

Unleashing Big Data in Entrepreneurship: Trends, challenges and prospects

Shefali Srivastava
Department of Management and
Humanities
Indian Institute of Information
Technology, Lucknow
Lucknow, India
rmh22102@iiitl.ac.in

Puneet Prashar
Rajiv Gandhi Institute Of Petroleum
Technology
Amethi, India
21mc3012@rgipt.ac.in

Priyank Kumar Singh
School of Management
Doon University
Dehradun, India
priyanksingh38@gmail.com

Krishna Mohan Shukla
Rajiv Gandhi Institute Of Petroleum
Technology
Amethi, India
21mc3009@rgipt.ac.in

Anugamani Priya Srivastava
Symbiosis Institute of Business
Management)
Symbiosis International Deemed
University
Pune, India
srivastavaanu0@gmail.com

Ta Huy Hung
Vietnam National University, Hanoi
Hanoi, Vietnam
hungth@vnuis.edu.vn

Mohit Yadav
Jindal Global Business School
O. P. Jindal Global University
Sonapat, India
mohitaug@gmail.com

Rahul Khurana
Faculty of Commerce and Management
SGT University
Gurugram, India
dr.rahulkhurana@gmail.com

Abstract—In this current study, a bibliometric analysis of 590 academic publications, published between 2012 and 2023, is offered to analyze the trends and future perspective directions for big data analytics in entrepreneurship. The purpose of the study is to examine the publishing trends, prominent publications, and noteworthy contributions within this field, including authors, institutions, and countries. Findings provide valuable assistance to researchers in selecting reputable journals for publication and present valuable insight into the key research issues, shedding light on emerging areas within this field. Researchers need to investigate the use of big data in emerging domains of entrepreneurship, including areas such as fostering diversity, addressing ethical concerns, safeguarding privacy, and inculcating inclusion. Subsequent investigations might prioritize the use of big data within the context of collective decision-making protocols, as well as the integration of big data analytics with the contemporary technological era.

Keywords— *bibliometric analysis, big data, entrepreneurship*

I. INTRODUCTION

The burgeoning field of Big Data in the realm of entrepreneurship has become a focal point of interest, as it offers a plethora of opportunities for analysis and insights. This emerging discipline encompasses the utilization of vast and complex datasets to discern patterns, trends, and correlations that can inform entrepreneurial decision-making processes [1]. By harnessing the power of advanced analytics and cutting-edge technologies, entrepreneurs can tailor their offerings to meet the specific needs and preferences of individual customers, enhancing customer satisfaction and loyalty. This personalized approach also extends to marketing strategies, allowing entrepreneurs to create targeted campaigns that resonate with their audience [2]. One significant impact of big data in entrepreneurship is the ability to personalize products and services. With the introduction of big data, the entrepreneur has seen a significant upheaval in

recent years. Big data has emerged as a transformative force in entrepreneurship, reshaping how businesses operate, make decisions, and create value. Entrepreneurs are increasingly leveraging the power of big data to gain unprecedented insights into market trends, customer behavior, and operational efficiency [3]. In the realm of entrepreneurship, big data refers to the massive volumes of structured and unstructured data generated by various sources, including social media, online transactions, sensors, and more. The ability to collect, process, and analyze this vast amount of information provides entrepreneurs with a competitive edge in understanding their target markets. Through advanced analytics, entrepreneurs can identify patterns, preferences, and emerging opportunities, enabling more informed and strategic decision-making [4].

