ISSN (ELECTRONIC): 2945-3135

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH & REVIEWS

journal homepage: www.ijmrr.online/index.php/home

A BIBLIOMETRIC METHOD TO REVIEWING MACHINE LEARNING FOR BIG DATA ANALYTICS INVESTIGATING THE DIMENSION DATABASE

Abhishek Kumar[1], Priyanshu Singh[2]

[1]Assistant Professor, Maulana Mazharul Haque A & P University, Patna, Bihar, India.

[2] Scholar, BCA, Maulana Mazharul Haque A & P University, Patna, Bihar, India.

How to Cite the Article: ***Kumar, Abhishek, & Singh, Priyanshu. (2024).*** *A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115*

DOI Information: https://doi.org/10.56815/IJMRR.V314.2024/103-115.

Keywords

Bibliometrics, Big data, Machine learning, Clustering,

Abstract

Big data has become more popular as a subject of study in recent years, and its rise has been widespread. The purpose of this work is to use text mining-based analytic techniques along with bibliometrics to capture the scientific structure and subject progression of big data research. From the Dimensions database, bibliographic information on big data journal papers published between 2024 and 2014 was gathered and examined. The findings indicate a notable increase in publications since 2014. The main journals, most referenced papers, most prolific authors, nations, and institutions are all highlighted by the study's findings. Second, an original method for locating and examining key research themes in big data papers was put forward. Each cluster of keywords was identified as a topic once it had been clustered. Additionally, to track the thematic progression, the papers were split into four sub-periods. The topic mapping indicates that big data analytics—which includes techniques, tools, supporting infrastructure, and applications—dominates big data research. Security and privacy are two other essential components of big data research. Big data is mostly generated by social networks and the Internet of things, and cloud computing's resources and services make big data processing and administration much easier.

1. INTRODUCTION

Many technical advancements have been observed in the last few decades in the areas of cloud computing, social sensing, data storage, cloud computing, sensor networks, Internet of Things (IoT), and medical and financial information systems. Data is the engine of the modern world and has a profound effect on society and the knowledge-based economy. drawn scholars from

interdisciplinary fields like the biological sciences, social sciences, natural sciences, and physical sciences (Tosi D. a., 2021) (Davoudian, 2020) (Kushwaha, 2020). As a result, data processing and management have entered a new era that is revolutionizing many scientific and technical domains. The information explosion that resulted from the broad use of information and communication technologies is where the idea of big data first emerged. In 2005, O'Reilly Media's Roger Mougalas first used the term "big data." (Van Rijmenam, 2019). The concept of big data is not new;

rather, it is made possible by advancements in memory and storage technologies, cloud and parallel computing, database technology, and cutting-edge analytic tools. Big data is defined as "a new generation of technologies and architectures, designed to economically extract value from huge volumes of a wide variety of data, by enabling high-velocity capture, discovery, and/or analysis," by the IDC, a pioneer in the field of big data research. (Parlina, 2020). It speaks about an enormous amount of data that is challenging to handle and analyze using conventional techniques and equipment. But using the term "big" to describe the size of the data is insufficient. In addition to large volumes, high-dimensional heterogeneous complex data that are probably noisy and incomplete are also frequently referred to as "big data." Despite the lack of agreement on a universally accepted definition, big data is commonly defined by the four Vs: volume, variety, velocity, and veracity. According to Doug Laney, there are three aspects to the potential and difficulties presented by the expansion of data: increasing volume, pace, and diversity (Laney, 2001). The enormous amount of data that is continuously produced in the digital age is referred to as volume. While variety refers to a growing range of data kinds or formats, including text, pictures, networks, computer simulations, and geographic maps, velocity refers to the speed at which data are created and transferred from one place to another. It is crucial for scientists to recognize the significance of research publications and research trends (Crespo J. A., 2014) (Wohlin, An analysis of the most cited articles in software engineering journals-2000., 2007).This presents a number of difficulties for traditional computing methods when it comes to processing and storing large data. (Ward, 2013) (De Mauro, 2016). In (Dobre, 2014), Dobre and Xhafa reportedly, 2.5 quintillion bytes, or exabytes (EB), or 1018 bytes, of data are produced day worldwide. The volume of data created increased dramatically in 2018. By 2020, 40 zettabytes (ZB, or 103 EB) of data will be created, 90% of which will be unstructured, predict (Gantz, 2012).According to Domo's seventh annual Data Never Sleeps infographic, 4,416,720 terabytes of internet data are used worldwide every minute (James, 2019). In recent years, there has been an expansion in scholarly literature on big data. Numerous scholarly works have been released addressing the theory, practices, infrastructure, and uses of big data technologies. A high-value research subject is to extract knowledge and relevant information from these big data papers already published in order to determine the present and future trends in this sector. (Xu, 2019). But because every speaker focuses on a different component of big data, it's challenging to come to an agreement on what big data actually is (Hu, 2014). One such technique for doing quantitative analysis of research articles is bibliometrics (Ho, op-cited articles in chemical engineering in Science Citation Index Expanded: A bibliometric analysis., 2012) (Ho, 2014). Bibliometrics uses citation counts as a primary criterion to determine the most significant works in a certain topic. Citations, in turn, facilitate the construction of new works on top of the body of knowledge by establishing a link between the innovative methods and those of their forebears. Citations can also be used to contextualize one's own work, give background information, and acknowledge the contributions of peers (Reuters, 2018). Another useful measure of a scientific

