GRIFFITH UNIVERSITY

Exploring the relationship between big data and firm performance

Author

Almeida, Fernando, Low-Choy, Samantha

Published

2021

Journal Title

Management Research and Practice

Version

Version of Record (VoR)

Rights statement

© 2021 Research Centre in Public Administration and Public Services, Bucharest University of Economic Studies. The attached file is reproduced here in accordance with the copyright policy of the publisher. Please refer to the journal's website for access to the definitive, published version.

Downloaded from

http://hdl.handle.net/10072/415698

Link to published version

http://mrp.ase.ro/no133/nr133f.htm

Griffith Research Online

https://research-repository.griffith.edu.au

EXPLORING THE RELATIONSHIP BETWEEN BIG DATA AND FIRM PERFORMANCE

Fernando ALMEIDA
ISPGAYA, University of Porto & INESC TEC, Porto, Portugal
almd@fe.up.pt

Samantha LOW-CHOY
Arts, Education and Law Group, Griffith University, Queensland, Australia
Environmental Futures Research Institute, Griffith University, Queensland, Australia
s.low-choy@griffith.edu.au

Abstract
Big data offers great potential to improve organizational performance and generate competitive advantages. In this sense, knowing this phenomenon is relevant and this study aims to explore the role of big data on firm performance through a process of synthesis of several studies that have been published in recent years in different organizational contexts. This study adopts a mixed-methods approach. Initially, a systematic literature review is performed to characterize the studies that adopt structural equation modeling to determine the relationship between big data and firm performance. Additionally, a meta-analysis method is used to quantify the association between these two phenomena. The findings reveal a moderate positive relationship between the adoption of big data in the firm performance. This ratio is estimated at 0.38 with a confidence interval between 0.32 and 0.44 for a significance level of 0.05. The results of this study also allow us to conclude that the performance of organizations is also determined by other factors such as human capital, the data-driven organizational culture, or the learning capacity of the organization. This study offers mainly implications for companies that intend to invest in big data to know the potential value of this technology in organizational performance.
Keywords: big data capabilities, performance, business value, competitive advantage, structural equation modeling

1. INTRODUCTION

Technological evolution and the consequent increase in the dependence of society and organizations on the Internet have led to an exponential growth in the volume and variety of data in recent years. Around 4 petabytes of data are generated daily on Facebook, 4 terabytes of data created by each networked car, 65 billion messages in WhatsApp, and 294 billion email messages (Desjardins, 2019). By the end of 2025, the World Economic Forum estimates 463 exabytes of data will be generated daily (Desjardins, 2019). This growth was mainly due to the growth of the Internet of Things (IoT) and the expansion of social networks and cloud computing, which caused an increase in both the number of devices connected to the Internet and the number of users (Dai et al., 2020; Islam & Reza, 2019; Kumar et al., 2019). These projections can still be considered excessively conservative, as the recent COVID-19 pandemic could exponentially increase the number of people teleworking and accelerate business digitalization (Berger, 2020). All these changes will necessarily require a paradigm shift in organizations, with the competitiveness of the market requiring the storage of a large volume of information from a variety of sources.

Big Data arises in this context and has been used to describe the exponential growth of data. This concept is related to a set of technologies, processes and practices that enable organizations to gain insights and improve decision-making processes from the analysis of the large volume of data (Almeida, 2018; Yaqoob et al., 2016). The concept of Big Data has evolved over time involving increasingly new perspectives. In 2001, Laney (2001) presented the perspective of data exploration in three dimensions: (i) volume, velocity, and variety. However, several authors have added other perspectives such as the relevance of knowledge exploration from the data (Manyika et al., 2011), the relevance of adopting Big Data in decision-making processes (Gartner, 2012), the truthfulness of the data to reduce the risks of uncertainty (Schroeck et al., 2012), or the importance of data distribution through the cloud to allow distributed processing (Wang et al., 2013). This evolution of the Big Data concept has led to the acceptance that currently Big Data consists of 10 dimensions: volume, velocity, variety, value, veracity, visualization, variability, validity, vulnerability, and volatility (Khan et al., 2018).

Management Research and Practice
Volume 13, Issue 3 / September 2021

ISSN
2067- 2462

Organizations seek to explore and seize the opportunities that Big Data Analytics can bring to the organizational performance of companies. Mazumdar et al. (2019) highlight that Big Data technologies offer significant cost advantages because they offer distributed storage of large volumes of data. This helps companies to identify new business opportunities more efficiently and faster. According to Hajli et al. (2020), the speed and agility provided by Big Data Analytics enable the analysis of a large diversity of data, both structured and unstructured. Furthermore, data can be collected from a wide variety of sources, such as Enterprise Resource Planning (ERP), Supply Chain Management (SCM), Customer Relationship Management (CRM), social networks, web pages, etc. (Holmlund et al., 2020; Maheshwari et al., 2021). Data emerges as a key element of the information society, but only generates business value if properly treated and analyzed.

Big Data changes not only technology and management processes but also strategic directions and culture within organizations. Nowadays, it is not possible to think and create business strategies in the same way as before, once this new resource is available and offers a whole range of new perspectives that can be exploited by companies. Several authors emphasize the role of big data in organizational performance (Maroufkhani et al., 2019; Mikalef et al., 2019; Müller et al., 2018). These studies use both quantitative and qualitative methodologies and mixed methods. Despite the growing interest in this phenomenon, the quantitative impact of the relationship between big data and firm performance is not fully consensual. Studies exploring this phenomenon quantitatively using Structural Equation Modelling (SEM) establish a relatively heterogeneous association between these two variables depending on the context in which it is performed (e.g., country, sector of activity, number of firms involved). In this sense, this study intends to correct this gap and seeks to characterize the quantitative impact between the adoption of big data and firm performance. For this purpose, a prior Systematic Literature Review (SLR) of studies using the SEM method is conducted to explore quantitatively the relationship between the two variables and, then, a Meta-analysis (MA) process is adopted to quantitatively determine the relationship between them. Our approach follows the recommendations proposed by Silva-Fernández & Carmona (2019), in which it is suggested the joint adoption of the SLR and MA to explore the contributions given by big data and also as a way to eliminate the noise caused by individual studies in the field.

This study is organized as follows: Initially, the phases carried out in the process of systematic review and meta-analysis are presented. After that, the results are presented. In the first phase, the PRISMA diagram of the systematic review is presented and the quality of the studies included in this review is assessed. After that, the relationship between big data and firm performance is explored through a meta-analysis process. Then, the results obtained are discussed considering the existing literature in the field. Finally, the conclusions are listed, and the limitations of the study are also discussed and some topics for future work are suggested.

