Pro-Metrics

Big data research in the business, management and accounting field: Revealing the thematic structure based on co-word analysis

**José Luis Ausejo Sánchez[1], Félix Gil Caro Soto[2],
Patricia Elena Ramos La Rosa[2], Damaris Faviola Medina Palma[2],
Gleny Amelia Ching Campos[2], Ana Juliani Rodriguez Cadillo[2]**

[1] Departamento de Administración y Gestión, Universidad Nacional José Faustino Sánchez Carrión, Perú.
   Email: jausejo@unjfsc.edu.pe **ORCID:** https://orcid.org/0000-0003-4674-6150.
   Corresponding author.
[2] Departamento de Administración y Gestión, Universidad Nacional José Faustino Sánchez Carrión, Perú.

ABSTRACT

Objective. This study aims to describe the thematic structure of big data research in business, management, and accounting.

Design/Methodology/Approach. A co-word analysis was performed on 12,345 documents retrieved from Scopus from 2014-2023 in the category "business, management, and accounting". Modularity algorithms were used to identify themes and subthemes, and the clustering of terms was qualitatively analyzed.

Results/Discussion. Five main themes were identified: (1) *Business and economic data analysis,* (2) *big data analytics in marketing, consumer behavior, and employee dynamics,* (3) *Scalable machine learning and big data optimization, smart cities and urban development,* (4) *IoT-Driven innovations in industry 4.0 for optimized supply chain management,* and (5) *Social media and sentiment analysis in modern tourism and hospitality.* The grouping of terms demonstrated the broad application of big data in healthcare, education, tourism, industry, organizational development, finance, social media, marketing, and hospitality.

Conclusions. Big data is a field of application. This is evident in each cluster, where there are sub-themes that are nothing more than applying big data principles in sectors such as manufacturing, tourism services, education, health, and urbanization. Generally, the findings here are similar to other studies that have analyzed broader or more selective literature.

Keywords: big data; business; management; accounting; bibliometrics; co-word analysis.

1. INTRODUCTION

THE TERM "big data" describes a vast set of complexly structured data processed for future decision-making based on hidden patterns (Sagiroglu & Sinanc, 2013). The term originated with the advent of computers and the large amounts of data generated by these devices (Yaqoob *et al.,* 2016). The academic and technological communities are increasingly focusing on big data, given its broad applicability in genetics, neuroscience, economics, and finance (Fan, Han, & Liu, 2014).

The advent of big data has profoundly impacted the management realm, which has led to a revolution in how decisions are made. The vast volume of data available translates into a wealth of knowledge that can be leveraged for informed decision-making (McAfee *et al.,* 2012). As outlined by Davenport and Dyché (2013), using big data can facilitate cost reduction, the reduction of time needed to develop computational tasks, products, and services, and internal business decision support. Big data represents a central tenet of modern science and business (Diebold, 2019). This is why so much attention has been devoted to the bibliometric description of the literature produced about it. Some studies have examined the relationship between big data and other topics. A case in point is the retrospective analysis of the intersection between big data and machine learning. Zhang *et al.,* (2021) conducted a comprehensive analysis of these areas' growth, distribution, and intellectual contributions. The authors revealed big data's leading trends, methodologies, and applicability with machine learning techniques. A similar study analyzing the synergies of big data with machine learning was that of El-Alfy and Mohammed (2020) and Lopez Belmonte (2020), who also emphasized the advances and integration of both areas.

