International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Trends and Pattern in Big Data: A Bibliometric Study

Dr. Radha Raghuramapatruni

Submitted: 29/11/2023 **Revised:** 09/01/2024 **Accepted:** 18/01/2024

Abstract:

Purpose –The goal of the present study is to provide bibliometric analysis in order to gain insight into the evolution of the literature in the field of big data science.

Methodology- This study leveraged bibliometric techniques to examine a dataset of 3255 articles published between 2015-2023 using Vosviewer software. The main focus of the study was the performance analysis and the scientific mapping of articles.

Findings- An overview of big data analysis and the patterns in publications over the last eight years were included in this study. It showed that "Big Data" were the most popular keywords, and that countries like the US, China, Thailand, Australia, and India have a lot of publications in this area. The highest number of publications on big data was in China followed by US, India, UK and South Korea. The analysis showed that computer science, Decision Science, and math are the main areas where more research has been done on big data and data science.

Practical implications – Researchers from a variety of fields, especially those studying big data, may find this study to be a useful tool for tracking the development of scientific publications over time in a particular area.

Originality/value – The purpose of this study is to do a detailed analysis of prior research in the field of big data, which has been ongoing for more than 8 years. This analysis should provide some insight into the key developments that have occurred in big data during that period of eight years. The techniques employed result in a more comprehensive and in-depth analysis.

Keywords- *Bibliometric analysis, Big data, Scopus, Thematic map, Citations, Co-occurrence of words*

Introduction:

The term "big data" is commonly used to refer to a large amount of data. Although the concept of "big data" is relatively new, the concept of data storage and analysis has been around for centuries. Prior to the advent of computers and databases, data was kept in paper registers. However, modern computers, the internet, and mobile technology have enabled the generation of vast amounts of data on a daily basis (Rodríguez et al., 2016; Ardagna, Ceravolo & Damiani, 2016; Baig, Shuib & Yadegaridehkordi, 2019). Big Data is generated from different social media sites, online shopping apps, different transactions in books of records, network devices, and educational records (Maroufkhani et al., 2020; Lutfi, 2022). Big data analytics is transforming the ways companies fight with one another. It is providing us with fresh perspectives on the details of data, enabling us to stay competitive, learn more, make wiser decisions, and be more productive. (Maroufkhani et al., 2020). Most of the time, big data is about what employees and customers do in their day-to-day lives, and it's stored in an organization's data structure that gives us actionable information, descriptions, predictions, and prescriptions. It can be challenging to extract meaningful information and data from big data because of its vastness, velocity, and diversity, regardless of its structure. (Ford et al., 2016; Ram et al., 2019). Bibliometrics helped us better understand large data by providing insights into its functioning and the ways in which various themes influence it. (Aparicio et al., 2019). Bibliometric analysis is a statistical and mathematical method of examining research activity. Another name for it is scientometrics. We can build various bibliometric networks to examine various kinds of connections between various analysis units. (Callon et al., 1991). In this case, the "co-occurrence" approach was our primary tool. In essence, it's the joining of two elements in a document. This type of research helps the researcher understand the main points of interest in

Associate Professor, GITAM School of Business, GITAM Deemed to be University, Visakhapatnam, Andhra Pradesh, India

radharaghuramapatruni@gmail.com

ORCID ID-0000-0002-4486-5458

any area of study, and it also gives them a better idea of what's going on in the brain (Cobo, 2012). The goal of this research is to identify possible directions for future investigation and to give academic, scientific, and other stakeholders a thorough grasp of the research interests in statistical, mathematical, psychological biases, and data analysis.

The primary objectives of this study are:

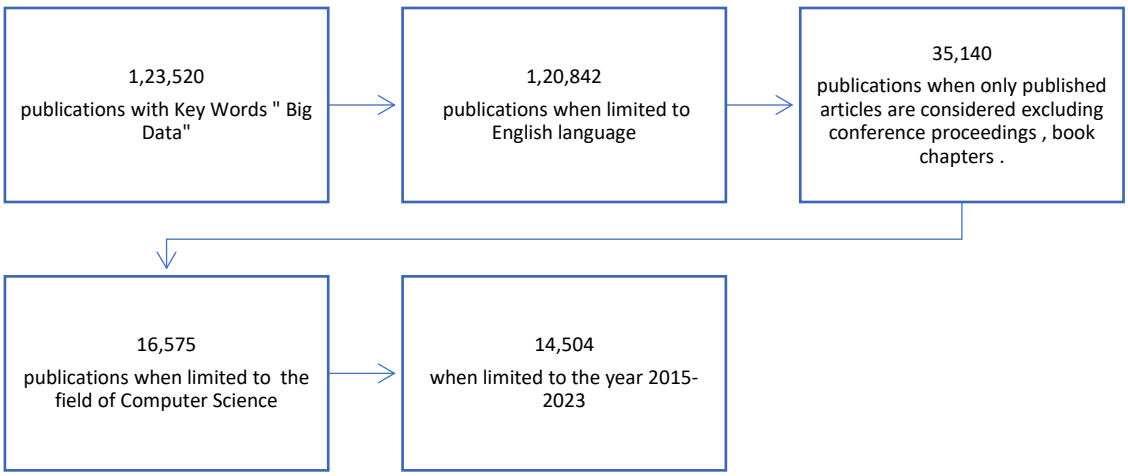
1) This article seeks to explore the latest research articles published in the fields of data research and its trend and patterns.
2) It explores the work of renowned authors, renowned countries, and renowned academic institutions, as well as illustrate the most commonly used terms and relevant research topics.
3) Additionally, it also focuses on the dominant countries and their research contributions, as well as explore the potential for future collaboration and objectives.