2. METHODOLOGY

The research methodology is presented in the next sub-sections. First, the stages of the systematic literature review process are presented. In this process, the research protocol, the adopted databases and the strategy of the research process, and the inclusion and exclusion criteria considered are highlighted. In a second phase, the meta-analysis process is presented.

2.1. Systematic literature review

SLR is a type of scientific research that brings together several original studies, synthesizing the results through strategies that limit bias and random errors. RLS is a retrospective, secondary study, which aims to gather similar published studies, and in which its quality can be critically evaluated. According to Booth et al. (2016), SLR has contributed to research planning in the area and helps in the decision-making process. SLR can also be combined with statistical methods to analyze and summarize the results of included studies.

Because an SLR is systematic it must be planned. To this end, Booth et al. (2016) state that the evidence to be included should be clearly indicated, the search area should be indicated and delimited, the terms that are used in the search process should be pointed out, subjective apprehension of the observed facts should be avoided, and the field and chronology of the review should be indicated. Furthermore, the SLR should be reproducible. Therefore, in contrast to the narrative or traditional literature review, the SLR responds to welldefined research questions and is characterized as methodologically comprehensive, transparent, and replicable.

Booth et al. (2016) refer to the existence of four essential criteria for an SLR: (i) it must be comprehensive; (ii) it must adopt a rigorous methodology; (iii) it must be comprehensive; and (iv) it must be conducted by at least two people. In line with these elements, research issues were previously defined, scientific databases were used, a protocol was adopted, and the process involved three researchers with an interdisciplinary background in the field of information technology, and research methodologies.

The SLR protocol is a document that can be used to formalize the primary study and define, monitor, and record all the steps to be performed in the secondary systematic review study. This study adopted the guidelines established by the Cochrane Handbook for SLR, in which the following steps are outlined: (i) formulation of research questions; (ii) search and selection of studies; (iii) data collection; (iv) critical evaluation of studies; (v) analysis and presentation of results; (vi) interpretation of results; and (vii) publication. The adoption of this protocol is fundamental to ensure the robustness of the process and its eventual replication.

In establishing the research questions, the guidelines proposed by Snyder (2019) were considered, which stipulate that research questions defined from an SLR should be specific and should provide a synthesis and comparison of the evidence. In accordance with these guidelines, the following research questions have been defined: (i) how many publications are there each year? (ii) how many citations per year? (iii) what is the industry focus of the studies? (iv) what is the size of the companies where the studies were conducted? (v) what is the distribution of the studies considering their geographical area of applicability? (vi) what is the quantitative estimation of the relationship between Big Data and Firm Performance?

Two databases were used for researching scientific articles: Web of Science (ISI) and Scopus. These two databases were used due to their high credibility and acceptance in the international scientific community. In addition to these two databases, the Association for Computing Machinery (ACM) and the Institute of Electrical and Electronic Engineers (IEEE) Xplore Digital Library were also initially considered due to their high acceptance in the field of computer science. However, it was found that this step would be unnecessary since all articles previously indexed in ISI and Scopus were also included in these two databases.

In the search for relevant articles several search terms were used: (i) "Big Data" AND "Structural Equation Modelling" AND "firm performance"; (ii) "Big Data" AND "Structural Equation Modelling" AND "business performance"; (iii) "Big Data" AND "Structural Equation Modelling" AND "organizational performance"; (iv) "Big Data" AND "Regression" AND "firm performance"; (v) "Big Data" AND "Regression" AND "business performance"; and (vi) "Big Data" AND "Regression" AND "organizational performance". These search terms were applied to the title, abstract, and keywords of each paper.

The following elements were considered as inclusion criteria: (i) articles only in English; (ii) articles with peer review; (iii) articles must be published by a reputable journal indexed in ISI or Scopus; and (iv) the full text of the article must be available online. On the opposite side, the elements below were considered exclusion criteria: (i) duplicate articles; (ii) articles that do not use the SEM method; (iii) articles outside the big data and firm performance scope; and (iv) articles published in 2020, as they could bias the statistical analysis since at the time of this study the first half of 2020 had not yet been completed.

Finally, there is a growing concern about the methodological quality of studies based on RLS. Jagannath et al. (2011) point out that more than 24 quality assessment instruments are available, but not all of them provide a rigorous analysis of the adopted process. One of the most recognized methods that have presented good evidence of construct validity and reliability is the Assessment of Multiple Systematic Reviews (AMSTAR). AMSTAR v2 is an instrument composed of 11 items in a question format that allow four types of answers: yes, no, I cannot answer, and no applicable (Shea et al., 2017). This set of answers enables the identification of failures and a critical analysis of the review process.

2.2. Meta-analysis

MA is a statistical technique used to combine data from multiple studies on a specific topic. Different individual studies are used to integrate them, combining, and summarizing their results. According to Borenstein et al.

(2011), this approach reduces the standard deviation and confidence interval, making the result more reliable. However, for this method to be applied, data must be groupable and standardized. The joint adoption of SLR with MA enables us to explore the accumulated evidence on the phenomenon and reduce the risk of biases (Ahn & Kang, 2018; Mengist et al. 2020). Additionally, MA allows exploring and answering questions not posed by individual studies and offers the opportunity to meet and resolve contradictory claims by these studies (Harris, 2019). The Review Manager (RevMan) software was used to maintain the systematic reviews and perform the meta-analysis. For this purpose, a confidence interval was determined for the estimation of the relationship between Big Data and firm performance considering fixed-effect estimates and random-effects estimates scenarios. In both cases, a forest plot of the risk ratio was performed.

3. RESULTS

A systematic review is a review of a clearly formulated set of one or more research questions that uses systematic and explicit methods to identify, select and critically evaluate relevant research, and collect and analyze data from those studies that are included in the review. In the process of analysis and systematization of results, PRISMA was adopted as recommended by Oláh et al. (2020) and Pati & Lorusso (2018). PRISMA presents the flowchart of the information collected and analyzed throughout the different phases of the systematic review. Figure 1 presents the records identified and the reasons for their inclusion and exclusion. In the first phase, Scopus and ISI databases were used to identify relevant papers published in the field considering previously established search criteria. A total of 167 publications were identified, of which 40 items were repeated because they are indexed in both databases. This resulted in 127 records after duplicates removed. In a second phase, records were analyzed according to a two-tier model: (i) analysis of the publications abstract, which resulted in the identification of 90 relevant publications; and (ii) analysis of the full-text articles, which resulted in 26 publications. Several reasons were identified for the exclusion of the articles, namely the issue of big data is not the focus of the publication, exploration in the technical dimension of big data, and its focus is not on the performance of the organization but looks at the social and organizational impact of big data. Finally, in the 3rd phase, the studies were included in the quantitative synthesis. In this process, 8 records were excluded because the publications do not present enough quantitative information about the process of applying SEM.