There are numerous examples of bibliometric studies on the application of big data in specific contexts, including the insurance industry (Ellili *et al.,* 2023), health and medicine (Liao *et al.,* 2018; Galetsi & Katsaliaki, 2020), the banking sector (Nobane e. al. (2021), sports (Šuštaršič *et al.,* 2022), the construction industry (Lu & Zhang, 2022), data science (Raban & Gordon, 2020), and finance (Ar-Raisi *et al.,* 2023). Other studies have performed bibliometric analyses from a more generalist perspective, addressing only the literature in databases such as Scopus and Web of Science (Kalantari *et al.,* 2017; Liu, 2020; Inamdar *et al.,* 2021). As previously stated, big data significantly impacts the management sector. Studies focused on this and related topics have also been developed. Ardito *et al.,* (2019) conducted a bibliometric analysis of research on big data analytics within the context of business and management. They aimed to examine the field's evolution, key research areas, and influential publications. The study discussed the impact of big data analytics on decision-making processes, business strategies, and organizational performance. Similarly, Khanra *et al.,* (2020) provided a bibliometric synthesis of the literature on big data analytics in enterprises. The researchers identified significant trends, methodologies, and applications. Their study emphasized the role of big data analytics in enhancing business operations, customer insights, and competitive advantage. Zhang *et al.,* (2021) examined a decade of research on big data in business. The authors analyzed the most influential studies, key research areas, and emerging trends. In a related vein, Batistič & van der Laken (2019) conducted a bibliometric analysis of the history, evolution, and future of big data and analytics, focusing on its relationship with organizational performance.

Based on the above research antecedents, this study aims to describe the thematic structure of big data research produced in business, management, and accounting. Following some bibliometric principles, we will employ a co-word analysis of the literature produced in the last ten years (2014-2023).

2. METHODOLOGY

The Scopus database was used to retrieve literature on the topic under study. The term used in the search equation was "big data" in the title, abstract, and keywords. The temporal coverage was ten years (2014-2023). The results were refined for the thematic category "Business, Management & Accounting" and the documentary typology: article, conference paper, chapter, review, and conference review. The equation is shown below:

TITLE-ABS-KEY (big AND data) AND PUBYEAR > 2013 AND PUBYEAR < 2024 AND (LIMIT-TO (SUBJAREA, "BUSI")) AND (LIMIT-TO (DOCTYPE, "ar") OR LIMIT-TO (DOCTYPE, "cp") OR LIMIT-TO (DOCTYPE, "ch") OR LIMIT-TO (DOCTYPE, "re") OR LIMIT-TO (DOCTYPE, "cr"))

By the above query, 14,345 documents were retrieved, the distribution of which is shown in Figure 1. Co-word analysis was performed on words in the title and abstract. The words with a frequency of ten or more (4942) were selected.

Based on the VOS method (Van Eck & Waltman, 2010), 60% of the terms were selected based on relevance. The terms were then normalized to remove possible irrelevant words or word families. The co-word network consisted of 2966 words and 360105 edges. Vincent *et al.*, (2008) defined the modularity algorithm to detect thematic structure. Centrality measures were then calculated to identify the importance of the nodes in the map (words) and the relationships between them (links). Finally, qualitative interpretations of the grouping of terms in their clusters were performed to identify topics and subtopics.
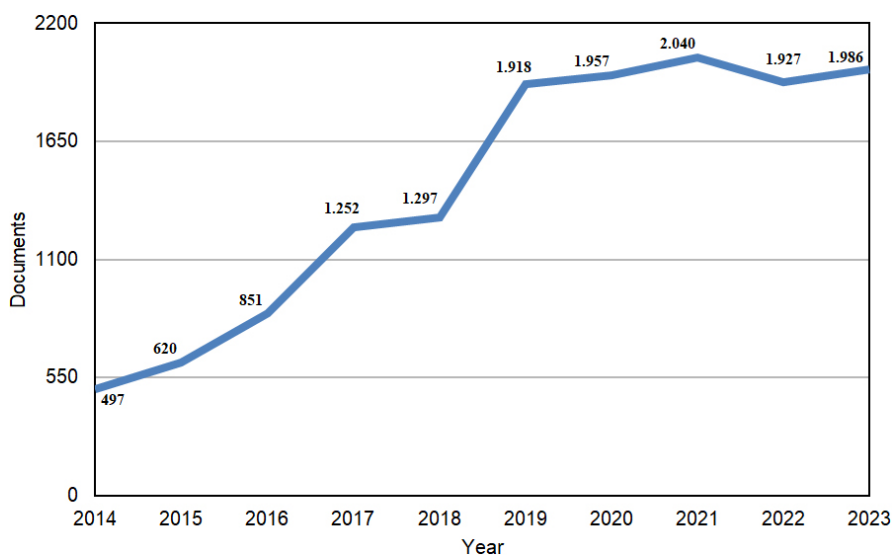

Figure 1. Documents per year.

