and the effects they have on each other.

To obtain these insights, we did an exhaustive review of the current literature within the strategic management field which treats the subject of big data analytics.

While the primary objective of the study is to get a comprehensive understanding of the answers provided by the academic literature, we also intend to develop future research questions.

In our research, we identified previous literature review papers which aimed to present an integrated approach in studying big data and management research together (Sheng et al. 2017), to present the opportunities and challenges of business intelligence and analytics (Chen et al. 2012) or to understand the big data challenges and the methods available to overcome them (Sivarajah et al. 2017).

Our study is novel in the methodology approached, in the way that we aimed to obtain insights from the current literature available on Web of Science with the help of text analytics algorithms in two ways. First, we extracted n-grams (trigrams and 4-grams) from the text body of our articles. N-grams are frequent sequence of n words within the text. Second, we performed an automatic text summarization through which we extracted the most relevant sentences from our articles based on the frequency of their words within the entire text.

Another expectation of our research is that big data as different implications based on the industry in which the company belongs.

2. Big data defined and explained

Big data is a special relatively new segment of business analytics. A comprehensive definition of this branch is given by Davenport and Harris (2007) who pointed out that analytics means "the extensive use of data, statistical and quantitative analysis, explanatory and predictive models, and fact-based management to drive decisions and actions". The business analytics domain originated at the same time with the appearance of computers in the late 1940s nurtured by the researchers from IBM (Evans, 2016) and it can be considered as the foundation for data science and big data (UpX Academy, 2016). This domain is closely related with

that of statistics, this being confirmed by the fact that business analytics is classified in three main categories: descriptive analytics, predictive analytics and prescriptive analytics (Davenport & Kim, 2013). Briefly explained, descriptive analytics displays "what happened?", predictive analytics identifies "what is probable to happen?", while prescriptive analytics advises "what should we do about it?"

Big data is a special segment of business analytics that has gained much attention in the last decade, in the beginning in the practitioner's world and later in academia. Its main particularity is given by the characteristics of data that is being used. Big data and big data analytics describe data sets and the operations that come with it that are very large (terabytes or even exabytes), complex and unstructured (e.g.: sensor data or social media data).

Gartner Inc. defines big data in the following terms: "Big data is high-volume, high-velocity and/or high-variety information assets that demand cost-effective, innovative forms of information processing that enable enhanced insight, decision making, and process automation (Gartner, n.d.)." We notice here that the focus in this definition is on the characteristics of big data, such as high in volume, velocity and variety. Also, on the fact that data represents a company's asset, in other words a resource with an economic value through which the owner expects to obtain future benefits. The second part of the definition refers to the state-of-the-art tools required to handle such data. The third part of the definition highlights the benefits of using big data, which are: "new discoveries", fact-based decisions and automations.

A good grasp of big data is ensured if we reflect about what characterizes big data. As in the above definition, there is an agreement among researchers about three characteristics of big data, volume, velocity and variety as defining (Laney, 2001; Zikopoulos et al. 2012; Chen and Zhang, 2014). Like the name speaks for himself, volume depicts the vast number of variables and observations from a dataset (George et al, 2016). Variety refers to the multitude forms under which data can be found either structured or unstructured, such as: text, audio, images, video, web activity, sensor data, transactions etc. (Miele and Shockley, 2013). Velocity addresses the time dimensions, meaning the speed at which data is being generated and the agility with which is being analyzed (UpX Academy, 2016). Other authors add veracity as a characteristic of big data (Miele and Shockley, 2013) which refers to the reliability and trustworthiness of the data. The fifth characteristic is value, meaning that the insights and patterns generated by big data are more relevant compared with those obtained from classical "small data" (Sharda et al. 2015). Van Rijmenam (2013), Seddon and Currie (2017), Mikalef et al. (2017) make a more comprehensive presentation of big data's characteristics adding two additional elements, variability, referring to the multiple ways in which data can be analyzed in order to obtain insights and visualization which is based on the beneficial effects of visual analytics in understanding patterns.

3. Big data's impact upon strategic management

In order to understand to implication of big data upon strategic management, we ask first what is the role of the strategy. The resolution for this query has been answered a long time ago by Porter (1996) who defined strategy as "the creation of a unique and valuable position, involving a different set of activities." To obtain above average returns, companies must develop a unique value proposition, the focus being on differentiation. Besides this, Porter argues that the essence of strategy is in doing trade-offs, in other words choosing what not to do. Decision makers should choose only those activities that reinforce each other. We understand from these definitions, that the big data strategy of a company, in order to succeed, must be integrated to the entire set of activities undertaken.