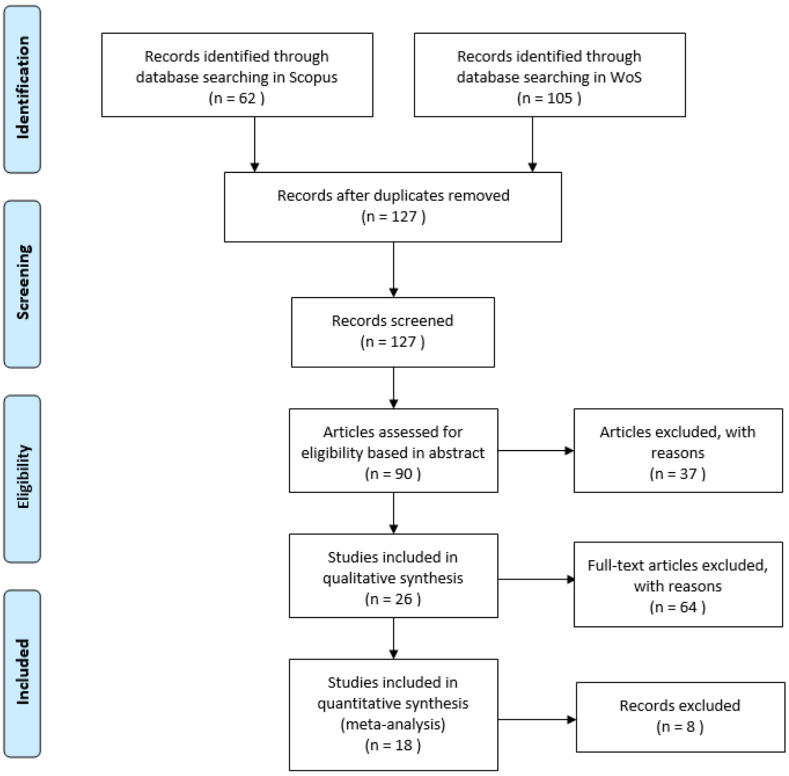

FIGURE 1 - PRISMA DIAGRAM

In order to assess the eligibility of the studies considered in the quantitative analysis, the following elements were identified for each study: (i) population description and setting; (ii) initial survey sample size; (iii) final sample size; (iv) response rate; (v) sampling method; (vi) survey instrument info; (vii) survey validation; (viii) time frame; (ix) risk of bias), (x) software for SEM; and (xi) funding source reported. Each study was evaluated in all these components and the critical flaws and non-critical weaknesses were identified. The same scale as the Assessment of Multiple Systematic Review (AMSTAR v.2) was adopted. This method establishes four global qualitative indicators to evaluate the eligibility of studies: (i) high - zero or one non-critical weakness; (ii) moderate - more than one non-critical weakness; (iii) low - one critical flaw with or without non-critical weaknesses; and (iv) critically low - more than one critical flaw with or without non-critical weaknesses. Only studies with a global evaluation of high or moderate were included in the SLR. Table 1 shows that two studies were excluded for presenting a low overall quality. In this sense, the final number of studies included was 16. Some non-critical weaknesses were also found mainly related to the lack of information on the initial sample and response rate, small sample size, or the focus on a specific industry only. However, these situations do not prevent the inclusion of these studies in the meta-analysis process.

TABLE 1 - ANALYSIS OF THE METHODOLOGICAL QUALITY

Study	Overall quality	Critical flaw	Non-critical weakness
Akhtar et al. (2019)	Moderate	N/A	Specific industry of agrifood Time frame not reported Software for SEM not reported
Akter et al. (2019)	Moderate	N/A	Response rate not reported Time frame not reported
Anwar et al. (2018)	Moderate	N/A	Survey validation mechanisms were not considered Time frame not reported
Caputo et al. (2019)	Moderate	N/A	Initial survey sample size not reported Response rate not reported Only high-tech sector companies involved
Chierici et al. (2019)	High	N/A	N/A
Ferraris et al. (2019)	Moderate	N/A	Time frame not reported Low size sample
Gupta et al. (2019)	Moderate	N/A	Initial survey sample size not reported Response rate not reported
Irfan & Wang (2019)	High	N/A	N/A
Kakhki & Palvia (2016)	Low	Survey instrument info and validation are not explicitly addressed	Initial survey sample size not reported Response rate not reported Time frame not reported
Lozada et al. (2019)	Low	Survey instrument info and validation are not explicitly addressed	Response rate is low
Mikalef et al. (2019)	High	N/A	Only large companies involved
Mishra et al. (2019)	High	N/A	Time frame not reported
Raut et al. (2019)	Moderate	N/A	Initial survey sample size not reported Response rate not reported Time frame not reported
Ren et al. (2017)	High	N/A	Time frame not reported
Shan et al. (2019)	High	N/A	N/A
Wamba & Akter (2019)	High	N/A	Time frame not reported
Wamba et al. (2017)	Moderate	N/A	Time frame not reported Only companies in e-commerce and m-commerce markets
Wamba et al. (2019)	High	N/A	N/A

The AMSTAR v.2 was adopted to evaluate the methodological quality of the systematic reviews included in the meta-analysis process (Shea et al., 2017). The AMSTAR v.2 presents a total of 16 items that allow evaluating the methodological quality as presented in Table 2. The authors' approach is presented for each question. AMSTAR v.2 does not seek to obtain a global measure but only to provide information on the quality of the studies included in the SLR.