4. RESULTS

Following the application of the methodological procedures, five thematic clusters were identified (See Figure 2 and Table 1). Five thematic subcategories are evident in the first cluster (551 words, orange nodes). The first one is related to accounting and auditing. Highly centralized terms stand out, such as accountant (437), audit (499), auditor (715), and audit quality (427). The high centrality of the terms 'audit' and 'auditor' highlights their pivotal role in financial oversight. The second subcategory refers to finance and banking, highlighting bank (1017), financial market (215), credit risk (162), and financial institution (344). Terms associated with banking and finance also show high centrality, particularly 'bank,' which is highly interconnected with other financial terms,

indicating its significant influence. The third subcategory refers to corporate governance. Terms like 'board' and 'corporate governance' have substantial centrality, underscoring the importance of governance structures in corporate management. The fourth subcategory refers to the economic indicators such as 'economic growth' and 'GDP,' which are highly central, reflecting their critical role in financial analysis and policy-making. Lastly, the fifth subcategory within this cluster involves data analysis and statistics, highlighting their importance in conducting rigorous data-driven research.

The second cluster is made up of 755 words (green nodes). This cluster contains several sub-themes. The first one focuses on the adoption, capabilities, and utilization of big data and analytics in organizations. It encompasses
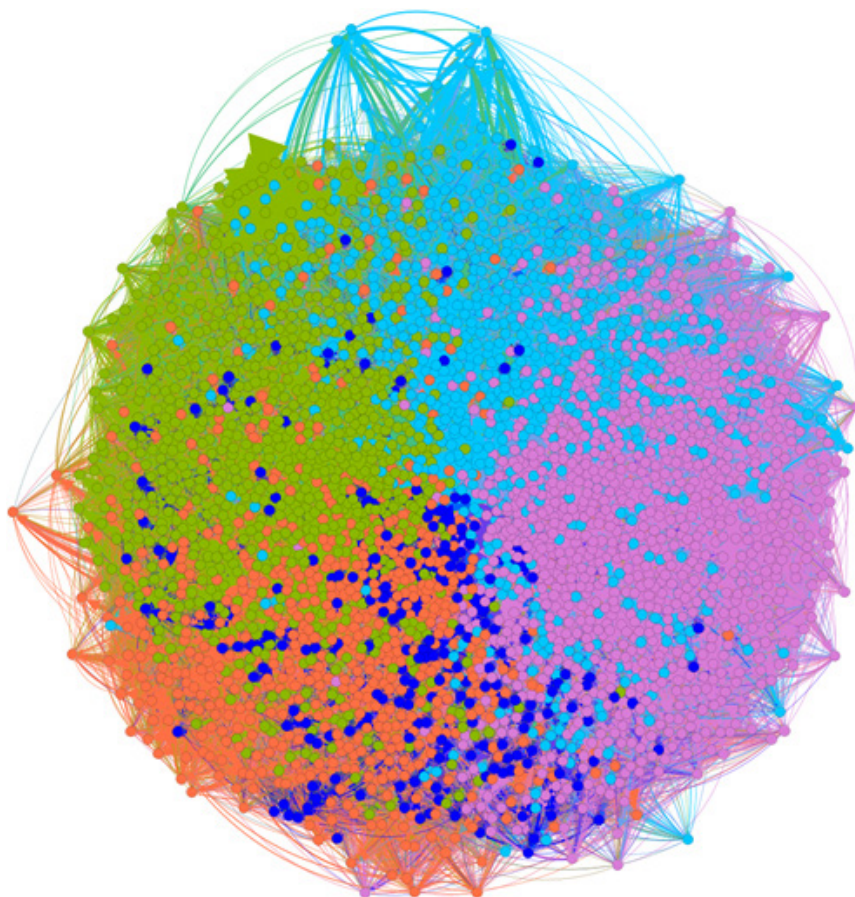
Figure 2. Map derived from the co-word analysis on big data research.