The type of research we conducted and our methodology are described in this section. We examined our methodology, our findings, and our limitations in terms of data analysis. We also discussed our research's primary conclusions and limitations. VoSViewers allowed us to generate a wide variety of maps and relationships, which is how we accomplished all of this.

Research Methodology:

In the current study, a bibliometric analysis is carried out to look at the trend in the earlier literature in the field of "Big Data" using the Vosviewer program. Multi-level filtering is utilized to obtain Scopus data in order to do the analysis. The time period considered for the study covers eight years i.e from the year 2015 to the year 2023. A number of network diagrams are prepared to capture the trend and conclude meaningfull insights. In the analysis presented in the next section , we tried to study the growth rate in the number of publications over the years along with identifying the crucial areas or subjects where Big Data is studied. An attempt is also made to study the contribution of each nation in the field of Big Data along with identifying influential authors. Bibliometric data gathered from Scopus database is analysed in different aspects to draw meaning full conclusions.

Fig 1: Total number of publications



Data Analysis And Interpretation

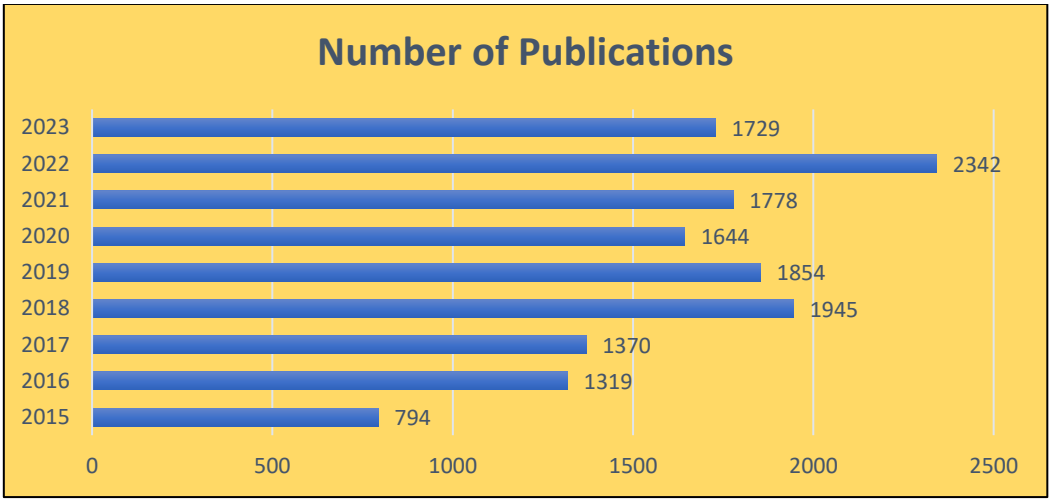
1. Publication Trend Analysis

Examining and evaluating the publication trend in the Big Data space is the main focus of this section. Initially publication on year-on-year basis was analyzed followed by investigating country wise publication. Later, investigation on finding the subject wise publication trend was carried out.

a. Publication Analysis Year On Year

Preceding investigation studies performed in the area of " Big Data " has an upward publication trend from the last decade as presented in figure No. 2 and table no. 1. In the year 2022 , maximum number of research work i.e. 2342 papers were published , followed by subsequent publications i.e. 1778 in the year 2021 and 1644 in the year 2020. Nearly 1729 papers have been published in the first seven month of the year 2023. Henceforward, it is a normally expected to have an upsurge in the publication numbers in the years to come. However, a researcher don't have access to all the publications as only a few papers are avaliable in the open access while others are usually charges a fee.

Fig 2: Year on Year Publication Analysis



(Source: Researchers' Output)

Table No. 1 Year on Year Publication Statistics

YEAR	*Number of Publications*
2015	794
2016	1319
2017	1370
2018	1945
2019	1854
2020	1644
2021	1778
2022	2342
2023	1729