TABLE 2 - AMSTAR V.2 QUESTIONS FOR ASSESSMENT

Questions	Approach
1. Did the research questions and inclusion criteria for the review include the components of PICO?	The inclusion criteria include population, intervention, comparator group, and outcome. Timeframe was also included.
2. Did the report of the review contain an explicit statement that the review methods were established prior to the conduct of the review and did the report justify any significant deviations from the protocol?	The established protocol includes the review questions, search strategy, inclusion/exclusion criteria, and risk of bias assessment. A meta-analysis plan was also included.
3. Did the review authors explain their selection of the study designs for inclusion in the review?	All the studies included in the review were explained and are assessed considering the inclusion criteria.
4. Did the review authors use a comprehensive literature search strategy?	Two databases were included: Scopus and ISI. The keywords and search strategy were presented. Publication restrictions were also presented.
5. Did the review authors perform study selection in duplicate?	Two independent reviewers were selected to identify eligible studies. Consensus was also established for all studies.
6. Did the review authors perform data extraction in duplicate?	Data were extracted from both reviews. Duplicated studies were identified and removed.
7. Did the review authors provide a list of excluded studies and justify the exclusions?	All studies excluded from the review were identified and compiled in a separate file. Justification for exclusion was also presented.
8. Did the review authors describe the included studies in adequate detail?	All included studies were assessed considering the population, outcomes, and research design.
9. Did the review authors use a satisfactory technique for assessing the risk of bias (RoB) in individual studies that were included in the review?	The risk of bias was identified for all studies. Studies that do not present risk of bias information were removed from the SLR.
10. Did the review authors report on the sources of funding for the studies included in the review?	The source of funding associated with each study was also registered. However, not all studies report this information. In these cases, we assumed that they did not get any financial support.
11. If meta-analysis was performed did the review authors use appropriate methods for statistical combination of results?	The SLR includes a meta-analysis process. The authors use the Review Manager software. The R square and the standard error were also determined. A 95% confidence interval was adopted.
12. If meta-analysis was performed, did the review authors assess the potential impact of RoB in individual studies on the results of the meta-analysis or other evidence synthesis?	The SLR only includes studies with high and moderate quality. Therefore, the potential impact of the risk of bias is low.
13. Did the review authors account for RoB in individual studies when interpreting/ discussing the results of the review?	The risk of bias is identified for each study. Its impact is assessed when interpreting the results of the review.
14. Did the review authors provide a satisfactory explanation for, and discussion of, any heterogeneity observed in the results of the review?	To decrease the heterogeneity of studies we only include SEM studies.
15. If they performed quantitative synthesis did the review authors carry out an adequate investigation of publication bias (small study bias) and discuss its likely impact on the results of the review?	The statistical tests were performed in the Review Manager. The forest and funnel plot were presented.
16. Did the review authors report any potential sources of conflict of interest, including any funding they received for conducting the review?	The authors report no competing interests.

Management Research and Practice
Volume 13, Issue 3 / September 2021

3.1. Distribution per year of publication

Figure 2 graphically presents the distribution of publications by year. 13 of the 16 publications were published in 2019, which represents 81.25% of the sample. In this sense, it can be observed that the analysis of the impact of big data on firm performance is an emerging theme.

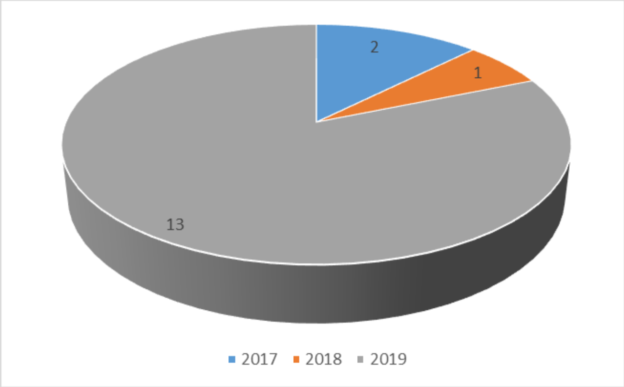

FIGURE 2 - DISTRIBUTION OF PUBLICATIONS PER YEAR

3.2. Number of citations

Figure 3 shows the number of citations for each study considered in the SLR. All studies have received citations. On average there are 64 citations per study. It is high considering that most studies were published in 2019. However, relevant asymmetries arise. Wamba et al. (2017) received a total of 654 citations, which is 10 times higher than the average number of citations. This study has received great recognition from the scientific community for being one of the main studies exploring the effects of big data analytics on firm performance.

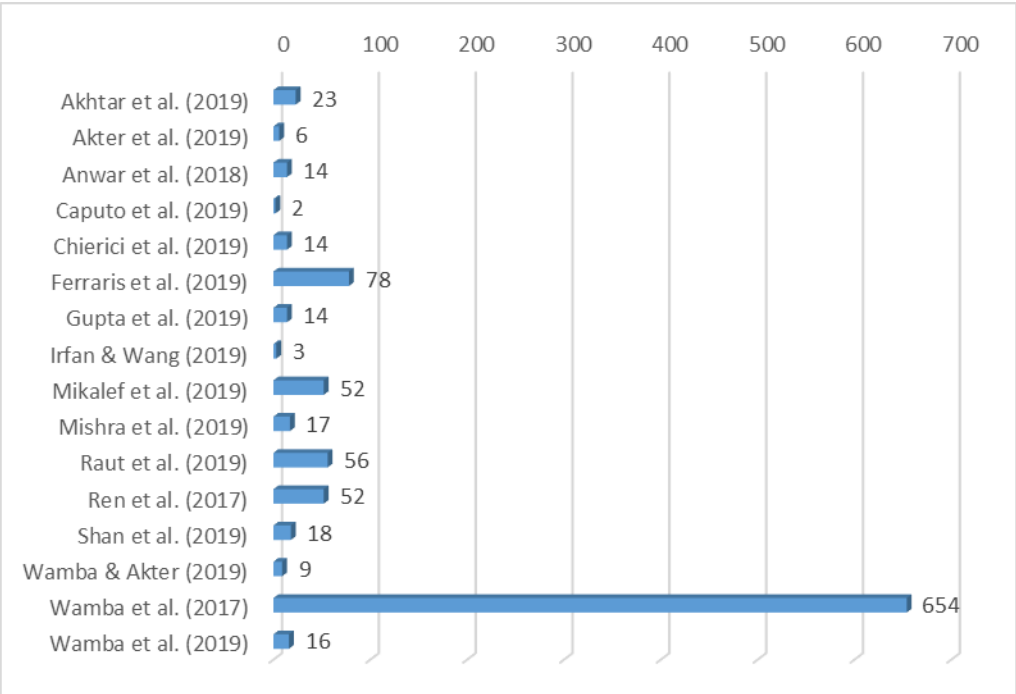

FIGURE 3 - NUMBER OF CITATIONS PER STUDY

Management Research and Practice
Volume 13, Issue 3 / September 2021

3.3. Industry focus, firm size and country-focus distribution

Table 3 presents the distribution of studies considering industry focus, firm size, and country focus. Most of the studies have a comprehensive perspective considering the size of the companies surveyed. Only in two of them, there is a focus on small and medium-sized enterprises (Ferraris et al., 2019) and on medium-sized and large firms (Anwar et al., 2018). The industry focus is presented as a more discriminating factor than the firm size. 9 of the 16 studies focus their analysis on a specific industry such as the service sector, supply chain, high-tech, e-business, among others. The distribution of the studies across countries is comprehensive by including both countries in Asia (i.e., China and India), North America (i.e., USA) and European Countries (i.e., France, Italy, among others).