the skills and technologies required to analyze large datasets and the strategic advantages these capabilities provide. Among the most highlighted terms, we find big data adoption (1105), big data analytics capability (431), big data capability (211), big data analytics adoption (134), and others. The second subtopic relates to various marketing aspects, including advertising, brand management, and consumer behavior. It explores how brands build awareness, loyalty, and equity and how consumers engage with and perceive brands and marketing efforts. Among the most important terms, in terms of centrality, we visualize brand (807), advertisement (345), marketing strategy (327), consumer behavior (227), and brand image (207). The third subtopic focuses on the strategic aspects of organizational performance, including how companies achieve and sustain competitive advantage, innovate their business models, and manage dynamic capabilities to adapt to changing environments. The key

terms are competitive advantage (1058), business performance (452), dynamic capability (275), business model innovation (210), organizational performance (170), and competitive strategy (132). The fourth subtopic encompasses research related to academic institutions, educational performance, and curriculum development. It includes studying higher education systems, educators' roles, and students' experiences in academic environments. The most highlighted terms are higher education (284), educator (282), college (264), curriculum (256), and educational institution (248). The last subtopic in this cluster covers various aspects of employee behavior and organizational dynamics, including employee engagement and performance, job satisfaction, organizational culture, and leadership styles. It focuses on how organizations manage and motivate their workforce. The most important terms here in terms of centrality are employee (1023), leadership (376), empowerment (260), organizational

culture (134), employee engagement (117), and job satisfaction.

The third cluster is composed of 906 terms (pink nodes). The first subtopic revolves around machine learning techniques and prediction models. It includes various algorithms, such as neural networks, support vector machines, random forests, and deep learning methods for prediction, classification, and optimization. Among the key terms, we find task (5549), prediction (4628), machine (4302), neural network (1213), support vector machine (1162), random forest (921), and machine learning algorithm. The second subtopic concerns the various aspects of data processing and storage. It includes handling large datasets, unstructured data, and the infrastructure required for efficient data storage and processing, such as HDFS and data warehouses. The third subtopic encompasses data and analytics in healthcare and medicine. It involves patient data, disease management, medical records, and the application of data processing techniques to improve healthcare outcomes and services. The most relevant terms related to this subtopic are patient (1949), healthcare (1360), hospital (705), disease (702), and medicine (596). The fourth subtopic deals with the challenges and techniques of big data processing. It includes frameworks like Hadoop and Spark and methods for ensuring scalability and performance in big data environments. The fifth subtopic focuses on the development and application of optimization algorithms. It includes techniques such as genetic algorithms, particle swarm optimization, and gradient-based methods to solve complex optimization problems.

Cluster 4 has 565 terms (light blue nodes). A first sub-theme, encompassing the broad and rapidly expanding Internet of Things (IoT) field, is evident. It includes developing, deploying, and managing connected devices and sensors, the systems and protocols that support them, and their various applications in smart cities, healthcare, manufacturing, and other industries. The second subtopic focuses on integrating smart technologies in urban environments to create efficient, sustainable, and livable cities. It includes smart grids, urban planning, sustainable urban development, and smart transportation systems aimed at improving the quality of life in cities. Key terms are smart

city (1469), urbanization (502), urban planning (447), smart grid (324), and traffic management (247). The third subtopic addresses the advancements in manufacturing technologies, often called Industry 4.0. It includes smart manufacturing processes, cyber-physical systems, the industrial Internet, and the integration of advanced technologies like IoT, robotics, and automation in manufacturing. The fourth subtopic covers the optimization and management of supply chains and logistics. It includes strategies for sustainable supply chain management, logistics operations, inventory management, and implementing advanced technologies to enhance efficiency and traceability in supply chains. Lastly, the fifth subtopic reflects the major themes within the dataset, highlighting the interplay between IoT, smart cities, manufacturing, supply chain management, and data analytics. Each topic represents a significant area of technological advancement and research, showcasing the integration of modern technologies in various domains.