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

publication's impact and visibility is its citation count. Numerous areas employ bibliometrics and citation analysis to determine the most significant works and researchers, as well as to examine and track the development of a particular research topic (Wohlin, 2007). It is used in many scientific fields, including medicine (Davis, Author-choice open-access publishing in the biological and medical literature: A citation analysis., 2009), physics (Gingras, 2010), social sciences (Ahmad, A decade of big data literature: analysis of trends in light of bibliometrics., 2020), and computer sciences (Garousi V. a., 2016). This work's goal is to use bibliometric analysis to look at how the big data literature has evolved. The purpose of this study is to investigate global research trends in the field of big data and related topics. R&D management in this subject, scientists, scholars, practitioners, and students may find the information and knowledge generated useful. The analysis comprises approximately 4378 data and is based on publications published over ten years (2014–2024). This is how the rest of the article is structured. A concise summary of research on the topic of bibliometrics is provided in Section 2. Research questions are presented in Section 3. A comprehensive examination of the data is presented in Section 4. and lastly, the task is concluded in Section 5.

2. LITERATURE REVIEW

Numerous studies in a variety of domains (Crespo J. A.-C., 2014), (Davis, 2009), (Ahmad, 2020) have focused on citation analysis. After a brief discussion of bibliometrics research in connection to "big data," we offer the most important academic works in citation analysis.

Examining 15 years of Big Data research, Davide Tosi's 2024 (Tosi D. R., 2024) study identifies analytics application domains, difficulties, and future research prospects. It highlights the requirement for machine learning integration and energy-efficient data collection, with ethical concerns in decision-making. Significant research has been conducted in the healthcare industry since 2012, and big data applications have spread to industrial systems and the Internet of Things. Integrated frameworks and approaches should be the main topic of future study.

The 2024 (Faaique, 2024) research by Muhammad Faaique explores the difficulties associated with using big data in astronomy and emphasizes the necessity for better data visualization, continuous data quality control, and photometric redshift estimation. It emphasizes the value of cooperation within scientific domains and the possibilities for using data mining and self-organizing maps to address problems like galaxy morphology.

The 2023 (Aydin, 2023) research by Ahmet Arif Aydin offers a big data value chain model, a comparative viewpoint on big data technologies, and recommendations for choosing the right technologies. It covers use scenarios, explains big data properties and processing paradigms, and discusses difficulties in developing data-intensive systems. The research does, however, admit many limitations, including the possible lack of public access and the omission of publications written in languages other than Turkish or English.

The objective of Nima Sedaghat's present by (Sedaghat, 2022) 2022 are to interpret the representations of deep convolutional neural networks, discover relationships with physical awareness, and acquire generic representations free from task bias. The network discovers six useful nodes by learning stellar properties without explicit task descriptions using a population sample of 270,000-star spectra. Four more nodes could provide fresh insights into astrophysical trends. More research is required to investigate possible novel patterns and the effect of disentanglement weight.