Big data refers to a significant amount of organized, semi-structured, or unstructured data that may be collected, processed, and analyzed using non-traditional ways. It involves the use of advanced algorithms for data analysis, decision-making, and prediction [5]. Utilizing big data in entrepreneurship has the potential to improve and expedite creativity and innovation. Furthermore, big data enables entrepreneurs to engage in predictive analytics, allowing them to foresee market trends, anticipate customer requests, and enhance supply chain management. Having this ability to anticipate future events enables entrepreneurs to take proactive measures to adjust to changing market circumstances, hence minimizing risks and strengthening the overall resilience of their firms [6]. The age of big data has also seen a thriving growth of collaborative ecosystems. Entrepreneurs can access external data sources, establish partnerships with other organizations, and engage in industry collaborations to improve their comprehension of the market and discover fresh prospects. It is crucial to acknowledge that while large data might possess a great level of accuracy, they

also have some limitations [7]. To achieve optimal performance, these models need a substantial volume of data for training. Additionally, there is a risk of overfitting the training data [8], which might result in less precise predictions when confronted with fresh data [9]. To fully explore the potential of big data in entrepreneurship, it is crucial to do a thorough bibliometric analysis of different types of big data. This analysis will help evaluate the current knowledge, find patterns in research, and establish the future trajectory of this topic.

A bibliometric analysis provides a systematic and quantitative approach to assess a large number of academic papers. It offers valuable insights into the current state of research, prominent researchers, influential journals, important research topics, and emerging areas of interest [10]. This approach enables the extended observation of patterns in a diverse array of domains. Big data analysis is another facet of bibliometric analysis. There has been a significant increase in the number of Bibliometric evaluations, particularly in the field of management and business. In addition, scientists use bibliometric analysis to examine disciplines such as artificial intelligence, machine learning, and networks. The proliferation and ease of use of bibliometric software such as Gephi, R-Bibliometix, Biblioshiny, and VOSviewer, together with the widespread availability of scientific databases like Scopus and Web of Science, are the primary drivers behind the increasing popularity of bibliometric research. Additionally, we conducted bibliometric research on the topic of big data in entrepreneurship.

Therefore, in this study we address the following Research Questions [RQ]: [RQ1] What are the enduring patterns and trends in the frequency and publication patterns of big data and related studies in the field of entrepreneurship from 2012 to 2023? [RQ2] Which academic journals have made substantial contributions, in terms of publication frequency, to the study of big data in the context of entrepreneurship? [RQ3] Which organizations, governments, and scholars exhibit the highest level of engagement within this particular domain of research? [RQ4] Which scholarly articles have garnered the highest number of citations, therefore exerting the most impact? By using bibliometric analysis, these research questions have the potential to provide valuable insights into the current status, patterns, and prominent contributors within the specific field of study. Our study significantly advances the use of big data in the field of entrepreneurship. This study presents a thorough bibliometric review that examines the methodologies used in this particular topic, providing valuable insights into the prevailing trends and advancements. Additionally, we propose a research agenda for future endeavors in order to guide further exploration and foster innovation.

II. DATA METHODOLOGY

Bibliometric approaches may be used by the researchers to derive their findings by analyzing aggregated data of bibliographic, given by the authors', who express their viewpoints via written works, collaboration, and citation practices. The collected data may be aggregated and examined in order to get valuable insights pertaining to the structure, networking, and areas of interest within the field. The following are the procedures for article selection, screening, and bibliometric analysis.

A. Data Selection

This study used the Scopus preserve, which is wellrecognized globally and offers comprehensive coverage of academic publications. Furthermore, it is a very suitable approach for conducting literature reviews within the realm of social science. The first step in the analytical procedure was doing a comprehensive search for relevant scholarly papers in the Scopus database. This search was conducted using specific keywords: TITLE-ABS-KEY ("big data*" AND "entrepreneur*") AND PUBYEAR 2011 AND PUBYEAR 2024 AND (LIMIT-TO (PUBSTAGE , "final")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (SUBJAREA , "COMP") OR LIMIT-TO (SUBJAREA , "BUSI") OR LIMIT-TO (SUBJAREA , "SOCT") OR LIMIT-TO (SUBJAREA , "ENGI") OR LIMIT-TO (SUBJAREA , "DECI") OR LIMIT-TO (SUBJAREA , "ECON") OR LIMIT-TO (SUBJAREA , "PSYC")). This study focuses on publications that have been published in peer-reviewed journals in the English language, since these articles are well recognized for their high level of quality and rigor.