The fifth cluster is the smallest, with 189 words (dark blue nodes). The first subtopic covers various aspects of tourism, including activities, destinations, industry practices, and research. It explores tourist behaviors, travel patterns, popular destinations, and the overall impact of tourism on different regions. The second subtopic focuses on sentiment analysis and its application on social media platforms. It includes analyzing sentiments from tweets, reviews, and other social media posts to understand public opinion, trends, and user engagement. The third subtopic covers the hospitality and hotel industry, focusing on various aspects such as hotel management, performance, guest satisfaction, and the overall role of hospitality in the tourism sector. The fourth subtopic explores the dynamics of urban and rural tourism, including the distribution and impact of tourism in different geographical areas. It covers urban planning, development, and the interaction between tourists and residents. Finally, the fifth subtopic addresses the significance of online reviews and digital presence for businesses, particularly tourism and hospitality. It includes analyzing customer reviews, review helpfulness, and the impact of platforms like TripAdvisor and Google on consumer decisions.

Topical cluster	Topic	Subtopics	Terms (Degree)
Cluster 1	Business and Economic data Analysis	Accounting & auditing, finance & banking, corporative governance, economic indicators, and data analysis & statistics	Variable (6922.0), firm (5192.0), evidence (3287.0), price (2464.0), return (2162.0), regression (2016.0), policymaker (1825.0), ratio (1792.0), regulator (1721.0), investor (1496.0).
Cluster 2	Big Data Analytics in Marketing, Consumer Behavior, and Employee Dynamics	Big Data Analytics & Capabilities, Marketing & Consumer Behavior, Organizational Performance & Strategy, Academic & Educational Research, Employee & Organizational Behavior	Methodology approach (9912.0), sample (7522.0), research limitations implication (6441.0), questionnaire (4081.0), respondent (3093.0), student (3005.0), perception (2934.0), SMEs (2158.0), personality trait (2139.0), validity (2130.0).
Cluster 3	Scalable Machine Learning and Big Data Optimization, Smart Cities and Urban Development	Machine Learning and Prediction Models, Data Processing and Storage, Healthcare and Medicine, Big Data and Scalability, Optimization and Algorithm Development	Task (5549.0), prediction (4628.0), machine (4302.0), processing (3613.0), feature (3083.0), storage (2814.0), dataset (2553.0), parameter (2357.0), optimization (2092.0), treatment (1998.0).
Cluster 4	IoT-Driven Innovations in Industry 4.0 for Optimized Supply Chain Management	Internet of Things (IoT) and Related Technologies, Manufacturing and Industry 4.0, Supply Chain Management and Logistics, Data Analytics and Research	Thing (10776.0), internet (6440.0), IoT (3594.0), sensor (2931.0), supply chain management (2139.0), systematic literature review (1981.0), transportation (1914.0), manufacturing (1585.0), smart city (1469.0), publication (1460.0).
Cluster 5	Social Media and Sentiment Analysis in Modern Tourism and Hospitality	Tourism and Tourist Activities, Sentiment Analysis and Social Media, Hospitality and Hotel Industry, Urban and Rural Tourism, Online Reviews and Digital Presence	Tourism (2612.0), tourist (1758.0), sentiment analysis (1566.0), sentiment (1542.0), twitter (1480.0), tweet (1225.0), visitor (1216.0), text mining (1107.0), rating (1059.0), tourism industry (950.0).

Table 1. Summary of topics and subtopics around the bid data research.

5. FINAL CONSIDERATIONS

This study identified the main lines of research on big data research in the areas of business, management, and accounting. Five main lines of research were identified: (1) Business and economic data analysis, (2) big data analytics in marketing, consumer behavior, and employee dynamics, (3)scalable machine learning and big data optimization, smart cities and urban development, (4) IoT-Driven innovations in industry 4.0 for optimized supply chain management, and (5) Social media and sentiment analysis in modern tourism and hospitality.