TABLE 3 - ANALYSIS OF INDUSTRY FOCUS, FIRM SIZE, AND COUNTRY-FOCUS DISTRIBUTION

Study	**Industry focus**	**Firm size**	**Country focus**
Akhtar et al. (2019)	Agrifood	Several	New Zealand plus several European countries
Akter et al. (2019)	Service sector	Several	USA and France
Anwar et al. (2018)	E-business sector	Medium-sized and large firms	China
Caputo et al. (2019)	High-tech sector	Several	European countries
Chierici et al. (2019)	Companies with social presence	Several	Italy
Ferraris et al. (2019)	Several	Small and medium-sized enterprises	Italy
Gupta et al. (2019)	Service sector	Several	Not specified
Irfan & Wang (2019)	Food and beverage	Several	Pakistan
Mikalef et al. (2019)	Several	Several	Norway
Mishra et al. (2019)	Several	Several	India
Raut et al. (2019)	Several	Several	India
Ren et al. (2017)	Several	Several	China
Shan et al. (2019)	Several	Several	Not specified
Wamba & Akter (2019)	Supply chain sector	Several	USA
Wamba et al. (2017)	Online market sector	Several	China
Wamba et al. (2019)	Several	Several	France and USA

3.4. Estimation the relationship between big data and firm performance

The relationship between big data and firm performance was explored using the RevMan 5.3 software. The forest plot and funnel plot were drawn as depicted in Figure 4. The studies were organized in alphabetical order and loaded into RevMan 5.3. The effect of big data analytics on firm performance was estimated to be 0.38 (moderate and positive) considering a significance level of 0.05 (interval confidence of 0.32-0.44). Test for overall effect indicates Z equal to 12.18 (p less than 1.10-5). The Tau2 = 0.01, Chi2 = 12.18 and I2 = 64% which indicates that the meta-analysis performed adequately for α = 0.05.

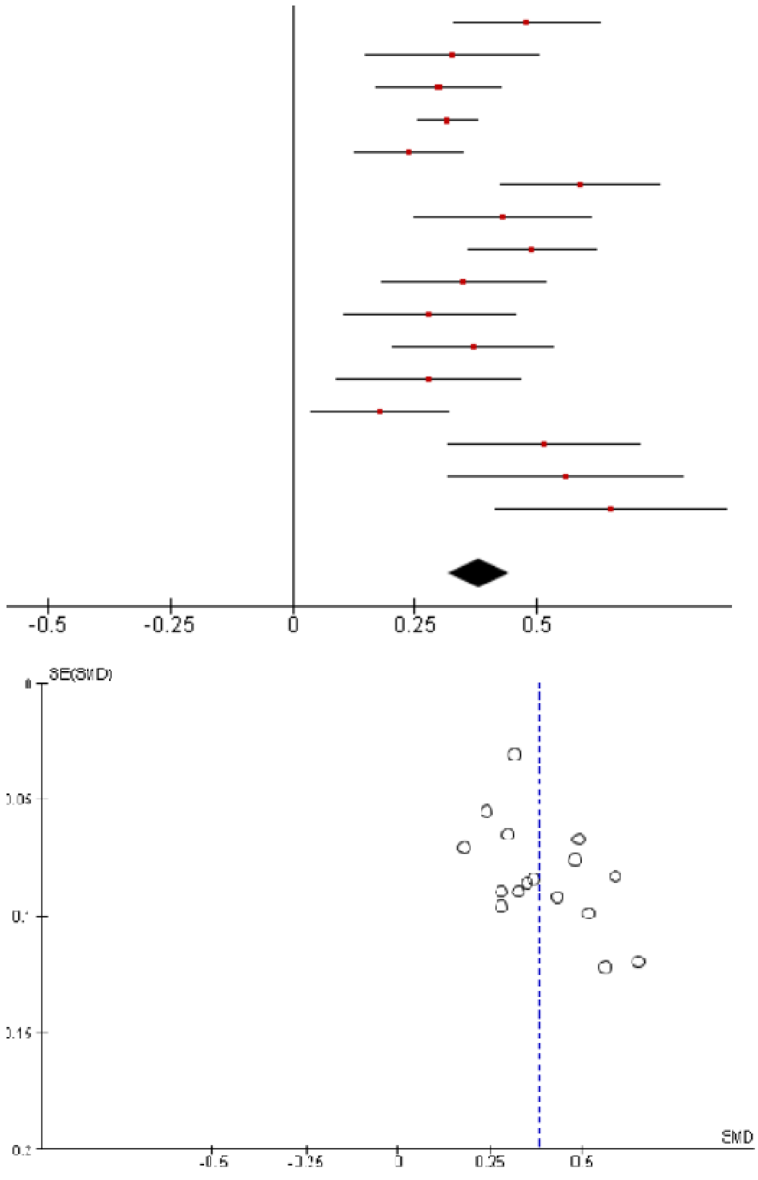

FIGURE 4 - FOREST PLOT AND FUNNEL PLOT

4. DISCUSSIONS

Big data has become an unquestionable asset for organizations. For this reason, big data has become one of the most discussed topics by researchers and experts in the field. Research in this area has focused in the first phase on the analysis of data distribution models, then moving on to large-scale data analysis models and the exploration of their benefits for organizational performance. The results of this study confirm this evidence by highlighting that 81.25% of studies exploring the relationship between big data and firm performance were published in 2019. The results of the study also allowed us to conclude that the importance of this phenomenon is addressed on a global scale involving both large corporations and small and medium-sized business companies.

The quantitative estimate of the relationship between big data and firm performance lies between 0.32 and 0.44 for a 95% confidence interval and was supported in 18 studies published between 2017 and 2019 that adopt the SEM technique to establish the relationship between the two phenomena. Although the estimation interval is relatively small it stands out studies that report an impact greater than 0.50 in Ferraris et al. (2019) and Wamba et al. (2019), while in other studies such as Chierici et al. (2019) and Shan et al. (2019) its impact is less than 0.25. A key potentiality of adopting the MA as reported by Harris (2019) is the identification of this divergent phenomenon from studies with various origins. The studies conducted by Chierici et al. (2019) and Shan et al. (2019) identify the innovative capabilities and dynamic capabilities of organizations as mediating elements and limiting the organizational performance of big data adoption. In Shan et al. (2019), it is explicit that organizational performance is more than twice affected by the strategic flexibility of organizations which involves a set of dynamic competencies of organizations such as the entrepreneurial leadership skills, the introduction of market disruptions, the creation of new innovative products/services, and new processes as highlighted in studies such as Ferreira et al. (2020), Sebhatu (2021), and Vu (2020) than IT technology capabilities. On the opposite, information quality reported in Wamba et al. (2019) and knowledge management highlighted in Ferraris et al. (2019) are elements that enhance greater organizational performance in companies through the adoption of big data.