The (Imran, 2021) 2021 research by Sohail Imran examines the features of big data in the healthcare

industry, pointing out obstacles in the field of big data analytics (BDA), examining BDA applications that make use of NoSQL databases, putting forth a Med-BDA architecture, and offering helpful implementation tactics. The benefits of BDA include more accurate diagnosis and financial savings, although there are still issues that need to be investigated. The report offers companies success techniques and a path for implementing BDA in the healthcare industry. Med-BDA will be implemented at Agha Khan University Hospital among other projects.

The 2020 study by Anne Parlina (Parlina, Theme mapping and bibliometrics analysis of one decade of big data research in the scopus database., 2020) examines the worldwide trends, future directions, subject evolution, big data research structure, and important research issues. The study highlights privacy, security, social networks, IoT, and cloud computing for big data management, drawing on 7274 journal papers published between 2009 and 2018. Nevertheless, more content analysis is required because the dataset is restricted to the Scopus database.

In his 2019 work, Amit Kumar Tyagi (Tyagi, 2019) examines the advantages and disadvantages of machine learning in the big data environment, as well as data mining technologies. It emphasizes the efficacy of different machine learning approaches, but it also draws attention to problems with security and privacy, integration, interpretation capacity limitations, and time limits. In spite of these obstacles, Tyagi encouraged researchers to investigate big data and machine learning.

Iftikhar Ahmad's 2018 (Ahmad, 2019) study identifies top cited papers in Big Data research, revealing that over 50% of publications receive no citations and the average citations per publication is 3.17. China surpasses the US in Big Data research contributions, with survey papers being the most common. The study also highlights limited diversity in research topics and low contributions from the Turkish software engineering community.

In this article (Zhou, 2017), Lina Zhou highlights exceptional studies, offers a taxonomy of ML algorithms, and addresses the potential and difficulties of machine learning (ML) on large data. She presents a framework for machine learning on large data (MLBiD) and emphasizes the need for fresh approaches to overcome technological obstacles. Additionally, covered in the study are the significance of explainable machine learning for successful user assistance, data privacy protection, and system decision comprehension.

The 2016 (Qiu, 2016) study by Junfei Qiu examines learning from large data sets, emphasizing big data research initiatives and problems. It discusses the properties of big data, such as volume, variety, velocity, truthfulness, and value, and looks at the relationships between machine learning and signal processing approaches. Future research directions and the necessity of enhanced learning methodologies are also covered in this article.

In her 2015 (Al-Jarrah, 2015) review, O.Y. Al-Jarrah presents new ideas for researchers and engineers on sustainable, energy-efficient machine learning. High computing costs prevent effective data exploration in current methods, which are inefficient for huge data quantities. The paper highlights the need for scalable machine-learning systems for many sectors by discussing model energy efficiency and computing needs.

In 2014, (Garrett, 2014) M.A. Garrett conducted a study to investigate the possibilities of Big Data analytics in astronomical research. The study focused on the effects of cognitive computing on data analysis, large- scale data processing, and SETI anomaly detection. To fully utilize the scientific potential of massive data sets, the research emphasizes the necessity of effective data visualization methods, predictive analytics, and cooperation between astronomers and data analytics specialists.

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

3. RESEARCH QUESTIONS

Following Garousi and Fernandes' methodology (Garousi V. a., 2014), we develop a series of research questions and utilize these as the foundation for our data analysis. Selecting research questions with the primary goal of identifying the most cited articles, the contributions of different nations to the advancement of big data research, and the identification of important study fields. The list of research questions looks like this:

a) To determine the time-related publishing patterns in the Big Data field?

b) To determine which journals, rank highest for Big Data publications.

c) To identify the top areas of Green HRM research with respect to publications

d) In terms of the number of publications, which nations and organizations have made the most contributions?

e) What is the trend in authorship? What is the typical author count for each publication?

4. RESULTS

We address the research questions presented in this part.

4.1 Trend of Publication from 2014 to 2024:

Provides Dimensions data for big data analytics articles in the period between 2014 and 2024 related to machine learning. For example, from three papers published in 2014 to sixty articles published in 2024, there has been an increase in publications. Publications on Big Data significantly between 2015 and 2023. During the previous three years, 3,670 papers, or 84 percent of the total, have been published out of 4378 publications. This increasing trend suggests that academics are becoming aware of the term "Big Data."