Big data is a field of application. This is evident in each cluster, where there are sub-themes that are nothing more than applying big data principles in sectors such as manufacturing, tourism services, education, health, and urbanization. These same patterns were identified by Inamdar *et al.,* (2021). On the other hand, one of the most representative clusters in terms of its nodes and the relationships between them was

the one referred to big data and machine learning. This topic reflects the need to solve general problems due to the large volume of data, mainly in organizations (Zhang *et al.,* 2021). Since the study developed by López Belmonte and colleagues in 2020, machine learning has been a consolidated topic within big data research (López Belmonte *et al.,* 2020). For example, cluster 5 has been of great relevance within the results. Zhang *et al.,* (2021) have already evidenced through term maps that sentiment analysis is an emerging area within big data.

The thematic structure identified here is also close to other findings, such as, for example, that of Mishra *et al.,* (2018). These authors also identified themes such as big data analytics (our cluster 1), big data and supply chain management (our cluster 4), and big data applications in specific sectors such as health and education. Generally, the findings here are similar to other studies that have analyzed broader or more selective literature. There are many benefits that the research and application of big data

can offer, and, as shown in Table 1, it is a growing area. At least, this is evidenced by scientific productivity. It would be good to apply other bibliometric techniques in the future to identify other research patterns around the topic addressed here. ●

REFERENCES

Agustí, M. A., & Orta-Pérez, M. (2023). Big data and artificial intelligence in the fields of accounting and auditing: a bibliometric analysis. *Spanish Journal of Finance and Accounting/Revista Española de Financiación y Contabilidad, 52*(3), 412-438. DOI: 10.1080/02102412.2023.2188828

Ar-Raisi, F. A., Sakti, E., Anggono, A., & Tarjo, T. (2023). Bibliometric Analysis of Big Data Research in Finance. *Journal Magister Akuntansi Trisakti, 10*(1), 1-18. DOI: 10.25105/jmat.v10i1.12607.

Ardito, L., Scuotto, V., Del Giudice, M., & Petruzzelli, A. M. (2019). A bibliometric analysis of research on Big Data analytics for business and management. *Management Decision, 57*(8), 1993-2009. DOI: 10.1108/MD-07-2018-0754

Batistič, S., & van der Laken, P. (2019). History, evolution and future of big data and analytics: a bibliometric analysis of its relationship to performance in organizations. *British Journal of Management, 30*(2), 229-251. DOI: 10.1111/1467-8551.12340.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment, 2008*(10), P10008. DOI 10.1088/1742-5468/2008/10/P10008.

Davenport, T. H., & Dyché, J. (2013). Big data in big companies. *International Institute for Analytics, 3*(1-31).

Fan, J., Han, F., & Liu, H. (2014). Challenges of big data analysis. *National Science Review, 1*(2), 293-314.

El-Alfy, E. S. M., & Mohammed, S. A. (2020). A review of machine learning for big data analytics: bibliometric approach. *Technology Analysis & Strategic Management, 32*(8), 984-1005. DOI: 10.1080/09537325.2020.1749991.

Diebold, F. X. (2019). *On the origin (s) and development of "big data": The phenomenon, the term, and the discipline.* Available at https://www.sas.upenn.edu/~fdiebold/papers/paper112/Diebold_Big_Data.pdf

Ellili, N., Nobanee, H., Alsaiari, L., Shanti, H., Hillebrand, B., Hassanain, N., & Elfout, L. (2023). The applications of big data in the insurance industry: A bibliometric and systematic review of relevant literature. *The Journal of Finance and Data Science,* 100102. DOI: 10.1016/j.jfds.2022.100102.

Galetsi, P., & Katsaliaki, K. (2020). Big data analytics in health: An overview and bibliometric study of research activity. *Health Information & Libraries Journal, 37*(1), 5-25. DOI: 10.1111/hir.12295.

Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2021). A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. *Journal of Enterprise Information Management, 34*(1), 101-139. DOI: 10.1108/JEIM-07-2019-0202.