Many are tempted to believe that the biggest barrier in implementing a digital strategy would be the technological one. Managers might fear that their organizations do not have the necessary access to hardware and software resources in order to adopt a big data analytics strategy. However, Davenport (2006), McAfee and Brynjolfsson (2012), support that the biggest challenges companies will encounter are related to changing the company culture, gaining top leadership support, aligning the organization towards a common big data initiative, finding the people with the right set of skills, changing the decision-making process, and yes, also, getting the right technology.

For this reason, company leaders must be careful not to fall into the trap of technology. Westerman (2018) supports that no matter in which form, technology is only valuable if it allows the company to do business in new ways.

Big data's influence upon strategy is directly related, depending on how close to the business model is big data analytics capability within that company. In this regard, Mazzei and Noble (2017) identify three tiers of value creation with the help of big data. The first one is big data as a tool in the traditional value chain. These organizations use big data analytics in order to drive efficiency on their operations. For the second tier, big data is a stimulus for new ventures and industry development. The business model of these companies is based on collecting and analyzing large amounts of data. They act as software or service providers for other companies for which the cost of developing an analytics capability would be too high. Tier three companies use big data capability as a driver for competitive strategy. These are learning organizations which adapt their strategy based on the knowledge generated by big data analyses. They put data at the core of their strategy development, as they recognize new opportunities informed by big data. Expansion is being treated as an opportunity to acquire new data sources. For this category of companies, data has an intrinsic value as it opens new doors for discovering new capabilities and new ways of doing business. Example of these companies are: Facebook, Google and Amazon. They are able to explore in new fields (e.g.: automotive industry, self-driving cars, virtual reality) without the need to immediately generate revenue

from it.

According to Woerner and Wixom (2015), organizations have the possibility to innovate their business models by using big data in two ways: data monetization and digital transformation. Data monetization means supplying the data collected to another party for which is valuable. Twitter has a business model like this, through which it sells the data collected to big data analytics companies. Another way in which this strategy works is when companies with similar activities choose to join together their data, because the larger the data the more representative and accurate it becomes. By digital transformation, companies extend their field of activity by connecting them to business ecosystems, networks of companies, customers and devices.

4. Scope and methodology of the research

Our analysis is performed in two phases. First, we did an exhaustive review of the literature within the strategic management domain related to big data analytics. We desired to identify the state of knowledge in this field, the implications of big data analytics upon strategy and possible directions for future research.

Secondly, we fulfilled a big data analysis on the articles published on Web of Science with the topic: "big data strategic management" during period 2010 - 2018. This will help us to gain insights from the available literature with the help of algorithms, and to test at work, the use of big data analytics methods. In this regard, we use two methods, one for identifying keywords from the body text of the published articles and the second to execute an automatic text summarization with the help of which we will extract the main phrases from these articles.

4.1. Data collection

The beginning of our research consisted in searching all the articles published on Web of Science during period January 2010 - May 2018 with the topic "big data strategic management". As a first bucket of results, we obtained a number of 168 articles. All these were downloaded and analyzed based on title, abstract and keywords so that the articles which did not treat the topics related to strategic management research together with big data were filtered out. As a result of this investigation we kept a number of 67 articles which met our results. A similar research methodology has been used by Sheng et al. (2017). In this regard the authors wanted to identify the level of awareness, in the management research literature, towards the topic of big data analytics.

After selection, the articles were sorted and classified based on multiple criteria: year of publication and sectors that were studied within the articles. After our classification, we merged the articles on the following sectors: healthcare, human resources, manufacturing, services, logistics, public sector and articles that were focused on strategic management without reference for the industry. The reason for this division is because we expected different roles that big data is playing based on industry in which the company is operating. The results of the sorting are presented in

Figure 1 and Figure 2.