Sr. No	Year	Publication
1.	2014	3
2.	2015	16
3.	2016	26
4.	2017	73
5.	2018	178
6.	2019	368
7.	2020	657
8.	2021	782
9.	2022	766
10.	2023	816
11.	2024	693

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

4.2 Leading Journals Publishing on Big Data

The second goal of our research was to identify the esteemed journals that disseminated the greatest number of papers about green HRM. Between 2014 and 2024, a total of 4378 papers were published in various journals. The list of top journals publishing on green HRM is shown in below. The top 15 journals have published 47% of all publications. The has IEEE access journal seen the highest number of research articles produced.

Sr. No	Journals	Publications	Citations
1.	IEEE access	850	38876
2.	Sustainability	458	15221
3.	Applied sciences	298	7399
4.	IEEE internet of things journal	220	13397
5.	IEEE transactions on industrial information	139	11770
6.	Electronics	195	3651
7.	Journal of cleaner production	94	7569
8.	Procedia computer science	246	3804

For instance, 3: The most prestigious publications that published articles on green human resource management between 2014 and 2024, taken from the Dimensions database

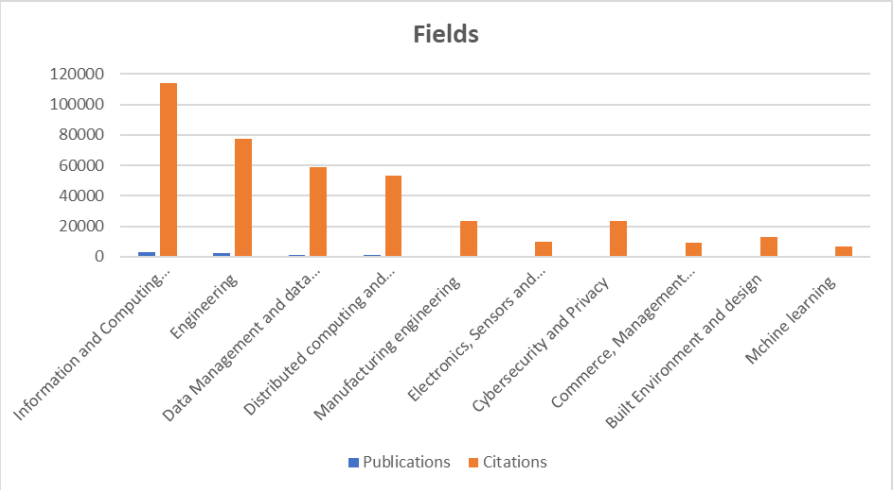
4.3 Top Areas of Green HRM Research

Instance 3 indicates that the domains leading the way in Big Data research are Information and Computing Sciences, Engineering, and Electronics, Sensors, and Digital Hardware. The relationships between the field of big data and fields like design and the built environment, cybersecurity and privacy,

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

as well as commerce, management, and tourism, are illustrated in Illustration 3. This implies that big data is an interdisciplinary area.

Sr. No.	Fields	Publications	Citations
1.	Information and Computing Sciences	3207	1,14,222
2.	Engineering	2217	77,429
3.	Data Management and data science	1346	58,747
4.	Distributed computing and Systems Software	1018	53,303
5.	Manufacturing engineering	468	23,286
6.	Electronics, Sensors and Digital Hardware	380	9,765
7.	Cybersecurity and Privacy	376	23,680
8.	Commerce, Management and tourism and	313	9,343
9.	Built Environment and design	292	13,037
10.	Mchine learning	196	6996



For instance, 3: Number of Green HRM publications in top subject areas during period 2014–2024, retrieved from Dimensions database.

4.4 Geographical and Institutional Contributions

Our final study objective was to determine which prestigious nation published the most articles on big data. A total of 4378 publications were published in different countries between 2014 and 2024. The top ten nations that publish on big data are shown below. The top 10 nations' annual contribution is shown. The top 15 countries have published 25% of all papers. Geographically, China has provided the most publications (520), the US (313), and India (194) in order of contribution.

