West Science Journal Economic and Entrepreneurship
Vol. 2, No. 03, August 2024, pp. 372~xx

Bibliometric Analysis to Understand Research Trends in Data-Driven Economics

Loso Judijanto[1], Komang Shanty Muni[2], Firlie Lanovia Amir[3]
[1]IPOSS Jakarta
[2,3]Institut Pariwisata dan Bisnis Internasional

Article Info

Article history:

Received August, 2024
Revised August, 2024
Accepted August, 2024

Keywords:

Data-Driven Economics
Bibliometric Analysis
VOSviewer

ABSTRACT

The bibliometric analysis conducted in this study offers a nuanced understanding of the evolution and current state of research within data-driven domains, highlighting "data" as a central theme interconnected with diverse fields like business, smart cities, and machine learning. This research provides valuable insights into the dominant trends, collaboration patterns, and thematic priorities that shape the data-driven research landscape, underscoring the critical role of data in advancing both theoretical knowledge and practical applications across various sectors. The findings suggest significant opportunities for enhancing data analytics capabilities, developing targeted educational programs, and fostering interdisciplinary collaborations that can bridge existing gaps in the literature. However, the study also notes limitations inherent in bibliometric analyses, including potential biases toward more frequently cited or recent publications and the exclusion of works outside selected databases. Overall, this analysis not only reflects the dynamic and impactful nature of data-driven research but also guides future academic and practical endeavors in the field.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Name: Loso Judijanto
Institution: IPOSS Jakarta
e-mail: losojudijantobumn@gmail.com

1. INTRODUCTION

The age of data-driven decision-making has ushered in significant transformations across various sectors, with economics being notably impacted [1]. In recent years, the proliferation of data sources and the advancement of analytical tools have shifted traditional economic analysis toward more dynamic, real-time methodologies that leverage big data [1], [2]. This paradigm shift is not merely a technical evolution but represents a broader change in how economic insights are derived and implemented [3]. As economies worldwide become increasingly interconnected and digitalized, the demand for data-driven economic research has surged, reflecting its critical role in shaping policy, business strategy, and financial models [4].

Furthermore, the integration of big data analytics into economics has allowed for a more nuanced understanding of complex phenomena that traditional models could not fully capture [5]. These phenomena include consumer behavior, market trends, and financial risks, all of which exhibit patterns that only extensive datasets can reveal [6]. Consequently, this has led to an enhanced

capability to forecast economic outcomes, tailor policy interventions, and optimize business operations, making the field not only more efficient but also more predictive and responsive to changes [7].

However, despite its growing importance, the field of data-driven economics is continually evolving, characterized by rapid developments in methodologies, technologies, and areas of application. This continuous evolution poses challenges for scholars and practitioners in keeping abreast of the most relevant research trends and technological advancements. As such, a comprehensive overview of the existing literature and a clear understanding of the trajectory of research in this field are essential for those engaged in economic analysis and policy-making.

This situation underscores the utility of bibliometric analysis, which offers a systematic method for mapping the scientific landscape of a research area. By analyzing publication patterns, citation structures, and co-authorship networks, bibliometric studies can identify the most influential works, prolific authors, and core topics, providing a structured comprehension of the field's development. Such analyses are crucial not only for recognizing the state of the art but also for uncovering emerging trends and gaps in the literature that might be pivotal for future investigations [8].

Despite the apparent advantages and the expanding scope of data-driven economics, there is a noticeable lack of comprehensive studies that consolidate the various strands of research within this domain. Existing literature reviews often focus on specific aspects like econometrics techniques, big data applications in finance, or economic forecasting, without integrating these components into a broader economic context. This fragmented approach may hinder the development of a cohesive understanding necessary for advancing the discipline. Therefore, there is a pressing need to conduct a bibliometric analysis to provide a holistic view of the research trends, key themes, and methodological approaches within the realm of data-driven economics. Such an analysis would not only highlight how the field has evolved but also pinpoint the areas that warrant further exploration or are on the cusp of breakthroughs.