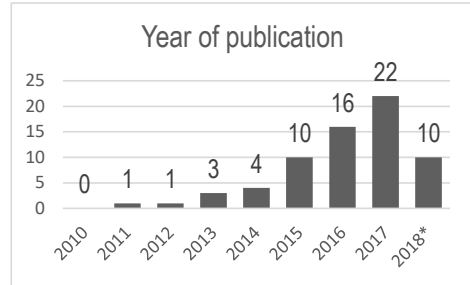


Figure 1. Year of publication for the analyzed articles

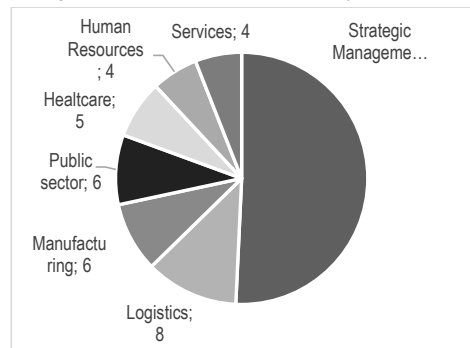


Figure 2. Distribution of the article topic based on industry

Excluding year 2018 for which we have available data only until month of May, we identified a continuous growth trend on average of 69% from year to year. Based on industry, we conclude that 34 from our total of 67 articles studied implications of big data in strategic management as a general field, while the rest of the papers had case studies which were focused on specific industries as resented in Figure 2. Because our searching criteria has been represented by strategic management articles which discuss the topic of big data, we consider it predictable the result of our finding, namely that 34 out of 67 articles treated this topic from a general perspective without the focus on a specific industry. However, the other half of results is focused on a specific sector as it follows: 8 articles on logistics sector, 6 articles on manufacturing, 6 articles on public sector applications, 5 articles on healthcare domain, 6 on human resources implications, and 4 articles are focused on implication upon services. This is relevant for our research because in our text summarization analysis we merged the body text of these articles based on their industry. The premise for this division is our hypothesis that big data has different implications depending on the sector in which the company belongs.

The text of the articles has been analyzed using Spyder 3 which is an open source platform for scientific programming in the Python language (Wikipedia, 2018).

After downloading, all the articles have been converted to .txt files in order to enhance text importing. Before importing the text, manual text cleansing has been performed, through

which we removed the following elements: references list, tables, figures and the headers and the footers of the articles.

We worked on text using the using the NLTK (Natural Language Package) package within Python, which offers a series of text processing libraries (NLTK, 2017).

Two main text analyses have been conducted. One to extract the trigrams and the four-grams from the body text compiled from the 34 articles which treated general aspects of strategic management and big data. The second analysis aimed to extract the most relevant sentences out of the articles merged based on each industry from our classification.

4.2. N-grams extraction

After importing the text, we did additional cleansing operations called in the practitioner’s field “text normalization” (Sarkar, 2016). The first one was to convert the entire text to small letters. This is needed because for example we want the text “Big Data” to be treated as equal with “big data”. Then, we removed all the text within parentheses operation allowed us to remove all the citations. Our algorithms are designed to identify as important the most frequent words. Because of this we would obtain as very important collocations of in the format “Author et al.”, “Author and Author” etc. Additionally, we removed the “et al” collocation because it has a high occurrence as well. Next, we removed the special characters these are punctuation marks and symbols. Another normalization operation is the removal of stop words. These are very common words such as “and”, “the”, “what” etc. which are removed because they provide little significance to the meaning of the text (Sarkar, 2016). The stop words dictionary that we used is part of the *nlk.corpus* called *stopwords.words* (English).

Once the data cleansing has been finished, we applied on our remaining text an algorithm built by Sarkar (2016) in order to obtain the bigrams and the trigrams from the text.

Table 1. Trigrams Frequency Distribution

Trigrams	Frequency
big data analytics	155
social media data	123
digital business strategy	120
value big data	87
use big data	64
open innovation strategies	35
using big data	35
big data analysis	34
big data open	33
data open innovation	32

Table 2. 4-Grams Frequency Distribution

4-Grams	Frequency
big data open innovation	32
realize value big data	28
customer insights social media	27
insights social media data	27
big data value realization	19
big data analytics capability	17
higher degree fit ba	16
degree fit ba organizational	15
organizations realize value big	15
value creation big data	13

N-grams are usually used to extract main topics out of large body texts. These would be more appropriate if the topics were unknown in advance so that the analysis will determine the main themes of the text. For our analysis, the results appear more like a confirmation of the main topic studied on the text composed out of the 34 articles which discussed big data analytics in the context of strategic management.

4.3. Key sentences extraction results

For the extraction of the key sentences from within the articles, the data cleansing process consisted in removing all the titles and the headings of the articles. In doing so we kept only the actual sentences of the articles. Following data cleansing, we run a text summarization algorithm which uses NLTK library. Part of the algorithm has been compiled using a source from The Glowing Python blog (2014). The algorithm will assign a score to each sentence based on how frequent its words are within the whole document. In other words, the sentence with the highest score will be the one that contains those words which aggregated are repeated the most within the document. The algorithm will ignore the bottom 10% words based on their frequency and the top 10%. The reason for the lower limit is because these words are insignificant to the entire document as they have a low occurrence, while the words above the higher limit are very common words, which can be considered stop words. Then the frequency of the words is computed and the top three highest ranking sentences are returned.

The full results of the analysis are presented within Appendices 1-6.