Big data can be characterized as a multidisciplinary phenomenon that includes different dimensions such as technology, data science, and knowledge economics. The technology involves the collection, analysis, relationship, and comparison of large data sets through the distribution of information processing and maximization of computational capacity (Avci et al., 2020; Yu & Zhou, 2019). Data analysis includes the identification of standards for better decision making (Jeble et al., 2016; Sivarajah et al., 2017). Finally, it becomes equally relevant, through a large data set, to achieve greater knowledge that is capable of generating accurate value, which would otherwise be unreachable (Müller et al., 2018; Shabbir & Gardezi, 2020).

When attempting to define big data, the first characteristic that stands out is the high volume of data. However, in recent years, other elements have been added such as speed, variety, veracity, and value (Kitchin & McArdle, 2016; Lee, 2017; Ramasamy & Chowdhury, 2020). The value is a key element so that the big data can translate into a benefit for organizations. According to Günther et al. (2017), value determines the capacity of data to produce value for a given process or activity (i.e., the importance of being able to draw economic value from the data collected). Data analysis is currently a necessity for organizations that intend to extract insights from their business. Therefore, a big data solution must be able to work with all the content produced and received by the organization. As Ajah & Nweke (2019) point out, a big data solution should enable technical cost reduction, time savings, and the development of new products/services.

Although the various dimensions of big data are distinct, they are related and support each other to achieve the organizations' business objectives. Organizations need to have an equally multidisciplinary set of skills to exploit the potential of big data. One of these dimensions is the big data technology capability, which is the technological capacity of the big data analytics platform to perform cross-analysis of data coming from multiple sources and involving both structured, semi-structured, and unstructured data (Mishra & Misra, 2017). Big data management capability is another essential aspect to ensure that the big data platform generates organizational value. Akter et al. (2016) state that in its implementation elements such as investment, planning, coordination, and control should be considered. Finally, the big data talent capability emerges. This dimension refers to the ability to have at their disposal professionals with the knowledge and analytical skills to perform tasks in a high data volume environment (Persaud, 2020).

The results of this study show the simultaneous role of these three perspectives and establish a moderate positive relationship of big data in firm performance. This conclusion is also based on several studies that explore the effects of big data on organizations. Faroukhi et al. (2020) argue that big data is transforming value chains and business models themselves. Prescott (2016) states that the use of big data by companies has become an important way to maintain a competitive advantage and improve their performance. Ramadan et al. (2020) also confirm this vision and add that big data enables the creation of a new organizational paradigm, changing the value of experience, management practices, and the capabilities of leaders to use and analyze a large volume of data at their disposal.

The findings also suggest that the organizational performance obtained by adopting the big data is conditioned by several factors, which allows understanding the moderate effects of the big data on firm performance. Several challenges are posed to organizations. Data analysis is not only a technical challenge but an organizational one as highlighted by Braganza et al. (2017). This view is also confirmed by Gupta & George who found that organizations' capacity to exploit big data is the result of tangible, intangible, and human-related elements. Other elements emerge such as a data-driven organizational culture and a propensity for organizational learning (Gupta & George, 2016). The role of the data scientist is not limited to handling the technological challenges of big data but finding strategic solutions for disruptive transformations as organizations insert big data into their decisions and management.

5. CONCLUSIONS

The use of systematic literature review and meta-analysis has allowed exploring the phenomenon of the relationship between big data and firm performance and quantifying its relationship. This approach has proven to be appropriate given the large number of recent studies that have been published in the field that use to describe this phenomenon in different contexts, namely in various countries, size of companies, or sectors of activity. This manuscript has revealed until the end of 2019 the years of publication of each study, the number of citations, and the industry focus, firm size, and country-focus distribution. The relationship between big data and firm performance was estimated at 0.38 with a confidence interval between 0.32 and 0.44 considering a level of significance of 0.05.

It can be concluded that big data offers great potential to organizations in the creation of new businesses, the development of new products and services, and the improvement of business operations. The systematic review has allowed us to realize that firm performance is measured over multiple perspectives such as revenue, cost reduction, and market share. Therefore, companies that have been able to innovate or use innovation to offer a better product or service will achieve higher performance. However, the study also reveals the existence of other factors that condition the potential of an organization to exploit the benefits of big data. Accordingly, the ability of a company to exploit big data is not restricted to the technical dimension but includes other tangible, intangible, and human-related factors. In this vision, organizational performance is also determined by the organizational data-driven culture and the learning capacity of the organization.

This study offers both theoretical and practical contributions. In the conceptual dimension, it was able to explore the findings obtained by 18 studies published in the area on the relationship between big data and firm performance in a diverse set of countries and organizations. The synthesis work carried out allowed us to explore the methodological quality of these studies and, through a process of meta-analysis, it was possible to determine the quantitative relationship between these two phenomena. In the practical dimension, the results of this study are relevant for the organizations that are taking the first steps in the adoption of big data. The conclusions of the study establish that big data has a moderate positive impact on the firm performance, but other factors related to human, cultural, and organizational capital emerge that are equally determinant in the firm performance. This study presents some limitations that become relevant to the debate. First, there is a great heterogeneity of the studies considered in the systematic review and which are highly conditioned by the context in which they were conducted. Therefore, and as future work, it would become relevant to explore the role of control variables such as firm size, industry type, or the knowledge management orientation of organizations. Another limitation is that the firm performance was not ramified in its multiple dimensions as the financial performance, marketing performance, operational performance, and environmental performance. Exploring the role of each of these dimensions in organizational performance is another interesting area of future work. Furthermore, it is recognized that the impact of big data on financial performance may have mixed results. Therefore, and as a future work, we suggest some research questions that may help to understand and characterize this phenomenon, in particular: Why still some organizations find big data analytics capability not useful? How organizational culture may influence the effects of big data analytics on firm performance? Why different countries have different outcomes? How national cultural dimensions have an effect on the path joining big data analytics and firm performance?

REFERENCES

Ahn, E., & Kang, H. (2018). Introduction to systematic review and meta-analysis. *Korean Journal of Anesthesiology*, *71*(2), 103-112.