The objective of this research is to conduct a bibliometric analysis on data-driven economics to elucidate the development patterns, key themes, and future directions of the field. Specifically, the study aims to identify the most impactful authors, institutions, and publications that have shaped the landscape of data-driven economic research. Additionally, this study seeks to uncover the thematic structures and collaborative networks within the literature, thereby providing insights into the intellectual and social structures that underpin the field. This comprehensive mapping will serve as a valuable resource for researchers, policymakers, and practitioners, guiding future research initiatives and informing economic strategies and decisions.

2. LITERATURE REVIEW

2.1 Evolution of Data-Driven Economics

The concept of data-driven economics is rooted in the broader evolution of econometrics and quantitative economics, which have historically utilized statistical methods to interpret economic realities. With the digital revolution, the scope of data sources expanded exponentially, leading to what many now refer to as the era of big data. [9] discuss the transformation brought about by big data in economic research, emphasizing its role in enabling more precise and scalable analyses. They highlight the shift from model-driven to data-driven approaches, where the availability of large datasets can reduce the reliance on strict assumptions that traditional econometric models often require.

[10] explore the implications of machine learning for economics, suggesting that these methods can uncover complex patterns in data that are not easily accessible through traditional statistical techniques. Their work is instrumental in bridging the gap between econometric methodologies and

machine learning, showing how predictive modeling can enhance economic forecasting and decision-making processes. This integration of machine learning into economics has not only expanded the analytical tools available to economists but has also prompted a reevaluation of economic theories and models in light of new data insights.

2.2 Key Areas of Research in Data-Driven Economics

Research in data-driven economics spans several key areas, including consumer behavior, financial markets, and macroeconomic forecasting. [11] provides an in-depth look at how big data is used to better understand consumer preferences and behavior, illustrating this through case studies in online commerce and advertising. The ability to analyze consumer data in real time allows for more responsive and personalized economic strategies, enhancing consumer satisfaction and business performance.

In the realm of financial economics, [12] examine the role of big data in improving financial monitoring and risk management. They discuss the potential of big data to enhance the accuracy of risk models and to provide financial institutions with the tools needed to manage complex portfolios in volatile markets. Similarly, macroeconomic forecasting has seen significant advancements due to the incorporation of big data. Their research demonstrates how data-driven models can improve the prediction of economic indicators, aiding policymakers in crafting more effective economic policies.

2.3 Bibliometric Studies on Data-Driven Economics

Bibliometric analyses have been applied to various fields to map research trends and patterns, and data-driven economics is no exception. For instance, [13] conduct a bibliometric study to analyze the publication patterns and citation networks within the field. Their findings reveal a rapid increase in the number of publications related to data-driven methods in economics, indicating a growing interest and investment in this research area. More recent studies, such as those by [14], focus on collaboration networks and thematic clusters in data-driven economics research. They use co-citation and content analysis to identify the most influential studies and authors, as well as the primary themes that shape the field. These themes often revolve around methodological innovations, applications of data-driven techniques in economic forecasting, and the ethical considerations of using large-scale data.

3. METHODS

This study employs a bibliometric analysis to explore the landscape of research within data-driven economics. We systematically collected peer-reviewed articles, conference papers, and reviews from Google Scholar, using keywords such as "data-driven economics," "big data in economics," and "economic data analysis." The time frame for the literature search was set from 1991-2024 to capture the most relevant and recent developments in the field. Utilizing VOSviewer software, we conducted co-citation, co-authorship, and keyword co-occurrence analyses to identify the most influential authors, seminal works, and prevailing themes within the dataset. This comprehensive approach not only highlights the evolution and intellectual structure of data-driven economics but also allows for the mapping of collaboration networks and thematic trends across the global research community. Data were analyzed quantitatively to ensure a rigorous and objective depiction of the field's trajectory and current focus areas.