5. Findings

As presented in Figure 2., our bucket of articles contains five articles which are concerned with the strategic implications of big data within the healthcare sector. From the results presented in Appendix 1., we notice that the first two constructs of sentences returned by the algorithm belong to an article written by Wang et al. (2017). These two are relevant for the text as they are situated within the first paragraph of the introduction. We comprehend from here

that healthcare companies are making significant investments in order to build data infrastructure and to nurture analytics capabilities. Big data has positive impacts upon patient safety, decreases administrative costs as well as readmission rates. The third highest ranking sentence from the "healthcare" bucket appears at Wang et al. (2018) as the opening sentence of the introduction. It mentions the IT challenges healthcare sectors faces such as: weak integration of healthcare systems, and inefficient information management.

For the human resources category, we compiled 4 articles and obtained the results represented in Appendix 2. The highest-ranking sentence is located within the "Discussion" part at Wang and Cotton (2018). Their study used recorded data over a period of 111 years of all the baseball players from the teams in Major League Baseball (MLB) taking into consideration a performance indication specific to baseball called *wins above replacement* (WAR). WAR is a statistical measurement that reflects the contribution of a player to the team and is calculated as the number of extra wins that a team has because of a player compared to the number of anticipated wins if that player is replaced with a minor league player (Wang and Cotton, 2018). The abundance of data from Major League Baseball is a key factor in enhancing big data analytics for these organizations. Within the second ranking sentence we grasp two specificities of baseball teams: a high turnover and a cross-employer career path, which play an important role for the team's organizational experience ties.

Krscynski et al. (2018) own the third highest ranking sentence from the human resources group. The authors present the characteristics of a person with analytical abilities such as: ability to assess causality between organizational components, develop metrics for components of the system and implement an insights-based decision-making process.

Within the manufacturing sector we analyzed six articles (Appendix 3). A sentence from the first paragraph of the introduction from an article by O'Connor and Kelly (2017) is the highest ranking from this group. The authors looked at the capacity of Small to Medium-Sized Enterprises (SMEs) to use big data in order to gain competitive advantages. SMEs target customer data in order to gain insights which will enhance their understanding about consumer behavior. Then our result highlighted a contribution of Cheah and Wang's (2017) study which exemplified ways in which big data can be used to determine business model innovations within the traditional industries in China. The last one highlighted from this category is an article by Zheng et al. (2017) which proves that a lifecycle-oriented maintenance repair and overhaul (MRO) business model which is founded on big data provides benefits such as intelligent diagnosis, production optimization and prompt services.

Logistics sector summed up 8 articles and our key sentences extraction is presented in Appendix 4. The highest ranking sentence contains the searching formula used by Huang et al. (2018) which employed a methodology of extracting empirical insights from multiple documents from the academic field and from the practitioner's field with

the aim to build an improved technology delivery system (TDS). Blackburn et al. (2015) assign as a possible future research opportunity the task to "develop a modeling language that is appropriate to describe scheduling problems." Our next insight is related to the potential functions of big data analytics with supply chain management: vendor risk management, demand forecasting, logistic planning and scheduling, inventory and transportation management and reverse logistics (Roßmann et al. 2018).

Although the services category has grouped together 4 articles all the insights obtained originated from a single article (see Appendix 5). Bone et al (2015) proposed to test the impact of peer to peer problem-solving (P3) communities upon the need for traditional customer service provided by the company. Peer to peer problem solving communities are forums where customer will post questions related to their problems and other users will solve their queries. This phenomenon has a significant impact in reducing the use of traditional customer service provided by the company.

Finally, our sixth domain that we studied is the public sector (see Appendix 6). From the six articles compiled together, two have been highlighted with their highest-ranking sentences. The highest-ranking sentence belongs to an article by Drewer and Miladinva (2017) and it contains the definition and the characteristics of big data. According to the same authors, the Europol Regulation allows use of large quantities of data for strategic purposes (e.g. crime prevention and combating). Kourtit and Nijkamp (2018) identified the essential factors which an i-city dashboard should contain: efficient operations within the urban areas, concrete actions which should reflect the city vision, key performance indicators, based on real-time data, in order to assess effectiveness of the vision implementation.

6. Discussions and conclusions

Within this paper we aimed to review the current literature which treats the topic of big data analytics within the sphere of strategic management. By presenting the state of the state of knowledge we set the ground for future research. Big data analytics definitions, characteristics and practices have been presented. Now we understand that big data analytics has the potential to become a major industrial revolution. This creates a challenging situation for the strategic management first of all because of the uncertainty it generates and second because of all the resources, skills and activates that organizations need to put in place in order to acquire big data analytics.