Kalantari, A., Kamsin, A., Kamaruddin, H. S., Ale Ebrahim, N., Gani, A., Ebrahimi, A., & Shamshirband, S. (2017). A bibliometric approach to tracking big data research trends. *Journal of big data, 4,* 1-18. DOI: 10.1186/s40537-017-0077-4.

Khanra, S., Dhir, A., & Mäntymäki, M. (2020). Big data analytics and enterprises: a bibliometric synthesis of the literature. *Enterprise Information Systems, 14*(6), 737-768. DOI: 10.1080/17517575.2020.1734241

Liu, X., Sun, R., Wang, S., & Wu, Y. J. (2020). The research landscape of big data: a bibliometric analysis. *Library Hi Tech, 38*(2), 367-384. DOI: 10.1108/LHT-01-2020-0025.

López Belmonte, J., Segura-Robles, A., Moreno-Guerrero, A. J., & Parra-González, M. E. (2020). Machine learning and big data in the impact literature. A bibliometric review with scientific mapping in Web of Science. *Symmetry, 12*(4), 495. DOI: 10.3390/sym12040495

Lu, Y., & Zhang, J. (2022). Bibliometric analysis and critical review of the research on big data in the construction industry. Engineering, *Construction and Architectural Management, 29*(9), 3574-3592. DOI: 10.1108/ECAM-07-2021-0652.

McAfee, A., Brynjolfsson, E., Davenport, T. H., Patil, D. J., & Barton, D. (2012). Big data: the management revolution. *Harvard Business Review, 90*(10), 60-68.

Mishra, D., Gunasekaran, A., Papadopoulos, T., & Childe, Liao, H., Tang, M., Luo, L., Li, C., Chiclana, F., & Zeng, X. J. (2018). A bibliometric analysis and visualization of medical big data research. *Sustainability, 10*(1), 166. DOI: 10.3390/su10010166.

Nobanee, H., Dilshad, M. N., Al Dhanhani, M., Al Neyadi, M., Al Qubaisi, S., & Al Shamsi, S. (2021). Big Data applications in the banking sector: A bibliometric analysis approach. *Sage Open, 11*(4), 21582440211067234. DOI: 10.1177/21582440211067234.

Raban, D. R., & Gordon, A. (2020). The evolution of data science and big data research: A bibliometric analysis. *Scientometrics, 122*(3), 1563-1581. DOI: 10.1007/s11192-020-03376-1.

Sagiroglu, S., & Sinanc, D. (2013, May). Big data: A review. In *2013 International Conference on Collaboration Ttechnologies and systems (CTS)* (pp. 42-47). IEEE.

S. J. (2018). Big Data and supply chain management: a review and bibliometric analysis. *Annals of Operations Research, 270,* 313-336.

Inamdar, Z., Raut, R., Narwane, V. S., Gardas, B., Narkhede, B., & Sagnak, M. (2021). A systematic literature review with bibliometric analysis of big data analytics adoption from period 2014 to 2018. *Journal of Enterprise Information Management, 34*(1), 101-139.

Šuštaršič, A., Videmšek, M., Karpljuk, D., Miloloža, I., & Meško, M. (2022). Big data in sports: A bibliometric and topic study. Business Systems Research: International *Journal of the Society for Advancing Innovation and Research in Economy, 13*(1), 19-34. DOI: 10.2478/bsrj-2022-0002.

Van Eck, N., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.

Yaqoob, I., Hashem, I. A. T., 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.

Zhang, J. Z., Srivastava, P. R., Sharma, D., & Eachempati, P. (2021). Big data analytics and machine learning: A retrospective overview and bibliometric analysis. *Expert Systems with Applications, 184,* 115561. DOI: 10.1016/j.eswa.2021.11556.

Zhang, Y., Zhang, M., Li, J., Liu, G., Yang, M. M., & Liu, S. (2021). A bibliometric review of a decade of research: Big data in business research-Setting a research agenda. *Journal of Business Research, 131,* 374-390. DOI: 10.1016/j.jbusres.2021.03.061.