4. RESULTS AND DISCUSSION

4.1 Metrics Data of Literature

Table 1. Citation Metrics

Publication years:	1991-2024

Citation years:	33 (1991-2024)
Papers:	690
Citations:	97620
Cites/year:	2958.48
Cites/paper:	141.48
Cites/author:	40193.50
Papers/author:	313.30
Author/paper:	3.09
h-index:	175
g-index:	299
hI,norm:	101
hI,annual	3.06
hA-index	66
Papers with ACC >= 1,2,5,10,20:	
584,530,435,383,274	

Source: Publish or Perish, 2024

Table 1 presents a comprehensive bibliometric analysis of publications in a specified field spanning from 1991 to 2024, with a total of 690 papers analyzed over 33 years. This body of work has accumulated a substantial 97,620 citations, averaging 2,958.48 citations per year, which indicates a high impact and relevance in the scholarly community. Each paper, on average, received 141.48 citations, reflecting significant individual paper influence. The analysis also shows a high level of collaboration, as evidenced by an average of 3.09 authors per paper, with each author contributing to approximately 313.30 papers, resulting in about 40,193.50 citations per author. The h-index of 175 signifies those 175 papers have been cited at least 175 times, underscoring the depth of influential research within this collection. Similarly, the g-index at 299 suggests that the top 299 papers have collectively received at least 89,401 citations. The normalized individual h-index (hI,norm) is 101, and the annualized individual h-index (hI,annual) is 3.06, both metrics further confirming the strong citation performance and enduring relevance of the research over time. The hA-index at 66 points to the high citation counts of papers authored by a single author. The breakdown of papers with at least 1, 2, 5, 10, and 20 citations also demonstrates the broad dissemination and acknowledgment of this body of work across the academic community.

4.2 Citation Analysis

Table 2. Top Cited Literature

Citation	Author and Year	Title	Findings
2518	[15]	Data-driven science and engineering: Machine learning, dynamical systems, and control	This paper demonstrates how data-driven techniques can be integrated into the fields of machine learning, dynamical systems, and control engineering, providing robust tools for predictive analysis and system behavior understanding.
2279	[16]	Data science and its relationship to big data and data-driven decision making	The authors explore the intersection of data science with big data and its implications for data-driven decision-making, emphasizing the
1904	[17]	Data-driven intelligent transportation systems: A survey	This comprehensive survey reviews the advancements in data-driven intelligent transportation systems, highlighting the

			significant improvements in efficiency, safety, and traffic management they facilitate.
1793	[18]	Should you be persuaded: Two samples of data-driven learning materials	This study evaluates the effectiveness of data-driven learning materials in influencing student engagement and learning outcomes, suggesting that such materials can significantly enhance educational practices.
1725	[19]	Dynamic mode decomposition: data-driven modeling of complex systems	The paper focuses on the application of dynamic mode decomposition in the modeling of complex systems, presenting it as a powerful method for extracting dynamic features from large sets of data.
1692	[20]	Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations	his research presents a novel approach to distributionally robust optimization using the Wasserstein metric, offering new insights into performance guarantees and practical reformulations that enhance computational tractability.
1662	[21]	A review on basic data-driven approaches for industrial process monitoring	The paper reviews various data-driven approaches for monitoring industrial processes, discussing their potential to improve process reliability and efficiency through enhanced real-time monitoring and fault detection.
1648	[22]	A review of data-driven building energy consumption prediction studies	This review analyzes studies on data-driven predictions of building energy consumption, emphasizing the importance of these approaches in enhancing energy efficiency and reducing operational costs.
1488	[23]	Survey on data-driven industrial process monitoring and diagnosis	This survey covers the development and application of data-driven methods in the monitoring and diagnosis of industrial processes, highlighting their effectiveness in identifying and addressing inefficiencies and anomalies.
1320	[24]	Strength in numbers: How does data-driven decisionmaking affect firm performance?	The authors assess the impact of data-driven decision-making on firm performance, finding that firms that leverage data-driven strategies typically see significant improvements in performance metrics across various dimensions.