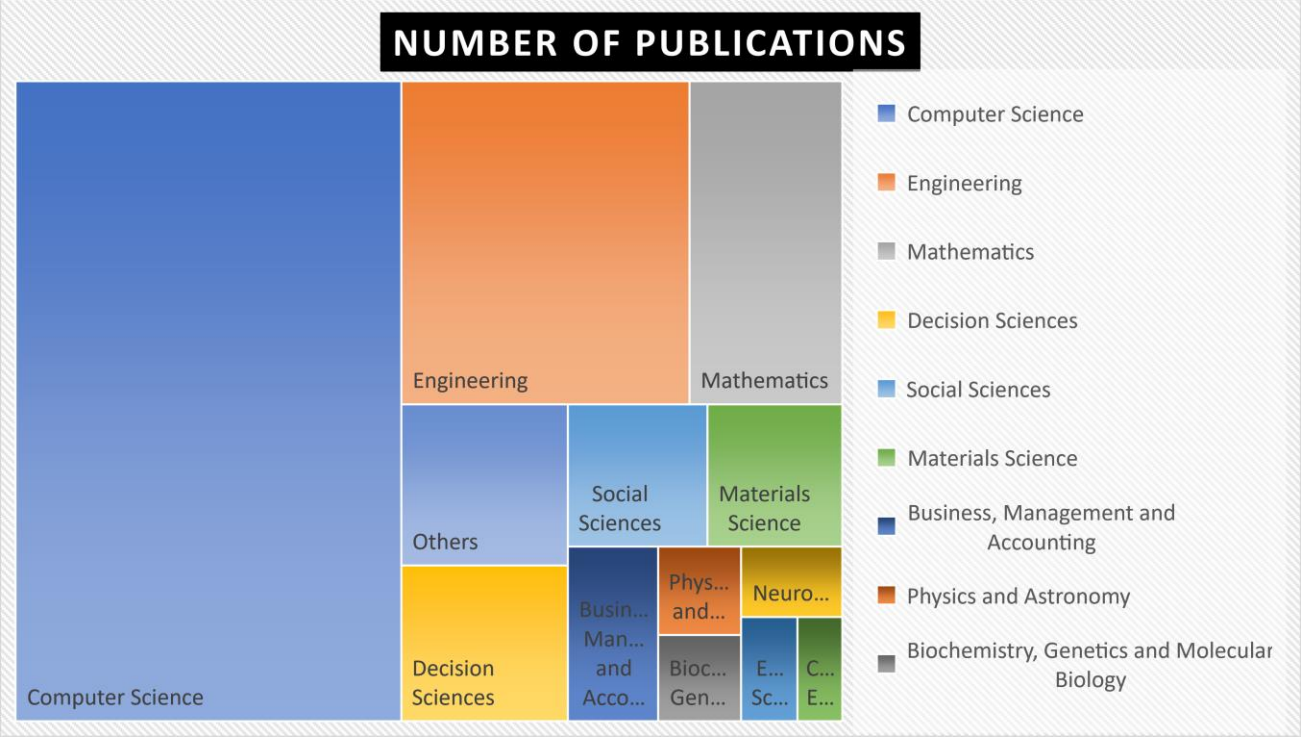
(Source: Researchers' Output)

b. Subject Wise Publications

Analysis is conducted in this part to investigate a variety of topics related to the study of "Big Data." Highest number of publications are undertaken in the area of Computer Science domain (14775) , followed by engineering subject under which 5571

papers are published, followed by mathematics under which approximately 2949 papers are published followed by others. A thorough examination is displayed in the table No. 1

Fig. 3 Subject Wise Analysis



(Source: Researchers' Output)

Table 2 Subject Wise Publication Trend

SUBJECT AREA	Number of Publications
Computer Science	14775
Engineering	5571
Mathematics	2949
Decision Sciences	1552
Social Sciences	1189
Materials Science	1147
Business, Management and Accounting	940
Physics and Astronomy	439
Biochemistry, Genetics and Molecular Biology	428
Neuroscience	425
Environmental Science	351
Chemical Engineering	278
Others	1607