B. Article Screening

A detailed database search was done in Dec 2023, and more than 700 papers were found. The initial sample was then filtered to extract articles about big data and entrepreneurship using a set of inclusion and exclusion criteria. Medical science-related articles, such as those on dentistry, were not included. Five hundred ninety articles were finally included if they had a big data and entrepreneurship focus. Each author reviewed the papers separately to guarantee the validity of the study. They convened later to talk about any issues and resolve any differences.

C. Article Screening

This analysis was conducted to get a macro-level perspective, as opposed to a systematic literature review. The complexity of bibliometric analysis is often acknowledged due to its involvement in several processes that use a wide range of analyses and mapping software tools. Moreover, considering that a majority of articles lack empirical evidence, it is more appropriate to use bibliometric analysis rather than doing a meta-analysis. The data analysis of the prominent authors and publications was conducted using the "Bibliometrix" module in the R programming language [10]. In addition, Biblioshiny and VOS viewer were used as supplementary tools to address the visual limitations of the "Bibliometrix" program.

III. KEY RESULTS

A. Affiliation statistics and geographic location

According to Table I, the most prolific institutions include Jilin University with 14 papers in Publication, Notreported has published 8 Papers, and Wuhan Donghu University has published 8 papers in the area of big data in entrepreneurship.

TABLE I: MOST PROLIFIC INSTITUTION

Affiliation	Articles
Jilin University	14
Notreported	8
Wuhan Donghu University	8
University Of Florence	7
Hunan University	6
Shanghai University	6
University Of Salento	6
Wuhan Business University	6
Zhejiang University City College	6

Table II indicates that the USA has the highest number of publications (168), followed by the USA (29) and the United Kingdom (22), with corresponding authors from each respective nation. Fig. 1 depicts the worldwide distribution of nations engaged in the field of big data in entrepreneurship. The MCP ratio refers to the percentage of an author's total papers that have been published in collaboration with authors from other nations.

TABLE II: CORRESPONDING AUTHORS' COUNTRY-BASED DISTRIBUTION

Country	Articles	SCP	MCP	Freq	MCP_Ratio
China	168	149	19	0.28571429	0.11309524
USA	29	24	5	0.04931973	0.17241379
UK	22	11	11	0.03741497	0.5
Italy	19	12	7	0.03231293	0.36842105
Germany	13	9	4	0.02210884	0.30769231
India	13	10	3	0.02210884	0.23076923
Australia	10	4	6	0.01700680	0.6

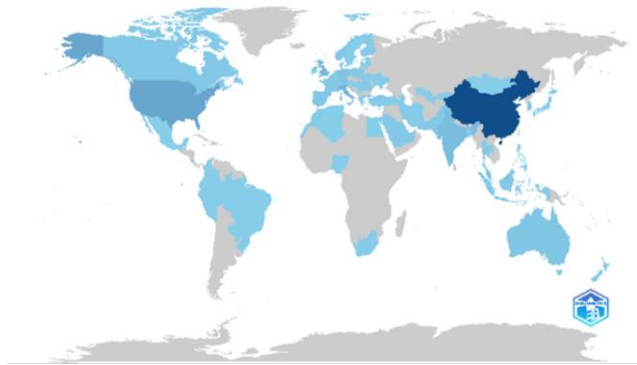


Fig. 1. Country-wise scientific production

B. . Most Cited Publication

Table III presents the 10 most often referenced papers on big data in entrepreneurship from 2012 to 2023. Among these 10 works, four were published in the year 2017.