4.3 Co-Word Network Visualization Analysis

1. Network Visualization

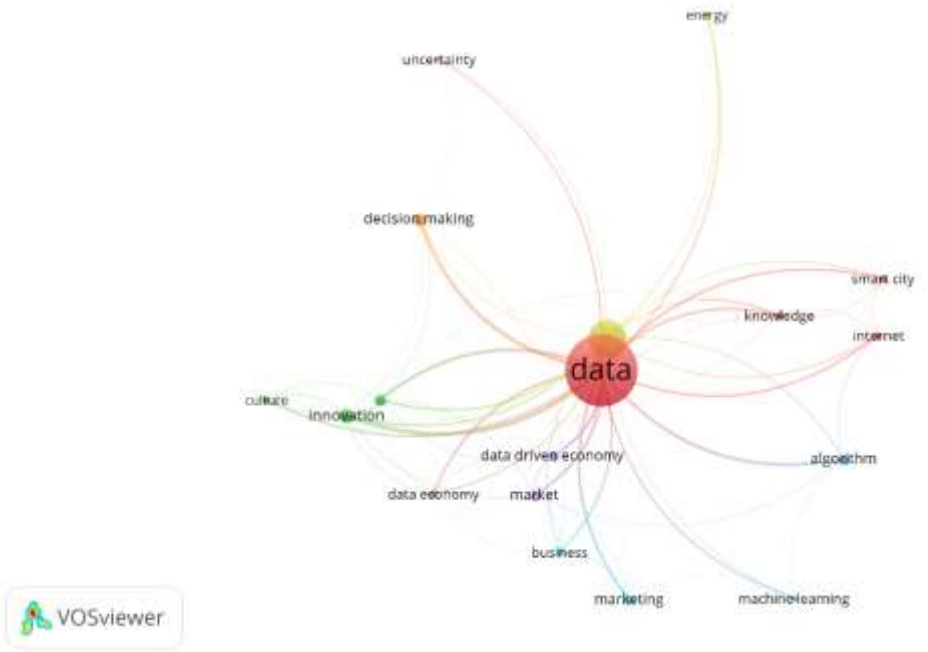

Figure 1. Network Visualization
Source: Data Analysis, 2024

The visual above illustrates the interconnected themes within the realm of data-centric research, focusing on how various concepts are interlinked through their relationships with "data." At the center, the term "data" acts as the core node, signifying its pivotal role in various domains and its central importance in the discourse analyzed. The size and color intensity of the "data" node emphasize the frequency and centrality of discussions around data in academic and practical contexts, highlighting its significance across multiple fields.

Surrounding the central node are clusters of related terms, each represented by different colors, suggesting thematic groupings or specific contexts in which data is heavily utilized or discussed. For instance, terms like "decision-making," "culture," "innovation," and "data economy" form one cluster, suggesting a focus on how data influences economic decisions, fosters innovation, and integrates into cultural contexts. This grouping may explore the impact of data on shaping business strategies, policy-making, and cultural shifts in the digital age. Another notable cluster includes terms like "machine learning," "algorithm," "internet," and "smart city." This cluster points to the technological and infrastructural aspects of data application, particularly in how data-driven technologies such as machine learning and algorithms are applied to urban development and internet technologies. The connection to "smart city" implies a focus on how data analytics and connected technologies enhance urban management and services, contributing to the development of more efficient, sustainable, and "smart" urban environments. Finally, the presence of terms like "energy," "market," and "business" in proximity to "data" reflects the cross-sectoral impact of data. These links suggest investigations into how data is leveraged in energy management, market analysis, and business operations, underlining the pervasive influence of data across various industries. The visual mapping provided by VOSviewer serves as a tool for identifying how scholarly discourse and research agendas are being shaped by data-driven innovations across different sectors, offering insights into the thematic priorities and collaborative networks within the research community.