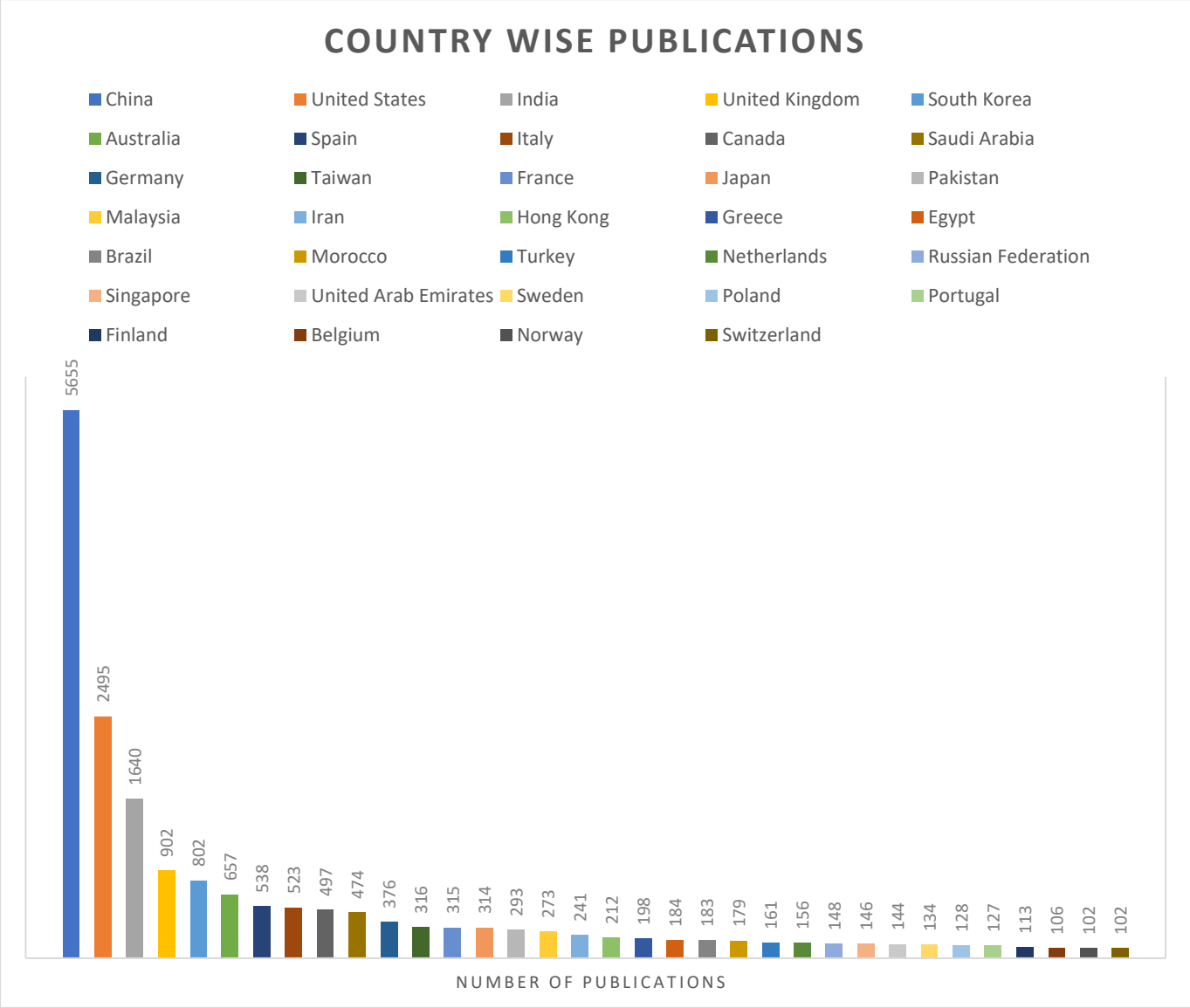
(Source: Researchers' Output)

c. Country Wise Publications

To find the most influential nation in the "Big Data" space, a country-by-country publishing analysis has been conducted in this section. The detailed country wise statistics is presented in the table No. 2. Country "China" stood at first position that has maximum number of publications (5655) to its credit, followed by 2495 by United States.

The network map created by Vos viewer software also emphasized the significant same results. Fig No. 4 represents the country wise publication trend.

Fig 4 Country Wise Publication Trend



(Source: Researcher's Output)

2. Keyword Analysis

By concentrating on the analysis of the author's keywords, this section of the research allowed the assessment of the title-associated essence of statistics. Measurements of comprehension excellence and the paper's relative content are required for such an inquiry. (Strozzi et al., 2017). The current research examines the publication trend in the field of "big data" with that goal. To inspect and analyse the trend of keywords , all 47343 keywords are analysed from a pool of 14,504 papers. Out of 47373 keywords only 4498 keywords are selected to study the trend as these keywords fullfill the criteria chosen i.e. minimum of five repitition. A network map is created with the help of Vosviewer software.

The analysis displayed " Big Data " as the greatest used keyword in all the papers with a total occurrence of 9839 times along with 66937 as the total link strength " D, followed by " Data Mining" with 1290 as frequency and 11468 as total link strength.This is followed by "Data Handling" with 1055 as frequency along with 9346 as total link strength. The word " Machine Learning " has the seocnd maximum total link strength. The keywords are comprehensively demonstrated in Table No. 3,

which also includes information on the keywords' total link strength and frequency of occurrence. The network diagram prepared with the vosvierwer software featured " Big Data " as the highest occuring word as presented in Fig no. 5

Fig 5 Keyword Analyis

(Source: Researcher's Output)

Table No. 3 Keyword Statistics

KEYWORD	OCCURRENCES	TOTAL LINK STRENGTH
big data	9839	66937
data mining	1290	11468
data handling	1055	9346
machine learning	1029	10171
data analytics	933	8980
internet of things	889	8460
digital storage	842	7806
information management	795	7362
learning systems	773	8406
deep learning	688	6999
artificial intelligence	687	6167
cloud computing	644	5896
decision making	575	5142
classification (of information)	557	6149
forecasting	522	5240