Ajah, I. A., & Nweke, H. F. (2019). Big Data and Business Analytics: Trends, Platforms, Success Factors and Applications. *Big Data and Cognitive Computing*, *3*(2), 1-30.

Akhtar, P., Frynas, J. G., Mellahi, K., & Ullah, S. (2019). Big Data-Savvy Teams' Skills, Big Data-Driven Actions and Business Performance. *British Journal of Management*, *30*(2), 252-271.

Akter, S., Wamba, S. F., Gunasekaran, A., Dubey, R., & Childe, S. J. (2016). How to improve firm performance using big data analytics capability and business strategy alignment? *International Journal of Production Economics*, *182*, 113-131.

Akter, S., Wamba, S. F., Barrett, M., & Biswas, K. (2019). How talent capability can shape service analytics capability in the big data environment? *Journal of Strategic Marketing*, *27*(6), 521-539.

Almeida, F. (2018). Big Data: Concept, Potentialities and Vulnerabilities. *Emerging Science Journal*, *2*(1), 1-10.

Anwar, M., Khan, S. Z., & Shah, S. Z. (2018). Big Data Capabilities and Firm's Performance: A Mediating Role of Competitive Advantage. *Journal of Information & Knowledge Management*, *17*(2), 1-28.

Avci, C., Tekinerdogan, B., & Athanasiadis, I. N. (2020). Software architectures for big data: a systematic literature review. *Big Data Analytics*, *5*(5), 1-53.

Berger, R. (2020). *How the current crisis will change the culture of our work*. Retrieved 12 January 2021, from https://www.rolandberger.com/en/Point-of-View/Digital-workplace-in-the-era-of-Covid-19.html

Booth, A., Sutton, A., & Papaioannou, D. (2016). *Systematic Approaches to a Successful Literature Review*. Thousand Oaks, California: SAGE Publications.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2011). *Introduction to Meta-Analysis*. Hoboken, New Jersey: Wiley.

Braganza, A., Brooks, L., Nepelski, D., Ali, M., & Moro, R. (2017). Resource management in big data initiatives: Processes and dynamic capabilities. *Journal of Business Research*, *70*. 328-337.

Caputo, F., Cillo, V., Candelo, E., & Liu, Y. (2019). Innovating through digital revolution: The role of soft skills and Big Data in increasing firm performance. *Management Decision*, *57*(8), 2032-2051.

Chierici, R., Mazzucchelli, A., Garcia-Perez, A., & Vrontis, D. (2019). Transforming big data into knowledge: the role of knowledge management practice. *Management Decision*, *57*(8), 1902-1922.

Dai, H. N., Wang, H., Xu, G., Wan, J., & Imran, M. (2020). Big data analytics for manufacturing internet of things: opportunities, challenges and enabling technologies. *Enterprise Information Systems*, *14*(9-10), 1279-1303.

Desjardins, J. (2019). How much data is generated each day?. *World Economic Forum*. Retrieved 19 Novemeber 2020, from https://www.weforum.org/agenda/2019/04/how-much-data-is-generated-each-day-cf4bddf29f/

Faroukhi, A. Z., Alaoui, I. E., Gahi, Y., & Amine, A. (2020). Big data monetization throughout Big Data Value Chain: a comprehensive review. *Journal of Big Data*, *7*(3), 1-22.

Ferraris, A., Mazzoleni, A., Devalle, A., & Couturier, J. (2019). Big data analytics capabilities and knowledge management: impact on firm performance. *Management Decision*, *57*(8), 1923-1936.

Ferreira, J., Cardim, S., & Coelho, A. (2020). Dynamic Capabilities and Mediating Effects of Innovation on the Competitive Advantage and Firm's Performance: the Moderating Role of Organizational Learning Capability. *Journal of the Knowledge Economy*, *12*, 620-644.

Günther, W. A., Mehrizi, M. H., Huysman, M., & Feldberg, F. (2017). Debating big data: A literature review on realizing value from big data. *The Journal of Strategic Information Systems*, *26*(3), 191-209.

Gupta, M., & George, J. F. (2016). Toward the development of a big data analytics capability. *Information & Management*, *53*(8), 1049-1064.

Gupta, S., Qian, X., Bhushan, B., & Luo, Z. (2019). Role of cloud ERP and big data on firm performance: a dynamic capability view theory perspective. *Management Decision*, *57*(8), 1857-1882.

Hajli, N., Tajvidi, M., Gbadamosi, A., & Nadeem, W. (2020). Understanding market agility for new product success with big data analytics. *Industrial Marketing Management*, *86*. 135-143.

Harris, C. (2019). *The Handbook of Research Synthesis and Meta-Analysis*. New York: Russell Sage Foundation.

Holmlund, M., Vaerenbergh, Y. V., Ciuchita, R., Ravald, A., Sarantopoulos, P., Ordenes, F. V., & Zaki, M. (2020). Customer experience management in the age of big data analytics: A strategic framework. *Journal of Business Research*, *116*, 356-365.

Irfan, M., & Wang, M. (2019). Data-driven capabilities, supply chain integration and competitive performance: Evidence from the food and beverages industry in Pakistan. *British Food Journal*, *121*(11), 2708-2729.

Islam, M., & Reza, S. (2019). The Rise of Big Data and Cloud Computing. *Internet of Things and Cloud Computing*, *7*(2), 45-53.

Jagannath, V., Mathew, J. L., Asokan, G. V., & Fedorowicz, Z. (2011). Quality assessment of systematic reviews of health care interventions using AMSTAR. *Indian Pediatrics*, *48*, 383-385.

Jeble, S., Kumari, S., & Patil, Y. (2016). Role of big data and predictive analytics. *International Journal of Automation and Logistics*, *2*(4), 307-331.

Kakhki, M. D., & Palvia, P. (2016). Effect of Business Intelligence and Analytics on Business Performance. *Proceedings of the Twenty-second Americas Conference on Information Systems*, San Diego, CA, 1-10.

Khan, N., Alsaqer, M., Shah, H., Badsha, G., Abbasi, A., & Salehian, S. (2018). The 10 Vs, Issues and Challenges of Big Data. *Proceedings of the 2018 International Conference on Big Data and Education - ICBDE '18*, Honolulu, USA, 52-56.

Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. *Big Data & Society*, *3*(1), 1-10.

Kumar, S., Tiwari, P., & Zymbler, M. (2019). Internet of Things is a revolutionary approach for future technology enhancement: a review. *Journal of Big Data*, *6*(111), 1-21.