Additionally, to reviewing the literature, we set to identify significant insights from the articles indexed on Web of Science on topic "strategic management big data" with the help of text analysis algorithms. In this regard we analyzed 67 articles through two operations.

First, we extracted trigrams and four-grams out of 34 articles containing a big data theme from a general strategic management point of view. The results of the three, respectively, four most frequent collocations came as a confirmation for our expectations regarding the most

approached topics within these articles.

Second, we extracted key sentences out of 33 articles about big data and strategic management grouped on the following sectors: healthcare, human resources, manufacturing, logistics, services and public sector. These analyses provided the focuses areas on each sector when it comes to big data.

The current study possesses a couple of limitations in the way that it examined only papers indexed on Web of Knowledge. Larger databases have the advantage of returning results that are more relevant to the topic. Second, there are only two text analysis tools used. There are other approaches such as: topic modeling, automated document summarization, clustering, sentiment analysis. All these can be used and could return unforeseen insights. Third, the sectors in which we applied our analyses are limited, by diversifying the research towards new industries, novel insights can be acquired.

All these limitations set the ground for future research.

References

- [1] Blackburn, R., Kallrath, J., & Klosterhalfen, S. T. (2015). Operations research in BASF's supply chain operations. *International Transactions in Operational Research*, 22(3), 385-405.
- [2] Cheah, S., & Wang, S. (2017). Big data-driven business model innovation by traditional industries in the Chinese economy. *Journal of Chinese Economic and Foreign Trade Studies*, 10(3), 229-251.
- [3] Chen, H., Chiang, R. H., & Storey, V. C. (2012). Business intelligence and analytics: from big data to big impact. *MIS quarterly*, 1165-1188.
- [4] Davenport, T. H., & Harris, J. G. (2007). *Competing on analytics: The new science of winning*. Harvard Business Press.
- [5] Davenport, T. H., & Kim, J. (2013). *Keeping up with the quants: Your guide to understanding and using analytics*. Harvard Business Review Press.
- [6] Davenport, Thomas H. "Competing on analytics." *Harvard Business Review* 84.1 (2006): 98.
- [7] Drewer, D., & Miladinova, V. (2017). The BIG DATA Challenge: Impact and opportunity of large quantities of information under the Europol Regulation. *Computer Law & Security Review*, 33(3), 298-308.
- [8] Drewer, D., & Miladinova, V. (2017). The BIG DATA Challenge: Impact and opportunity of large quantities of information under the Europol Regulation. *Computer Law & Security Review*, 33(3), 298-308.
- [9] Evans, James R., 1950-. (2017). *Business analytics: methods, models, and decisions*. Boston: Pearson,
- [10] George, G., Osinga, E. C., Lavie, D., & Scott, B. A. (2016). Big data and data science methods for management research. *Academy of Management Journal*, 59(5), 1493-1507.
- [11] Huang, Y., Porter, A. L., Cunningham, S. W., Robinson, D. K., Liu, J., & Zhu, D. (2017). *A technology delivery system for characterizing the supply side of technology emergence: Illustrated for Big Data & Analytics*. Technological Forecasting and Social Change.
- [12] Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities—An experiment on Stockholm. *Land Use Policy*, 71, 24-35.
- [13] Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M., & Russell, G. (2018). Analytical abilities and the performance of HR professionals. *Human Resource Management*, 57(3), 715-738.
- [14] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. *META Group Research Note*, 6(70).
- [15] Mazzei, M. J., & Noble, D. (2017). Big data dreams: A framework for corporate strategy. *Business Horizons*, 60(3), 405-414.
- [16] McAfee, A., Brynjolfsson. (2012). Big data: the management revolution. *Harvard business review*, 90(10), 60-68.
- [17] Miele, S., & Shockley, R. (2013). *Analytics: The real-world use of big data*. IBM Institute for Business Value.
- [18] Mikalef, P., Pappas, I. O., Krogstie, J., & Giannakos, M. (2017). Big data analytics capabilities: a systematic literature review and research agenda. *Information Systems and e-Business Management*, 1-32
- [19] O'Connor, C., & Kelly, S. (2017). Facilitating knowledge management through filtered big data: SME competitiveness in an agri-food sector. *Journal of Knowledge Management*, 21(1), 156-179.
- [20] Porter, M. E. (November–December 1996). What Is Strategy? *Harvard Business Review*, 61–78.
- [21] Roßmann, B., Canzaniello, A., von der Gracht, H., & Hartmann, E. (2017). *The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study*. Technological Forecasting and Social Change.
- [22] Sarkar, D. (2016). *Text Analytics with Python: A Practical Real-world Approach to Gaining Actionable Insights from Your Data*. Apress.
- [23] Seddon, J. J., & Currie, W. L. (2017). A model for unpacking big data analytics in high-frequency trading. *Journal of Business Research*, 70, 300-307.
- [24] Sharda, R., Delen, D., & Turban, E. (2013). *Business intelligence: a managerial perspective on analytics*. Prentice Hall Press.
- [25] Sheng, J., Amankwah-Amoah, J., & Wang, X. (2017). A multidisciplinary perspective of big data in management research. *International Journal of Production Economics*, 191, 97-112.
- [26] Sivarajah, U., Kamal, M. M., Irani, Z., & Weerakkody, V. (2017). Critical analysis of Big Data challenges and analytical methods. *Journal of Business Research*, 70, 263-286.
- [27] Van Rijmenam, M. (2013). Why the 3v's are not sufficient to describe big data. Big data startup, <http://www.bigdata-startups.com/3vs-sufficient-describe-big-data>.
- [28] Wang, L., & Cotton, R. (2018). Beyond Moneyball to