Sr. No	Country	Publication	Citations

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

1.	China	520	28837
2.	Saudi Arabia	245	8801
3.	United States	313	18169
4.	Pakistan	162	7971
5.	United Kingdom	212	11408
6.	India	194	9439
7.	South Korea	207	7595
8.	Australia	153	11503
9.	Canada	116	6609
10.	Spain	161	5584
11.	Italy	158	6888
12.	Malaysia	99	3806
13.	United Arab Emirates	63	2486
14.	Germany	84	5057
15.	Taiwan	82	2821

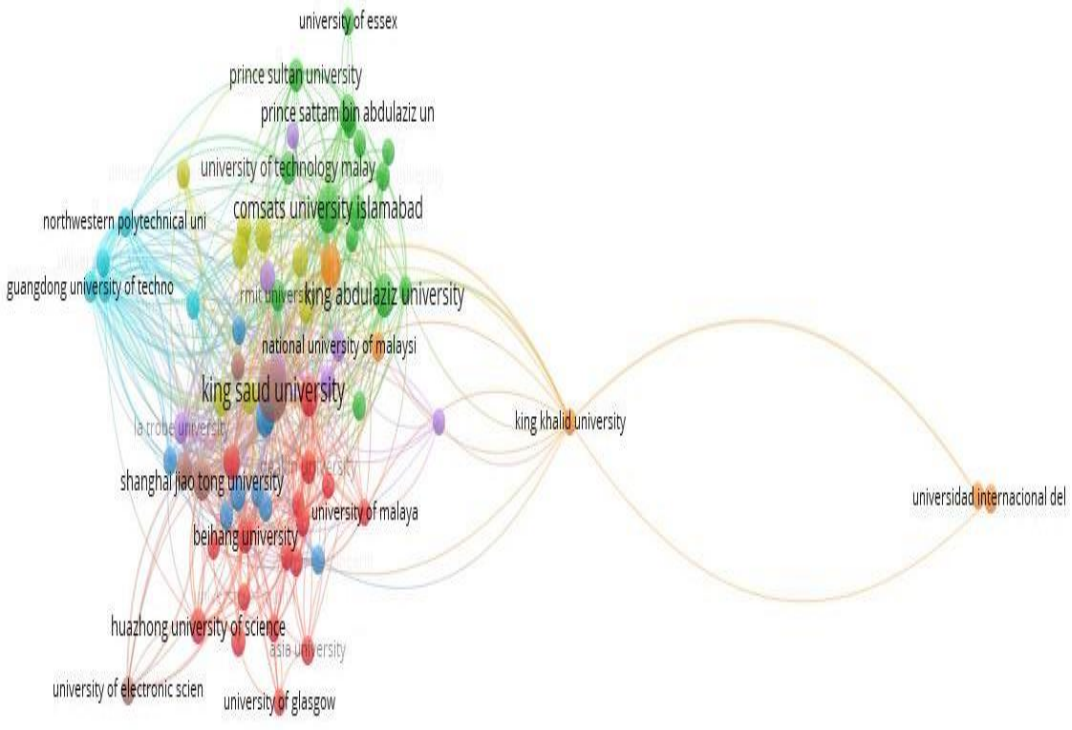
Institutional:

Instance 4 indicates that when it comes to Big Data research, King Saud University's organizations are at the top. The relationships between the Institute of Big Data and fields like machine learning, Comsats university Islamabad, Shanghai Jiao Tong University, and South China University of Technology are

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

illustrated in Illustration 4. This implies that big data is an interdisciplinary field. Illustration 4 illustrates an unusual reality that other universities, like Deakin University and Prince Sultan University, lack literature.

Sr. No	Organisation	Publications	Citations
1.	King Saud university	63	3008
2.	South China university of technology	20	2250
3.	Shanghai jiao tong university	24	2024
4.	Comsats university Islamabad	36	1556
5.	Deakin university	17	1472
6.	Sejong university	28	1713
7.	University of technology Sydney	22	919
8.	Hong Kong polytechnic university	18	893
9.	Prince sultan university	18	302
10.	Prince sattam bin Abdulaziz university	22	280



Authorship Trends:

Instance 5 indicates that the top researchers in big data are Wan, Jiafu, Li, di, Hossain, and M. Shamim, Kumar, and Neeraj. Figure shows how the author of Big Data is related to experts like Gehlot, Anita, Kaddoum, Georges, Priyadarshi, Neeraj, Twala, and BhekiSipho. This implies that Big Data's creator

is greater than. Illustration 5 highlights the odd fact that there aren't many researchers in other fields, like Garcia-Ortiz, Joselin, Sanchez-Vetteri, and Santiago.