Laney, D. (2001). Deja VVVu: Gartner's Original "Volume-Velocity-Variety" Definition of Big Data. Retrieved 26 November 2020 from https://community.aiim.org/blogs/doug-laney/2012/08/25/deja-vvvu-gartners-original-volume-velocity-variety-definition-of-big-data

Lee, I. (2017). Big data: Dimensions, evolution, impacts, and challenges. *Business Horizons*, *60*(3), 293-303.

Lozada, N., Arias-Pérez, J., & Perdomo-Charry, G. (2019). Big data analytics capability and co-innovation: An empirical study. *Heliyon*, *5*(10), 1-17.

Maheshwari, S., Gautam, P., & Jaggi., C. K. (2021). Role of Big Data Analytics in supply chain management: current trends and future perspectives. *International Journal of Production Research*, *59*(6), 1875-1900.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. (2011). *Big data: The next frontier for innovation, competition, and productivity*. New York: McKinsey Global Institute.

Maroufkhani, P., Wagner, R., Ismail, W. K., Baroto, M. B., & Nourani, M. (2019). Big Data Analytics and Firm Performance: A Systematic Review. *Information*, *10*(7), 1-21.

Mazumdar, S., Seybold, D., Kritikos, K., & Verginadis, Y. (2019). A survey on data storage and placement methodologies for Cloud-Big Data ecosystem. *Journal of Big Data*, *6*(15), 1-37.

Mengist, W., Soromessa, T., & Legese, G. (2020). Method for conducting systematic literature review and meta-analysis for environmental science research. *MethodsX*, *7*, 1-11.

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*, *98*, 261-276.

Mishra, D., Luo, Z., Hazen, B., Hassini, E., & Foropon, C. (2019). Organizational capabilities that enable big data and predictive analytics diffusion and organizational performance: A resource-based perspective. *Management Decision*, *57*(8), 1734-1755.

Mishra, S., & Misra, A. (2017). Structured and Unstructured Big Data Analytics. *Proceedings of the 2017 International Conference on Current Trends in Computer, Electrical, Electronics and Communication (CTCEEC)*, Mysore, India, 740-746.

Müller, O., Fay, M., & vom Brocke, J. (2018). The Effect of Big Data and Analytics on Firm Performance: An Econometric Analysis Considering Industry Characteristics. *Journal of Management Information Systems*, *35*(2), 488-509.

Oláh, J., Krisán, E., Kiss, A., Lakner, Z., & Popp, J. (2020). PRISMA Statement for Reporting Literature Searches in Systematic Reviews of the Bioethanol Sector. *Energies*, *13*, 1-35.

Pati, D., & Lorusso, L. N. (2018). How to Write a Systematic Review of the Literature. *HERD: Health Environments Research & Design Journal*, *11*(1), 15-30.

Persaud, A. (2020). Key competencies for big data analytics professions: a multimethod study. *Information Technology & People*, *34*, 178-203.

Prescott, M. E. (2016). Big Data: Innovation and Competitive Advantage in an Information Media Analytics Company. *Journal of Innovation Management*, *4*(1), 92-113.

Ramadan, M., Shuqqo, H., Qtaishat, L., Asmar, H., & Salah, B. (2020). Sustainable Competitive Advantage Driven by Big Data Analytics and Innovation. *Applied Sciences*, *10*, 1-14.

Ramasamy, A., & Chowdhury, S. (2020). Big Data quality Dimensions: A Systematic Literature Review. *JISTEM – Journal of Information Systems and Technology Management*, *17*, 1-13.

Raut, R. D., Mangla, S. K., Narwane, V. S., Gardas, B., Priyadarshinee, P., & Narkhede, B. E. (2019). Linking big data analytics and operational sustainability practices for sustainable business management. *Journal of Cleaner Production*, *224*, 10-24.

Ren, S., Wamba, S. F., Akter, S., Dubey, R., & Childe, S. J. (2017). Modelling quality dynamics, business value and firm performance in a big data analytics environment. *International Journal of Production Research*, *55*(17), 5011-5026.

Schroeck, M., Shockley, R., Smart, J., Romero-Morales, D., & Tufano, P. (2012). Analytics: The real-world use of big data. Retrieved 24 November 2020 from https://www.informationweek.com/pdf_whitepapers/approved/1372892704_analytics_the_real_world_use_of_big_data.pdf

Sebhatu, S. (2021). Managerial capabilities and firms' sustainable performance: Evidence from Chinese manufacturing small and medium-sized enterprises. *Frontiers in Management and Business*, *2*(1), 74-86.

Shabbir, M. O., & Gardezi, S. B. (2020). Application of big data analytics and organizational performance: the mediating role of knowledge management practices. *Journal of Big Data*, *7*(47), 1-17.

Shan, S., Luo, Y., Zhou, Y., & Wei, Y. (2019). Big data analysis adaptation and enterprises' competitive advantages: the perspective of dynamic capability and resource-based theories. *Technology Analysis & Strategic Management*, *31*(4), 406-420.

Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: a critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ*, *358*.

Silva-Fernández, L., & Carmona, L. (2019). Meta-analysis in the era of big data. *Clinical Rheumatology*, *38*, 2027-2028.

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.

Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, *104*, 333-339.

Vu, H. M. (2020). A Review of Dynamic Capabilities, Innovation Capabilities, Entrepreneurial Capabilities and Their Consequences. *The Journal of Asian Finance, Economics and Business*, *7*(8), 485-494.

Wamba, S. F., & Akter, S. (2019). Understanding supply chain analytics capabilities and agility for data-rich environments. *International Journal of Operations & Production Management*, *39*(6/7/8), 887-912.

Wamba, S. F., Akter, S., Trinchera, L., & De Bourmont, M. (2019). Turning information quality into firm performance in the big data economy. *Management Decision*, *57*(8), 1756-1783.

Wamba, S. F., Gunasekaran, A., Akter, S., Ren, S., Dubey, R., & Childe, S J. (2017). Big data analytics and firm performance: Effects of dynamic capabilities. *Journal of Business Research*, *70*, 356-365.

Wang, W., Zhou, X., Zhang, B., & Mu, J. (2013). Anomaly detection in big data from UWB radars. *Security and Communication Networks*, *8*, 2469-2475.

Yaqoob, I., Hashem, I., Gani, A., Mokhtar, S., Ahmed, E., Anuar, N. B., & Vasilakos, A. V. (2016). Big data: From beginning to future. *International Journal of Information Management*, *36*(6), 1231-1247.

Yu, J. H., & Zhou, Z. M. (2019). Components and Development in Big Data System: A Survey. *Journal of Electronic Science and Technology*, *17*(1), 51-72. .