social capital inside and out: The value of differentiated workforce experience ties to performance. *Human Resource Management*, 57(3), 761-780.

[29] Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. *Journal of Knowledge Management*, 21(3), 517-539.

[30] Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. *Technological Forecasting and Social Change*, 126, 3-13.

[31] Westerman, G. (2018). Your Company Doesn't Need a Digital Strategy. *MIT Sloan Management Review*, 59(3), 1-5.

[32] Wikipedia, (2014): Spyder (software): [https://en.wikipedia.org/wiki/Spyder_\(software\)](https://en.wikipedia.org/wiki/Spyder_(software))

[33] Woerner, S. L., & Wixom, B. H. (2015). Big data: extending the business strategy toolbox. *Journal of Information Technology*, 30(1), 60-62.

[34] Zhang, Y., Ren, S., Liu, Y., Sakao, T., & Huisingh, D. (2017). A framework for Big Data driven product lifecycle management. *Journal of Cleaner Production*, 159, 229-240.

[35] Zikopoulos, P. C., Eaton, C., DeRoos, D., Deutsch, T., & Lapis, G. (2012). *Understanding big data*. New York et al: McGraw-Hill, 5(8).

[36] <https://www.gartner.com/it-glossary/big-data/>

[37] <https://glowingpython.blogspot.com/2014/09/text-summarization-with-nltk.html>

[38] <http://www.nltk.org/>

[39] <http://upxacademy.com>

Appendix 1. Healthcare Articles

Sentences	Source
"BA is consist of a number of different analytics techniques such as descriptive, predictive, and prescriptive analytics (Delen, 2014) and visualization techniques that use to extract meaningful information from big data. BA also holds the potential to help transform the healthcare system (Chen et al., 2012; Watson, 2014) and to address the cost and service quality pressures, including the reduction of readmission rate (Bardhan et al., 2015) and management cost (Bates et al., 2014), and the improvement of patient safety (Ferranti et al., 2010)."	Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. <i>Journal of Knowledge Management</i> , 21(3), 517-539.
"A large amount of clinical data being produced by various healthcare systems (e.g., electronic health records system, pharmaceutical R&D reporting system, and insurance claims/billing system) galvanizes healthcare organizations toward making considerable investments in business analytics (BA) to strengthen their data governance and analysis	Wang, Y., & Byrd, T. A. (2017). Business analytics-enabled decision-making effectiveness through knowledge absorptive capacity in health care. <i>Journal of Knowledge Management</i> , 21(3), 517-539.

capabilities (Ghosh and Scott, 2011; Ferranti et al., 2010; Raghupathi and Raghupathi, 2014)."	
"Information technology (IT)-related challenges such as inadequate integration of healthcare systems and poor healthcare information management are seriously hampering efforts to transform IT value to business value in the U.S. healthcare sector (Bodenheimer, 2005; Grantmakers In Health, 2012; Herrick et al., 2010; The Kaiser Family Foundation, 2012)."	Wang, Y., Kung, L., & Byrd, T. A. (2018). Big data analytics: Understanding its capabilities and potential benefits for healthcare organizations. <i>Technological Forecasting and Social Change</i> , 126, 3-13.