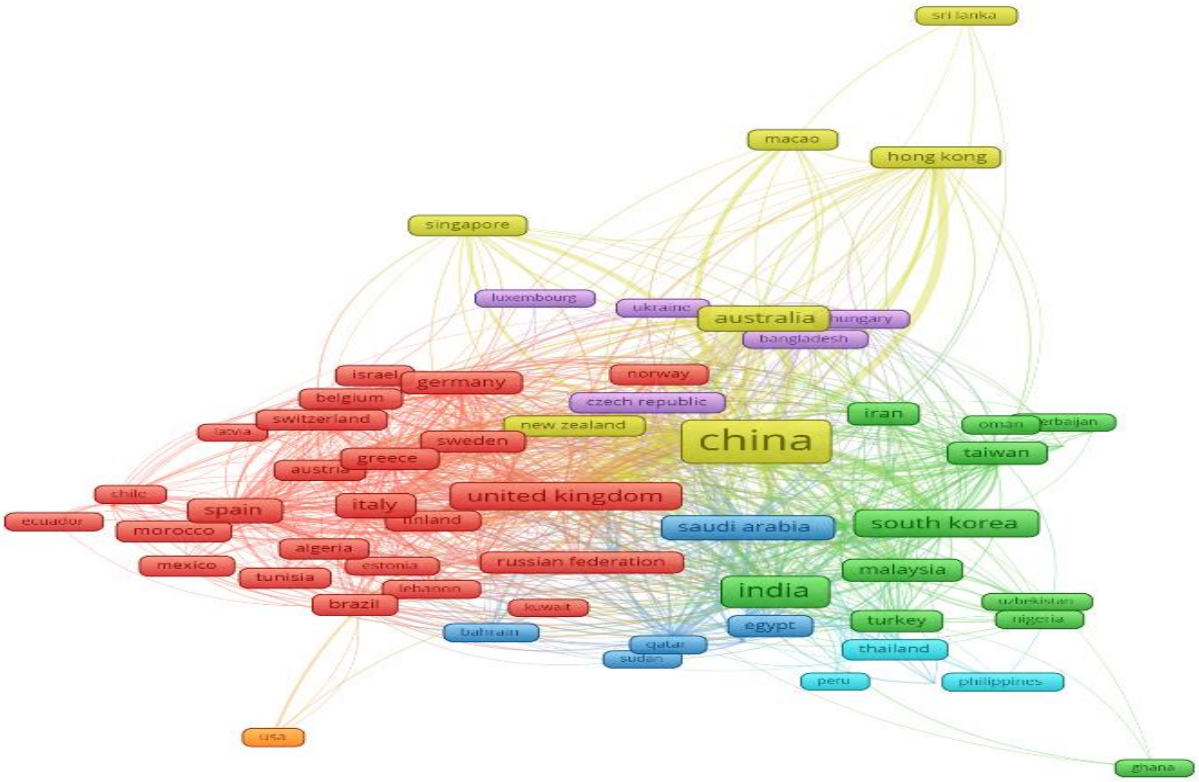
learning algorithms	453	5071
article	438	5967
human	427	6025
clustering algorithms	412	3733
social networking (online)	402	3604
big data analytics	398	3893
algorithms	358	4120
data technologies	336	2526
students	326	2767
advanced analytics	317	2960

(Source: Researcher's Output)

Co-Occurrence Analysis: **Country Trend**

Every one of the 197 nations has been taken into consideration when completing co-authorship for countries investigation. A network map is constructed, showing every nation where notable publications in the "Big Data" subject have been made. Within VOSviewer, the interconnectedness amid any two countries will be computed by measuring the distance amongst them . Smaller the distance , higher the connection. The analysis has bifurcated all 197 nations into seven (07) clusters whereas there are eighty seven countries who have published more than five (05) papers in the area of " Big Data". The most influential country in terms of publications seems to be China , followed by United Kingdom , Russian Federation , South Korea , India amongst others. Country wise network map is presented in Fig No. 6

Fig 6 Country Wise Network Map



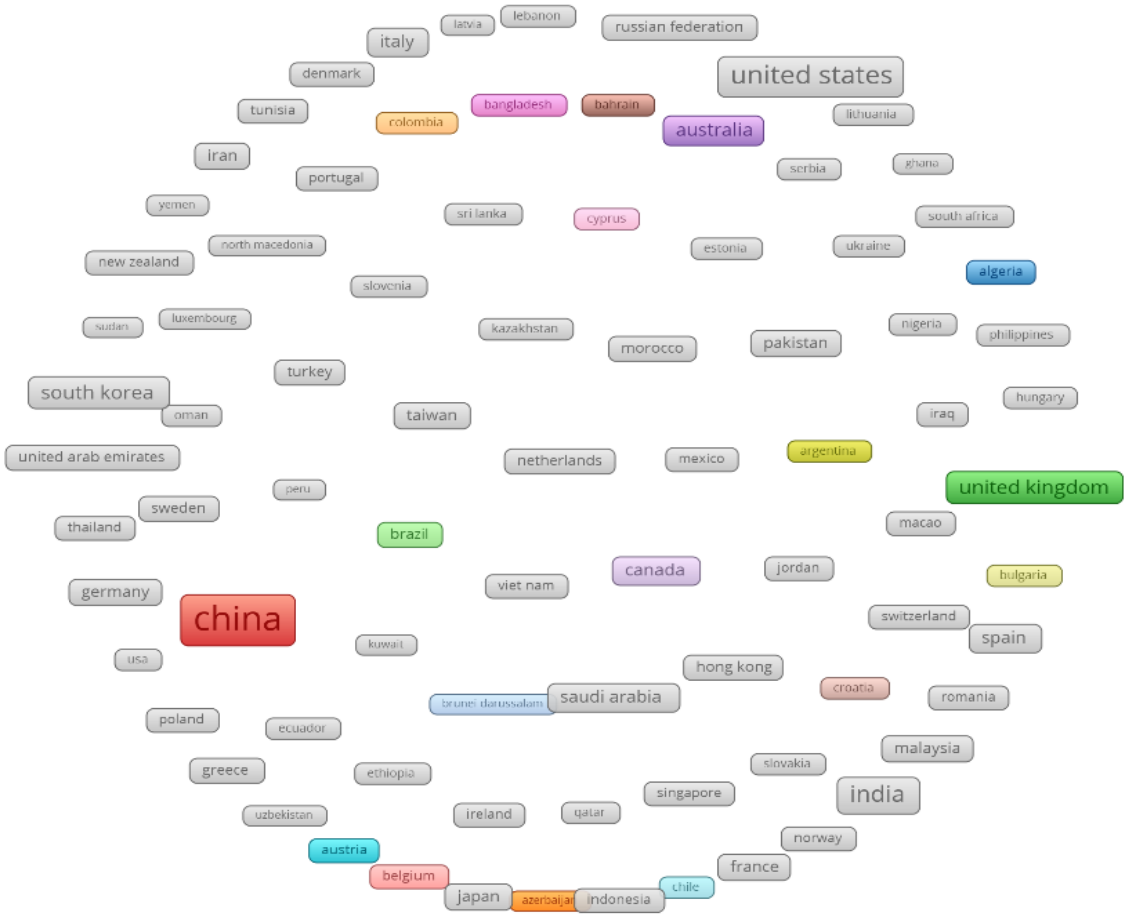
(Source: Researcher's Output)