C. Keyword Analysis

The primary objective of bibliometric big data is to facilitate the analysis of extensive and intricate bibliographic data by visually representing fundamental elements of the data. In addition, we conducted a visual bibliometric analysis of the keywords found in the papers on big data in entrepreneurship, as seen in Fig. 2 Keyword analysis yields a total of four clusters. The first research cluster focuses on investigating decision-making in large data pertaining to IoT and AI, as well as subgroup variations, particularly in the realm of innovation. This cluster largely focuses on understanding the functioning of big data, particularly in several subcategories such as IoT, AI, economics, entrepreneurship, and innovation. This series of research investigates many aspects of cognitive capacity, achievement, and innovation. The objective is to get a deeper understanding

of the functioning of big data across numerous domains and their impact on diverse decision-making and data analytics processes.

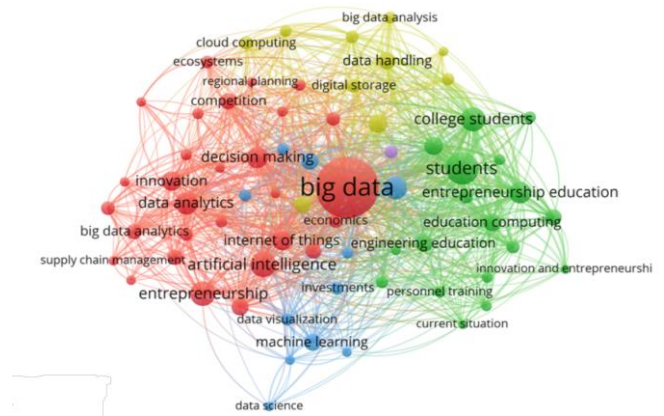


Fig. 2. Co-occurrence of Scopus Keywords using VOS viewer

The Second cluster centers on college students and the present state of big data. This cluster focuses on advanced research on big data that goes beyond the scope of the first cluster. It includes studies on entrepreneurship and engineering education. Research conducted in this cluster examines the present state and effectiveness of big-data analytics in educational settings. They may engage in the innovation and analysis of educational computer models in order to comprehend their operation. Through the analysis of extensive educational data, academics strive to acquire a more profound comprehension of basic concepts that may be implemented in many fields, including entrepreneurship-based education. The third cluster is on the utilization of big data in investing and the application of machine learning techniques. This cluster focuses on the correlation between large-scale data and investing practices.

This analysis explores the use of big data in data visualization and its correlation with machine learning. Research in this cluster may focus on the use of big data in specialized financial operations, such as investing approaches or data science. The goal is to understand how big data may improve outcomes or get insights into the underlying processes of certain investments or data visualization. The cluster emphasizes the practical uses of big data research in machine learning, particularly in the context of financial behavior. This research aims to establish knowledge of human behavior that can be used to financial analysis. The last cluster focused on the utilization of big data in cloud computing and the management of data. This cluster focuses on the correlation between big data and big data analytics, data management, and data storage. This study explores the use of big data in cloud computing and its correlation with data analytics, management, and storage. Research in this cluster may focus on the use of big data in specialized analytical techniques, such as data processing methods or storage processes.

TABLE III: TOP LOCAL CITED ARTICLES

Document	DOI	Year	Total Citation	TC Per Year	Normalized TC
Kitchin R, 2014, Geojournal	10.1007/S10708-013-9516-8	2014	1623	162.3	11.3232558
Makridakis S, 2017, Futures	10.1016/J.Futures.2017.03.006	2017	673	96.1428571	18.3583385
Dubey R, 2020, Int J Prod Econ	10.1016/J.Ijpe.2019.107599	2020	273	68.25	22.7699727
Moore P, 2016, New Media and Society	10.1177/1461444815604328	2021	200	25	8.3333333
Ciampi F, 2021, J Bus Res	10.1016/J.Jbusres.2020.09.023	2021	191	63.6666667	34.5388788
Rippa P, 2019, Technol Forecast Soc Change	10.1016/J.Techfore.2018.07.013	2019	174	34.8	11.8416667