Appendix 2. Human Resources Articles

Sentences	Source
"In terms of workforce analytics, this study is notable in that we utilized the innovative measure of WAR, a sophisticated, call inclusive means of capturing talent contribution to team results that is customized by strategic capability area and which allows realistic and objective comparison of individual performance across industry firms and over time (Baumer & Zimbalist, 2014). Second, this study also took advantage of MLB's long history of capturing performance outcome data on each and every strategic chance as it applies to its core strategic capability areas, again cementing this industry role in being at the forefront of workforce analytics."	Wang, L., & Cotton, R. (2018). Beyond Moneyball to social capital inside and out: The value of differentiated workforce experience ties to performance. <i>Human Resource Management</i> , 57(3), 761-780.
"In baseball, it would seem especially important to monitor support team organizational experience ties because with high turnover and cross-employer career paths (Sonnenfeld & Peiperl, n1988), support team members may become quick successors to the strategic team (Becker et al., 2009) or they may be traded to other teams or demoted to make roster space for a newly needed support team member, possibly causing significant changes to the support team's organizational experience profile."	Wang, L., & Cotton, R. (2018). Beyond Moneyball to social capital inside and out: The value of differentiated workforce experience ties to performance. <i>Human Resource Management</i> , 57(3), 761-780.
"We thus define an individual's analytical ability as that individual ability to (a) develop causal logic connecting critical components of the organizational system, (b) leverage appropriate analytics to test causal relationships in the data, (c) ensure appropriate measures for the components of the system, and (d) ensure a process for incorporating insights into organizational decision making."	Kryscynski, D., Reeves, C., Stice-Lusvardi, R., Ulrich, M., & Russell, G. (2018). Analytical abilities and the performance of HR professionals. <i>Human Resource Management</i> , 57(3), 715-738.

Appendix 3. Manufacturing Articles

Sentences	Source
"It specifically examines the ability of Small to medium-sized businesses (SMEs) to overcome barriers to big data use, to enable SMEs to engage with big data, and to impact positively on small business competitiveness. The big data of interest is that of big data consumer analytics, defined as the extraction of hidden insights about consumer behavior from big data and the exploitation of that insight through advantageous interpretation (Erevelles et al., 2015, p.897)."	O'Connor, C., & Kelly, S. (2017). Facilitating knowledge management through filtered big data: SME competitiveness in an agri-food sector. <i>Journal of Knowledge Management</i> , 21(1), 156-179.
"Second, our study has contributed to the theoretical lens of value by contextualizing the value components of a business model (Bowman and Ambrosini,2000) and providing an integrated framework of big data-driven business model innovations comprising perspectives, business model processes and big data-driven business model innovations."	Cheah, S., & Wang, S. (2017). Big data-driven business model innovation by traditional industries in the Chinese economy. <i>Journal of Chinese Economic and Foreign Trade Studies</i> , 10(3), 229-251.
"With the transformation of manufacturing and service mode, and application of lifecycle-oriented MRO business model based on Big Data analytics, an intelligent and comprehensive health management for the units can be achieved by the case company (Shaangu, 2014)."	Zhang, Y., Ren, S., Liu, Y., Sakao, T., & Huisingh, D. (2017). A framework for Big Data driven product lifecycle management. <i>Journal of Cleaner Production</i> , 159, 229-240.

Appendix 4. Logistics Articles

Sentences	Source
"The final search strategy we applied was: TS = ((“Big Data” or Bigdata) OR ((Big Near/1 Data or Huge Near/1 Data) or “Massive Data” or “Data Lake” or “Massive Information” or “Huge Information” or “Big Information” or “Large-scale Data” or Petabyte or Exabyte or Zettabyte or “Semi-Structured Data” or “Semi structured Data” or “Unstructured Data”) AND (“analytic” OR “analyze” OR “analys*”))."	Huang, Y., Porter, A. L., Cunningham, S. W., Robinson, D. K., Liu, J., & Zhu, D. (2017). A technology delivery system for characterizing the supply side of technology emergence: Illustrated for Big Data & Analytics. <i>Technological Forecasting and Social Change</i> .
"In the first place, it would be an important research task to develop a modeling language that is appropriate to describe scheduling problems. Similar to algebraic modeling languages, which use the key objects variables, constraints, and objective functions, and are thus ideal to implement mathematical optimization problems, the development of a scheduling language would be required that maps the key objects units, resources, states, tasks, network	Blackburn, R., Kallrath, J., & Klosterhalfen, S. T. (2015). Operations research in BASF's supply chain operations. <i>International Transactions in Operational Research</i> , 22(3), 385-405.

Sentences	Source
topology, logical relations, etc."	
"Potential applications of BDA in SCM are manifold and include, inter alia, supplier risk management (Rajesh, 2016; Rajesh and Ravi, 2015; Sanders, 2016), demand forecasting (Chong et al., 2017; Huang et al., 2014; Jun et al., 2014; Sagaert et al., 2016; Yu and Tseng, 2014), logistics planning and scheduling (Zhong et al., 2015), inventory and transportation management (Bertsimas et al., 2016; Huang and Van Mieghem, 2014; Waller and Fawcett, 2013) and reverse logistics (Morgan et al., 2016)."	Roßmann, B., Canzaniello, A., von der Gracht, H., & Hartmann, E. (2017). The future and social impact of Big Data Analytics in Supply Chain Management: Results from a Delphi study. <i>Technological Forecasting and Social Change</i> .