2. Overlay Visualization

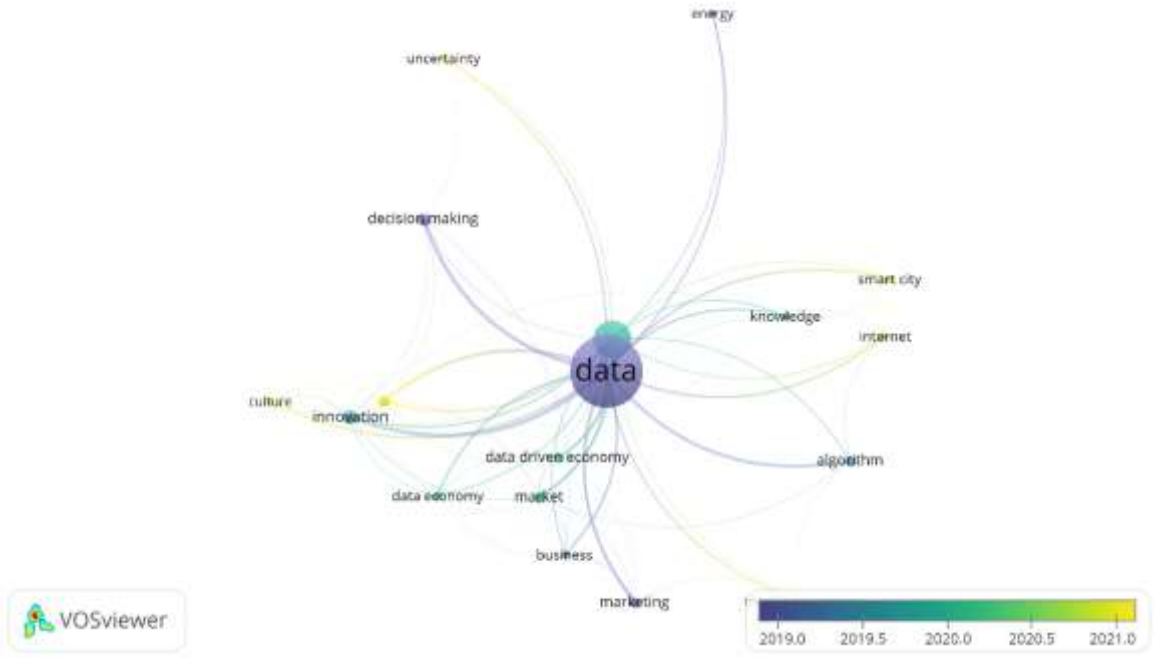

Figure 2. Overlay Visualization
Source: Data Analysis Result, 2024

The second VOSviewer visualization incorporates a temporal dimension, evident from the color gradient ranging from blue to yellow, representing the progression from 2019 to 2021. The central node, "data," remains pivotal, now enriched with a time-sensitive overlay that shows the evolution of associated themes over recent years. This temporal aspect allows us to trace the development and shifting focus within the research on data-driven fields, highlighting how certain topics have gained or waned in emphasis over time.

The color gradient indicates that newer discussions, depicted in warmer tones (yellow), have increasingly centered around "smart city," "internet," and "algorithm." This suggests a more recent focus on the application of data-driven technologies in urban development and online environments, as well as the advancement of algorithms in managing and analyzing vast data sets. Conversely, themes like "decision-making" and "culture," shown in cooler tones (blue), may represent areas of focus that were more prominent at the beginning of the depicted time frame, indicating foundational research areas that set the stage for later developments.

Moreover, the interconnections among the nodes suggest a complex network of interdisciplinary research that has dynamically evolved. The visualization provides a macroscopic overview of how data-related research interacts with various sectors such as energy, business, and marketing, and how these interactions have shifted over time. It underscores the broadening scope of data applications and reflects a research landscape that is increasingly collaborative and integrated across domains, adapting to new technologies and societal needs as they emerge.