Citation Analysis

Country Wise

Investigating country-specific citation analysis is the main goal of this section. In order to complete the analysis, citations from all 197 nations are taken into account. A network map is created for country wise analyzed which displays China holds maximum citations in the domain of "Big Data," followed by United Kingdom, United States and Russian Federation amongst others. The network map for countries is presented in Fig No. 7

Fig 7 Country Wise Citation Analysis



(Source: Researcher's Output)

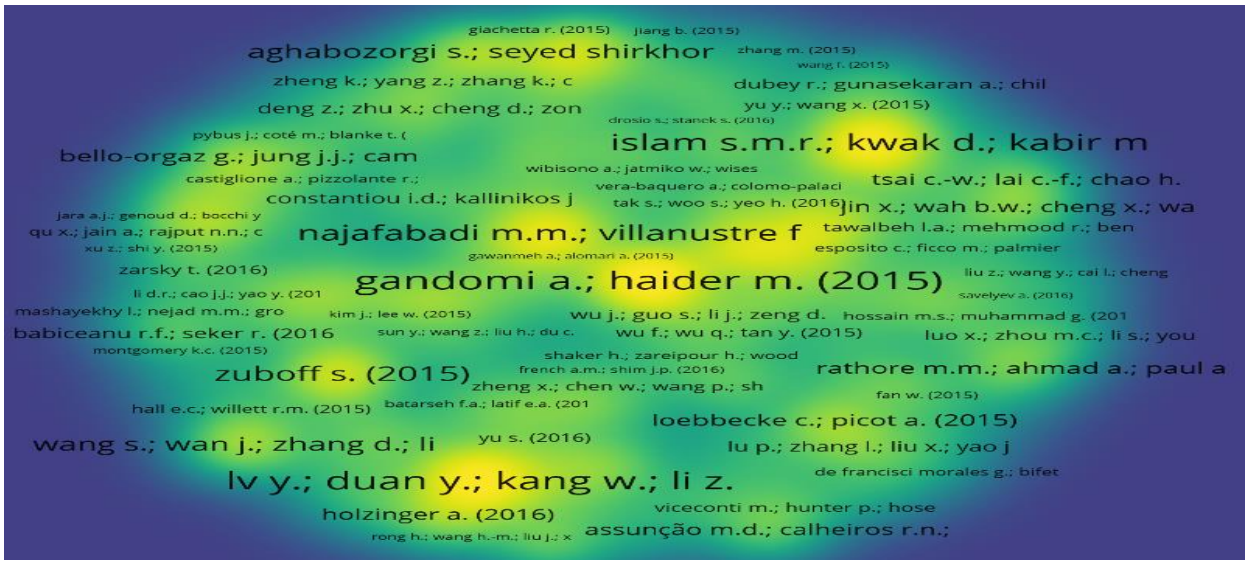
Documents

Citation Analysis

Document-wise citation is explained in more detail in the section that follows. All 12288 documents were examined and studied in order to complete the analysis, with the requirement that each document contain at least 10 citations. Out of all the documents, only 3255 documents full-filled the condition of having at least 10 citations to its credit. The network map for documents having minimum 10 citations have been prepared and presented in Fig No.8 Through the network diagram, document authored by Haider m in the year 2015 have the highest citations, followed by Kwak D, Duan Y and Kang; Li amongst others

.