Sr. No	Author	Publication	Citations
1.	Wan, Jiafu	10	2087
2.	Li, di	8	1011
3.	Hossain, m. Shamim	11	966
4.	Liu, chengliang	6	912
5.	Kumar, Neeraj	13	575
6.	Gehlot, Anita	9	368
7.	Kaddoum, Georges	6	338
8.	Priyadarshi, Neeraj	7	246
9.	Twala, BhekiSipho	6	224
10.	Garcia-Ortiz, joselin	6	33

5. CONCLUSION

Big data publication citations were employed in the study to solve many research issues. It was found that from 2014 to 2024, not a single publication in the top 10 was released. Big data research is dominated by big data analytics, which includes methodologies, auxiliary infrastructure, and applications. Privacy and security concerns are essential to big data research. Social networks and the Internet of Things are the main sources of big data, while cloud computing resources enable its processing and management. Among the papers with the most references is a survey study on the subject. After China, the US comes in second place in terms of publications. An average publication has 11.8 citations, while an average publication has 7.14 authors. Citation count is the research paper's statistic, and each citation is given an equal evaluation. It is advisable to identify important works and do context-based citation analysis. The unique method for identifying the primary research questions may be used to various academic and technological fields. However, the study has limitations as well, including search words, bibliographic database selections, and publication-to-citation latency. Future research should develop a uniform bibliometric schema and provide a framework for knowledge extraction in order to facilitate data integration and transfer between indexing systems.

6. AUTHORS CONTRIBUTION

The authors agreed to have no connections or engagements with any group or body that provides financial and non-financial assistance for the topics and resources covered in the article.

7. ACKNOWLEDGEMENT

Individuals / resources participated in the work are acknowledged properly.

Kumar, Abhishek, & Singh, Priyanshu. (2024). A Bibliometric method to reviewing Machine learning for Big Data analytics Investigating the Dimension Database. International Journal of Multidisciplinary Research & Reviews. 3(04), 103-115.

8. SOURCES OF FUNDING

The authors received no financial aid to support the study.

9. PLAGIARISM POLICY

The authors declare that any kind of violation of plagiarism, copyright, and ethical matters will be handled by all authors. Journalists and editors are not liable for the aforesaid matters.

10. CONFLICT OF INTEREST

The authors declared that no potential conflicts of interest concerning the research, authorship, and/or publication of this article.

REFERENCE:

[1] Ahmad, I. G. (2019). A decade of big data literature: analysis of trends in light of bibliometrics. The Journal of Supercomputing, 76, 3555-3571.

[2] Ahmad, I. G. (2020). A decade of big data literature: analysis of trends in light of bibliometrics. The Journal of Supercomputing, 2020, 3555-3571.

[3] Al-Jarrah, O. Y. (2015). Efficient machine learning for big data: A review. Big Data Research, 2(3), 87-93.

[4] Aydin, A. A. (2023). A comparative perspective on technologies of Big Data value chain. IEEE Access.

[5] Crespo, J. A. (2014). The effect on citation inequality of differences in citation practices at the web of science subject category level. Journal of the Association for Information Science and Technology, 65(6), 1244-1256.

[6] Davis, P. M. (2009). Author-choice open-access publishing in the biological and medical literature: A citation analysis. Journal of the American Society for Information Science and Technology, 60(1), 3- 8.

[7] Davoudian, A. a. (2020). Big data systems: A software engineering perspective. ACM Computing Surveys, 53(5), 1-39.

[8] De Mauro, A. M. (2016). Beyond data scientists: a review of big data skills and job families. Proceedings of IFKAD, pp. 1844-1857.

[9] Dobre, C. a. (2014). Intelligent services for big data science. Future generation computer systems, 37, 267-281.

[10] Faaique, M. (2024). Overview of big data analytics in modern astronomy. International Journal of Mathematics, Statistics, and Computer Science, 2, 96-113.