Appendix 5. Services Articles

Sentences	Source
"We found that when customer motivation to participate in the P3 was problem resolution, customers were more satisfied with the P3 community than when customers motivation was knowledge acquisition or social. In sum, the robustness of our results, when accounting for multiple forms of heterogeneity as well as customer motives, increases our confidence in the results of the study."	Bone, S. A., Fombelle, P. W., Ray, K. R., & Lemon, K. N. (2015). How customer participation in B2B peer-to-peer problem-solving communities influences the need for traditional customer service. <i>Journal of Service Research</i> , 18(1), 23-38.
"Our model tests the effect of customer P3 community participation, static knowledge search, community log-in frequency, and membership breadth on the customer's future use of traditional customer support service provided by the firm, while controlling for previous traditional customer support service behaviors and community expertise (see Figure 1)."	Bone, S. A., Fombelle, P. W., Ray, K. R., & Lemon, K. N. (2015). How customer participation in B2B peer-to-peer problem-solving communities influences the need for traditional customer service. <i>Journal of Service Research</i> , 18(1), 23-38.
"After controlling for past traditional customer support service behavior and customer community expertise, we demonstrate that the problem-solving activities of helping oneself (posting questions) and helping others (responding to questions) in a peer-to-peer problem-solving community were significant predictors and primary drivers of reducing the customer's use of traditional customer support service."	Bone, S. A., Fombelle, P. W., Ray, K. R., & Lemon, K. N. (2015). How customer participation in B2B peer-to-peer problem-solving communities influences the need for traditional customer service. <i>Journal of Service Research</i> , 18(1), 23-38.

Appendix 6. Public Sector Articles

Sentences	Source
"The term big data describes a phenomenon in which vast amounts of	Drewer, D., & Miladinova, V.

<p>information are collected, pooled, and analyzed to provide empirical insights that help explain the present and the past, and predict the future. Although it lacks a universally accepted definition, the term has been used to describe very large data sets that are difficult to filter, analyze and interpret using standard database management tools or traditional data processing applications. As pointed, out in the 2014 EDPS Preliminary Opinion on the topic, currently, 57% of businesses analyzed consider themselves to be managing big data in the sense of very large datasets which can include streaming data from machines, sensors, web applications and social media."</p>	<p>(2017). The BIG DATA Challenge: Impact and opportunity of large quantities of information under the Europol Regulation. Computer Law & Security Review, 33(3), 298-308.</p>	<p>available sources, particularly sources on the internet and clear overview of trends."</p>	
<p>"Under the Europol Regulation, in light with the objectives stipulated in paragraphs 3, 12, 13, n23, 24 and 42 of the Preamble, it can be implied that large quantities of information may potentially be used in cases of strategic nature if necessary and proportionate data protection safeguards are in place so as to legitimize the overarching purpose and aim of Europol business objectives.⁶³ These provisions talk about the purposes of preventing and combating crime, fullest and most up-to-date information possible, improve Europol effectiveness, new technologies to process data, cross-matches of data, information obtained from publicly</p>	<p>Drewer, D., & Miladinova, V. (2017). The BIG DATA Challenge: Impact and opportunity of large quantities of information under the Europol Regulation. Computer Law & Security Review, 33(3), 298-308.</p>	<p>"The essential key factors of the i-city dashboard are: (i) an efficient operation of a sustainable urban area by addressing the challenges of complex urban systems for anticipating changes and managing policies, strategies and related actions in urban areas (see Kourtit, 2015); n(ii) the effective translation and strategic alignment between an established city vision for the future and related plans to be implemented on the one hand, and priorities on strategic goals, critical conditions and predefined key performance indicators (KPIs) on the other hand; n(iii) the intelligence (making vision and strategy understandable via smart KPIs) to pull n(real-time and digital) big data so as to unlock valuable insights from multiple data sources at various levels in the urban system (a multi-level information management approach), and to transform this into operational information and into concrete actions which stimulate (captured and created) high added-value outcomes leading to the highest city performance (similar to the previously defined XXQ concept introduced by Nijkamp, 2008, with a view to the achievement of the highest quality of urban life)."</p>	<p>Kourtit, K., & Nijkamp, P. (2018). Big data dashboards as smart decision support tools for i-cities—An experiment on stockholm. Land Use Policy, 71, 24-35.</p>