Fig 8: Documents Citations Analysis



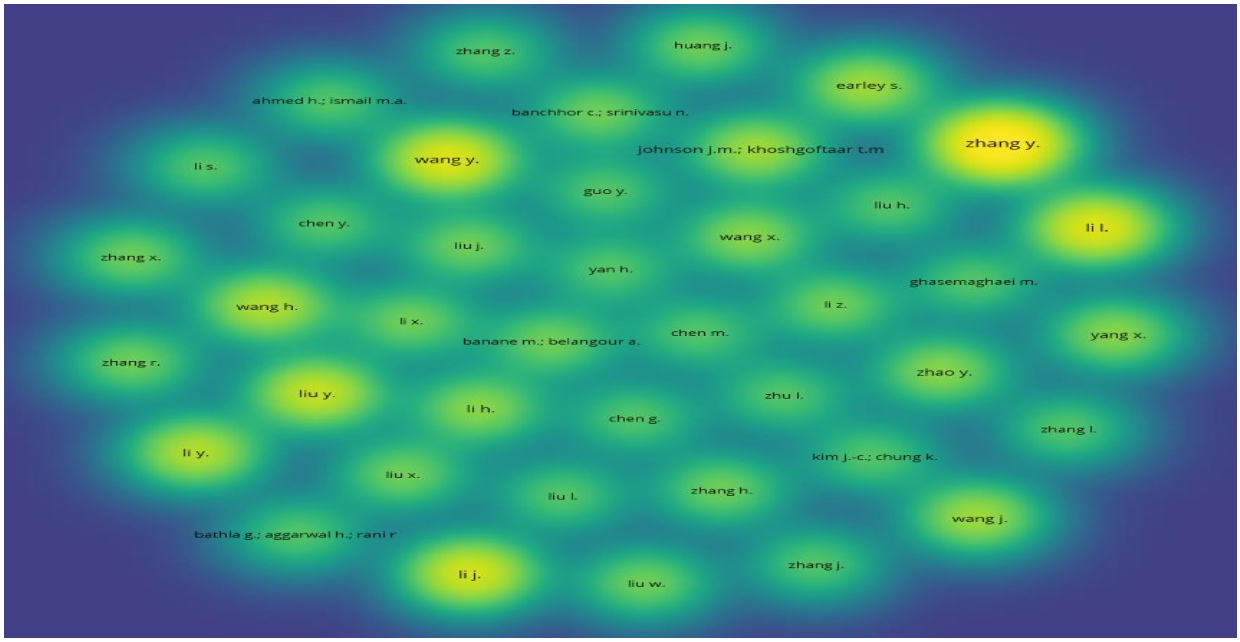
(Source; Researcher's Output)

Author

Density

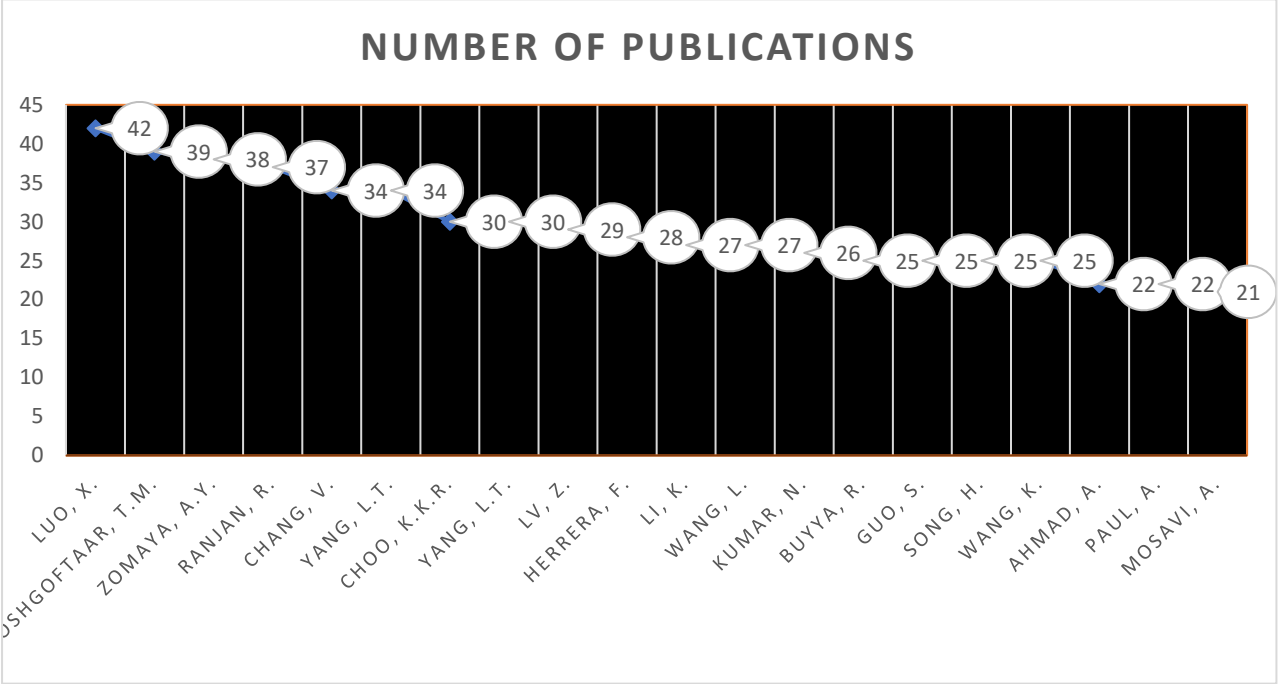
An analysis is carried out in this section of the paper while taking the publication's writers into consideration. The most important writers are shown in Figure No.9, along with the total number of publications each author has contributed. Fig No.10 showcase the author wise publication like Luo, X. has forty-two publications to his credit followed by Khoshgoftaa, T.M. who had thirty-nine publications to his credit whereas Zomaya, A.Y., Ranjan, R have thirty-eight and thirty-seven publications to their credit respectively. Table No. 4 presents additional data illustrating author-by-author publication. The young scholars in the "Big Data" discipline will find this analysis helpful in learning about important writers related to their profession.