[11] Gantz, J. a. (2012). The digital universe in 2020: Big data, bigger digital shadows, and biggest growth in the far east. IDC iView: IDC Analyze the future 2007, 1-16.

[12] Garousi, V. a. (2014). highly-cited papers in software engineering: The top-100. information and Software Technology, 71, 108-128.

[13] Garousi, V. a. (2016). Citations, research topics and active countries in software engineering: A bibliometrics study. Computer Science Review, 19, 56-77.

[14] Garrett, M. A. (2014). Big Data analytics and cognitive computing–future opportunities for astronomical research. In IOP conference series: materials science and engineering,, 67, pp. 012-017.

[15] Gingras, Y. a. (2010). Why it has become more difficult to predict Nobel Prize winners: a bibliometric analysis of nominees and winners of the chemistry and physics prizes (1901–2007). Scientometrics, 82(2), 401-412.

[16] Ho, Y.-S. (2012). op-cited articles in chemical engineering in Science Citation Index Expanded: A bibliometric analysis. Chinese Journal of Chemical Engineering, 20(3), 478-488.

[17] Ho, Y.-S. (2014). Classic articles on social work field in Social Science Citation Index: a bibliometric analysis. Scientometrics, 98(1), 137-155.

[18] Hu, H. Y.-S. (2014). Toward scalable systems for big data analytics: A technology tutorial. IEEE access, 2, 652-687.

[19] Imran, S. T. (2021). Big data analytics in healthcare– A systematic literature review and roadmap for practical implementation. IEEE/CAA Journal of Automatica Sinica, 8(1), 1-22.

[20] James, J. (2019). Data Never Sleeps 7.0. Retrieved from https://www.domo.com/learn/data-never- sleeps-7

[21] Kushwaha, A. K. (2020). Language model-driven chatbot for business to address marketing and selection of products. In Re-imagining Diffusion and Adoption of Information Technology and Systems: A Continuing Conversation: IFIP WG 8.6 (pp. 16-28). Springer International.

[22] Laney, D. (2001). 3D data management: Controlling data volume, velocity and variety. META group research note, 6(70), 1.

[23] Parlina, A. K. (2020). Theme mapping and bibliometrics analysis of one decade of big data research in the scopus database. Information, 11(2), 69.

[24] Qiu, J. Q. (2016). A survey of machine learning for big data processing.". EURASIP Journal on Advances in Signal Processing, 1-16.

[25] Reuters, T. (2018). Using Bibliometrics : A guide to evaluating research performance with citation data. Retrieved from www.ips.clarivate.com/m/pdfs/325133 thomson.pdf

[26] Sedaghat, N. M.-X. (2022). Machines learn to infer stellar parameters just by looking at a large

number of spectra. Monthly Notices of the Royal Astronomical Society, 501(4), 6026-6041.

[27] Tosi, D. a. (2021). How schools affected the covid-19 pandemic in Italy: Data analysis for Lombardy region, Campania region, and Emilia region. Future Internet, 13(5), 109.

[28] Tosi, D. R. (2024). 15 years of Big Data: a systematic literature review.". Journal of Big Data, 11(1), 73.

[29] Tyagi, A. K. (2019). Machine learning with big data." In Machine Learning with Big Data. International Conference on Sustainable Computing in Science, Technology and Management (SUSCOM),. Jaipur: Amity University Rajasthan.

[30] Van Rijmenam, M. (2019). M. A Short History of Big Data.

[31] Ward, J. S. (2013). Undefined By Data: A Survey of Big Data Definitions. arXiv preprint arXiv:1309.5821.

[32] Wohlin, C. (2007). An analysis of the most cited articles in software engineering journals-2000. Information and Software Technology, 49(1), 2-11.

[33] Wohlin, C. (2007). An analysis of the most cited articles in software engineering journals-2000. Information and Software Technology, 49(1), 2-11.

[34] Xu, Z. a. (2019). A Bibliometrics analysis on big data research (2009–2018). Journal of Data, Information and Management, 1, 3-15.

[35] Zhou, L. S. (2017). Machine learning on big data: Opportunities and challenges. Neurocomputing, 237, 350-361.