3. Density Visualization

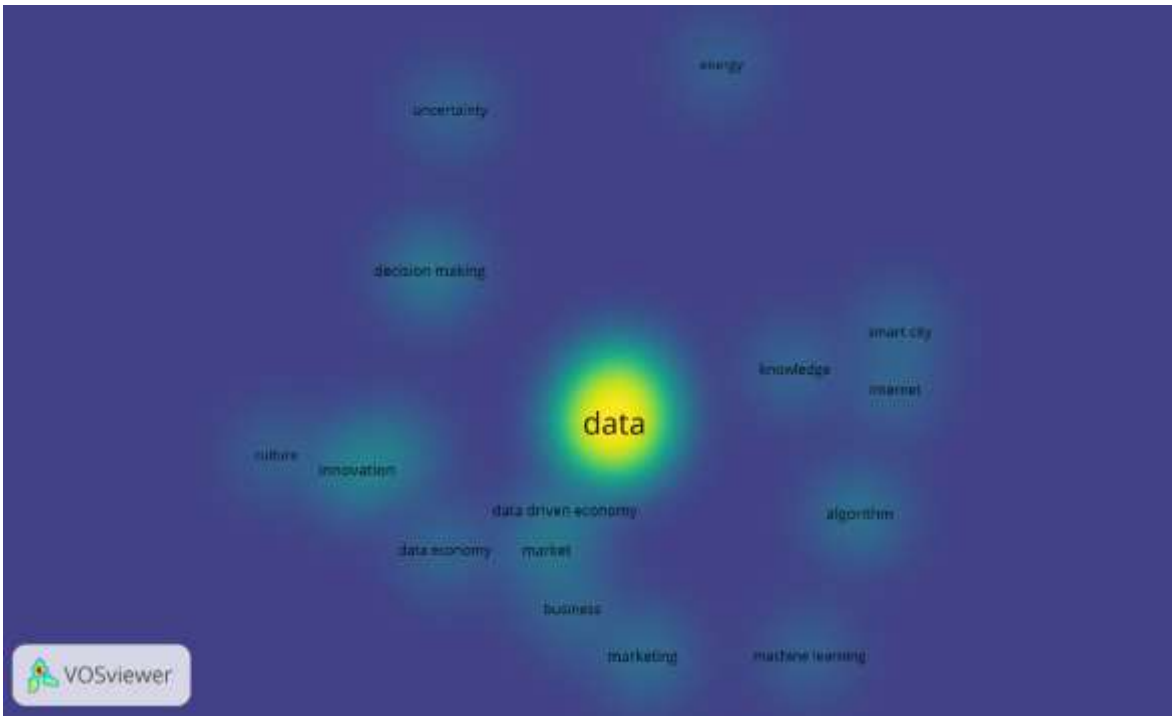

Figure 3. Density Visualization
Source: Data Analysis Result, 2024

This VOSviewer visualization employs a heat map overlay to emphasize the intensity and centrality of the term "data" within a field of interconnected concepts, marked by its vibrant yellow core fading into green and surrounded by a vast expanse of blue. The intensity of the color at the center highlights "data" as the focal point of extensive academic and practical discussions, signifying its paramount importance in the research landscape. The gradient spread of color towards the peripheries where terms like "smart city," "internet," "algorithm," and "machine learning" are positioned suggests these areas, while significant, are conceptually and perhaps empirically derivative of or influenced by core data methodologies and frameworks.

Surrounding "data," the terms are distributed in a manner that reflects their relative conceptual proximity to the central theme. The placement of terms like "business," "market," and "marketing" close to "data" underscores the practical applications of data-driven strategies in commercial contexts, indicating a strong focus on how data analytics powers decision-making processes in economic sectors. Conversely, the terms "culture" and "innovation" are somewhat distanced from the center, suggesting that while they are relevant, their integration with data-centric research is less direct or perhaps more nuanced. This layout provides a visual representation of how various domains either converge on or orbit around the pivotal concept of data, offering insights into the interdisciplinary nature of data-driven research.

4.4 Co-Authorship Network Analysis

Figure 4. Authorship Collaboration
Source: Data Analysis Result, 2024

This VOSviewer visualization illustrates the co-authorship networks among researchers, segregated into two distinct clusters, each indicated by a different color. The green cluster comprises researchers "liu_y," "zhang_y," and "wang_s," indicating a collaborative group where these authors frequently co-author papers. Similarly, the red cluster consists of "lim_mk," "tseng_ml," and "bui_td," highlighting another group of researchers working together. The separation of clusters with no interlinking lines between them suggests that while each group is internally collaborative, there is no direct co-authorship or research collaboration between the two clusters. This type of visualization is useful for identifying collaborative patterns within a field and understanding how different research groups or communities are interconnected or isolated within their specific research domains.