Ehret M, 2017, J Mark Manage	10.1080/0267257X.2016.1248041	2017	161	23	4.39181649
Carolan M, 2017, Sociol Ruralis	10.1111/Soru.12120	2017	154	22	4.20086795
Lytras Md, 2017, Int J Semant Web Inf Syst	10.4018/IJSWIS.2017010101	2017	145	20.7142857	3.95536268
Krivy M, 2018, Planning Theory	10.1177/1473095216645631	2018	136	22.6666667	12.2013129

IV. CONCLUSION AND FUTURE AGENDA

The integration of big data analytics in entrepreneurship has proven to be a transformative force, offering unprecedented opportunities for growth, innovation, and strategic decision-making. In conclusion, this research thoroughly examines the current body of research on big data in the field of entrepreneurship. An extensive analysis of 590 articles from Scopus using bibliometric methods has provided valuable insights into the present level of knowledge, emerging patterns, and future prospects in this field. The data provide valuable insights into publication trends, prominent authors, distribution channels, and emerging study areas. Academics, professionals, and decision-makers may use these insights to identify specific topics for additional study, viable areas for cooperation, suitable publications to target and to identify research gaps. The report proposes many prospective research domains for future scholars in consideration of the discoveries. The abundance of data produced and analyzed, using big data technology has enabled businesses to acquire a more profound understanding of customer behavior, market trends, and operational effectiveness. The use of big data has empowered entrepreneurs to make choices based on facts, reduce risks, and discover novel business prospects.

Analyzing massive datasets has become crucial for both startups and existing firms, as it promotes agility and adaptation in the always-changing business environment. Further investigation is necessary to examine the impact of big data on optimism, culture, and organizational results. The scope of data collection is a possible limitation of this research. For the study, only the Scopus database and Englishlanguage papers were used as sources. The lack of relevant research produced in languages other than English or indexed in diverse databases may be attributed to the exclusion of different databases and languages. As a consequence, the study may have overlooked important contributions from non-English literature and other academic databases, which limits the comprehensiveness and applicability of the study.

REFERENCES

- [1] M. Obschonka and D. B. Audretsch. Artificial intelligence and big data in entrepreneurship: a new era has begun. *Small Business Economics*, 55:529–539, 2020.
- [2] H. D. Mohammadian and F. Rezaie. Big data for entrepreneurship towards csr and sustainable development. In *Big Data for Entrepreneurship and Sustainable Development*, pages 21–48. CRC Press, 2021.
- [3] M. A. Barchiesi and A. F. Colladon. Big data and big values: When companies need to rethink themselves. *Journal of Business Research*, 129:714–722, 2021.
- [4] A. Schwab and Z. Zhang. A new methodological frontier in entrepreneurship research: Big data studies, 2019.
- [5] D. Acemoglu and P. Restrepo. Artificial intelligence, automation, and work. In *The economics of artificial intelligence: An agenda*, pages 197–236. University of Chicago Press, 2018.
- [6] P. Church, H. Mueller, C. Ryan, S. V. Gogouvitis, A. Goscinski, H. Haitof, and Z. Tari. Scada systems in the cloud. *Handbook of Big Data Technologies*, pages 691–718, 2017.
- [7] A. Agrawal, J. Gans, and A. Goldfarb. Prediction, judgment, and complexity: a theory of decision-making and artificial intelligence. In *The economics of artificial intelligence: An agenda*, pages 89–110. University of Chicago Press, 2018.
- [8] X. Ying. An overview of overfitting and its solutions. In *Journal of physics: Conference series*, volume 1168, page 022022. IOP Publishing, 2019.
- [9] P. Davidsson, J. Recker, and F. V. Briel. External enablement of new venture creation: A framework. *Academy of Management Perspectives*, 34(3):311–332, 2020.
- [10] M. Aria and C. Cuccurullo. bibliometrix: An r-tool for comprehensive science mapping analysis. *Journal of informetrics*, 11(4):959–975, 2017