4.5 Practical Implication

The findings from this bibliometric analysis offer several practical implications for both academic and industry stakeholders engaged in data-driven fields. Firstly, the central positioning of "data" in the research landscape and its connections to various sectors such as "business," "marketing," and "smart cities" underscores the necessity for organizations to invest in robust data analytics capabilities. Businesses and municipal governments should consider these insights to guide their investments in technologies that enable effective data analysis, aiming to enhance decision-making processes and operational efficiencies. For instance, companies can leverage data-driven strategies to tailor marketing efforts that are more responsive to consumer behaviors and preferences, thereby increasing ROI and customer satisfaction.

Secondly, the emergent focus on "algorithms" and "machine learning" within the research suggests a growing trend towards automation and advanced analytical techniques in handling large datasets. This trend presents a strong case for educational institutions and professional training programs to incorporate advanced data analytics, machine learning, and algorithmic programming into their curricula. By doing so, they can prepare a workforce that is equipped with the necessary skills to thrive in an increasingly data-centric world. Furthermore, the rise of smart city initiatives highlighted in the analysis indicates a need for urban planners and policymakers to adopt data-driven approaches in city management and infrastructure development, ensuring smarter, more sustainable urban environments.

Lastly, the bibliometric analysis also reveals potential areas for future research, particularly in bridging the gaps between the identified clusters of collaboration. Academics and researchers should consider exploring interdisciplinary studies that merge insights from both technology-focused and business-oriented domains. Collaborative projects could be initiated to explore how data-driven strategies can be optimized across different fields, potentially leading to innovations that could transform industries. Engaging in such cross-disciplinary research could not only fill existing knowledge gaps but also spur the development of new technologies and methodologies that further enhance the efficacy and application of data analytics.

4.6 Limitation of Study

This bibliometric analysis, while comprehensive, has several limitations that should be acknowledged. Firstly, the reliance on specific databases, such as Google Scholar, may restrict the scope of publications analyzed, potentially omitting relevant studies published in sources not indexed by these databases. Additionally, bibliometric analyses inherently focus on quantitative metrics such as citation counts and publication rates, which may not fully capture the qualitative impact or the innovative nature of the research. This could lead to an underrepresentation of emerging topics that have not yet gained substantial traction in the literature. Furthermore, the analysis is constrained by the selected keywords and

time frame, which might limit the ability to detect broader or evolving trends outside of the predefined search parameters. Lastly, the lack of inter-cluster connectivity observed could be partly due to methodological limitations in tracking more informal or less documented forms of collaboration, such as interdisciplinary research or industry partnerships, which are often crucial for innovation but less visible in academic publications.

5. CONCLUSION

The bibliometric analysis conducted in this study offers a nuanced understanding of the evolution and current state of research within data-driven domains, highlighting "data" as a central theme interconnected with diverse fields like business, smart cities, and machine learning. This research provides valuable insights into the dominant trends, collaboration patterns, and thematic priorities that shape the data-driven research landscape, underscoring the critical role of data in advancing both theoretical knowledge and practical applications across various sectors. The findings suggest significant opportunities for enhancing data analytics capabilities, developing targeted educational programs, and fostering interdisciplinary collaborations that can bridge existing gaps in the literature. However, the study also notes limitations inherent in bibliometric analyses, including potential biases toward more frequently cited or recent publications and the exclusion of works outside selected databases. Overall, this analysis not only reflects the dynamic and impactful nature of data-driven research but also guides future academic and practical endeavors in the field.

REFERENCES

[1] C. I. Michael *et al.*, "Data-driven decision making in IT: Leveraging AI and data science for business intelligence," *World J. Adv. Res. Rev.*, vol. 23, no. 1, pp. 472–480, 2024.

[2] O. Abdul-Azeez, A. O. Ihechere, and C. Idemudia, "Enhancing business performance: The role of data-driven analytics in strategic decision-making," *Int. J. Manag. Entrep. Res.*, vol. 6, no. 7, pp. 2066–2081, 2024.

[3] O. T. Joel and V. U. Oguanobi, "Data-driven strategies for business expansion: Utilizing predictive analytics for enhanced profitability and opportunity identification," *Int. J. Front. Eng. Technol. Res.*, vol. 6, no. 02, pp. 71–81, 2024.

[4] A. Pabedinskaitė, V. Davidavičienė, and P. Milišauskas, "Big data driven e-commerce marketing," 2014.

[5] B. De Langhe and S. Puntoni, *Decision-Driven Analytics: Leveraging Human Intelligence to Unlock the Power of Data*. University of Pennsylvania Press, 2024.

[6] A. Kusiak, "Innovation: A data-driven approach," *Int. J. Prod. Econ.*, vol. 122, no. 1, pp. 440–448, 2009.

[7] O. Troisi and G. Maione, "Data-Driven Decision Making: Empowering Businesses through Advanced Analytics and Machine Learning," *J. Environ. Sci. Technol.*, vol. 3, no. 1, pp. 515–525, 2024.

[8] N. Donthu, S. Kumar, D. Mukherjee, N. Pandey, and W. M. Lim, "How to conduct a bibliometric analysis: An overview and guidelines," *J. Bus. Res.*, vol. 133, pp. 285–296, 2021.

[9] L. Einav and J. Levin, "Economics in the age of big data," *Science (80-.).*, vol. 346, no. 6210, p. 1243089, 2014.

[10] S. Mullainathan and J. Spiess, "Machine learning: an applied econometric approach," *J. Econ. Perspect.*, vol. 31, no. 2, pp. 87–106, 2017.

[11] H. R. Varian, "Big data: New tricks for econometrics," *J. Econ. Perspect.*, vol. 28, no. 2, pp. 3–28, 2014.

[12] D. Bholat, "Big data and central banks," *Big Data Soc.*, vol. 2, no. 1, p. 2053951715579469, 2015.

[13] D. Zhao and A. Strotmann, "The knowledge base and research front of information science 2006–2010: An author cocitation and bibliographic coupling analysis," *J. Assoc. Inf. Sci. Technol.*, vol. 65, no. 5, pp. 995–1006, 2014.

[14] L. Marinelli, "Tempo di bilanci. Dalla" lingua e letteratura polacca" agli" studi polacchi" e oltre," *Eur. Orient.*, vol. 39, pp. 9–22, 2020.

[15] S. L. Brunton and J. N. Kutz, *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press, 2022.

[16] F. Provost and T. Fawcett, "Data science and its relationship to big data and data-driven decision making," *Big data*, vol. 1, no. 1, pp. 51–59, 2013.

[17] J. Zhang, F.-Y. Wang, K. Wang, W.-H. Lin, X. Xu, and C. Chen, "Data-driven intelligent transportation systems: A survey," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 4, pp. 1624–1639, 2011.

[18] T. Johns, *Should you be persuaded: Two samples of data-driven learning materials*, vol. 4. na, 1991.

[19] J. N. Kutz, S. L. Brunton, B. W. Brunton, and J. L. Proctor, *Dynamic mode decomposition: data-driven modeling of complex systems*. SIAM, 2016.

[20] P. Mohajerin Esfahani and D. Kuhn, "Data-driven distributionally robust optimization using the Wasserstein metric: Performance guarantees and tractable reformulations," *Math. Program.*, vol. 171, no. 1, pp. 115–166, 2018.

[21] S. Yin, S. X. Ding, X. Xie, and H. Luo, "A review on basic data-driven approaches for industrial process monitoring,"

IEEE Trans. Ind. Electron., vol. 61, no. 11, pp. 6418–6428, 2014.

[22] K. Amasyali and N. M. El-Gohary, "A review of data-driven building energy consumption prediction studies," *Renew. Sustain. Energy Rev.*, vol. 81, pp. 1192–1205, 2018.

[23] S. J. Qin, "Survey on data-driven industrial process monitoring and diagnosis," *Annu. Rev. Control*, vol. 36, no. 2, pp. 220–234, 2012.

[24] E. Brynjolfsson, L. M. Hitt, and H. H. Kim, "Strength in numbers: How does data-driven decisionmaking affect firm performance?," *Available SSRN 1819486*, 2011.