Fig.9: Network map of Author's Density



(Source: Researcher's Output)

Fig 10: Author Wise Publication Statistics



(Source: Researcher's Output)

Table No. 4 Author Wise Publication Statistics

AUTHOR NAME	Number of Publications
Luo, X.	42
Khoshgoftaar, T.M.	39
Zomaya, A.Y.	38
Ranjan, R.	37
Chang, V.	34
Yang, L.T.	34
Choo, K.K.R.	30
Yang, L.T.	30
Lv, Z.	29
Herrera, F.	28
Li, K.	27
Wang, L.	27
Kumar, N.	26
Buyya, R.	25
Guo, S.	25
Song, H.	25
Wang, K.	25
Ahmad, A.	22

Paul, A.	22
Mosavi, A.	21

Conclusion:

The big data analysis trends and publications over the last eight years were briefly reviewed in this study. It showed that "Big Data" were the most popular keywords, and that countries like the US, China, Thailand, Australia, and India have a lot of publications in this area. Big data as a keyword has been used 66937 times in last few years. It also compared India to other countries like the US and China, as well as other countries like Canada and Saudi Arabia. The highest number of publications on big data was in China followed by US, India, UK and South Korea. The analysis showed that computer science, Decision Science, and math are the main areas where more research has been done on big data and data science. Future researchers will find this study helpful as it provides a concise summary of previous research.

Limitations And Future Scope

According to the research methods section, multi-level filtration has been done, thus the current study did not incorporate all of the previously published papers and material that is now available in the field of "Big Data." Only Scopus data base was considered for the analysis which serves as a major limitation. The existing study has measured and presented the statistics regarding publication trend for the years 2015 to 2023 only. Therefore, in the future, other database for extracting the publication statistics can be considered. Time horizon of the study can be enlarged to conclude more meaningful insights. Since Vosviewer is being used for this investigation, different software might be taken into consideration for later research.

Acknowledgements

Funding

This research did not receive any specific grant from funding agencies in the public.

Data Availability Statement

The database generated and /or analysed during the current study are not publicly available due to privacy, but are available from the corresponding author on reasonable request.

Declarations

Author declares that all works are original and this manuscript has not been published in any other journal.

References

[1] Aparicio, G., Iturralde, T., & Maseda, A. (2019). Conceptual structure and perspectives on entrepreneurship education research: A bibliometric review. *European research on management and business economics*, 25(3),105-113.

[2] Baig, M.I., Shuib, L., & Yadegaridehkordi, E. (2019). Big data adoption: State of the art and research challenges. *Information Processing & Management*, 56(6),102095.

[3] Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemsitry. *Scientometrics*, 22,155-205.

[4] Cobo Martín, M.J. (2012). SciMAT: herramienta software para el análisis de la evolución del conocimiento científico. Propuesta de una metodología de evaluación. *Granada: Universidad de Granada*.

[5] Ford, J.D., Tilleard, S. E., Berrang-Ford, L., Araos, M., Biesbroek, R., Lesnikowski, A. C., & Bizikova, L. (2016). Big data has big potential for applications to climate change adaptation. *Proceedings of the National Academy of Sciences*, 113(39),10729-10732.

[6] Lutfi, A., Alsyouf, A., Almaiah, M. A., Alrawad, M., Abdo, A. A. K., Al-Khasawneh, A. L., & Saad, M. (2022). Factors influencing the adoption of big data analytics in the digital transformation era: Case study of Jordanian SMEs. *Sustainability*, 14(3), 1802.

[7] Maroufkhani, P., Tseng, M. L., Iranmanesh, M., Ismail, W. K. W., & Khalid, H. (2020). Big data analytics adoption: Determinants and performances among small to medium-sized enterprises. *International journal of information management*, 54,102190.

[8] Ram, J., Afridi, N. K., & Khan, K. A. (2019). Adoption of Big Data analytics in construction: Development of a conceptual

model. *Built Environment Project and Asset Management*, *9*(4),564-579.

[9] Strozzi F., Colicchia C., Creazza A., Noè C. (2017). Literature review on the 'Smart Factory'concept using bibliometric tools. *International Journal of Production Research*, 55(22),6572-6591.

[10] Rodríguez-Mazahua, L., Rodríguez-Enríquez, C. A., Sánchez-Cervantes, J. L., Cervantes, J., García-Alcaraz, J. L., & Alor-Hernández, G. (2016). A general perspective of Big Data: applications, tools, challenges and trends. *The Journal of Supercomputing*, 72,3073-3113.

[11] Ardagna, C. A., Ceravolo, P., & Damiani, E. (2016). Big data analytics as-a-service: Issues and challenges. In *2016 IEEE international conference on big data (big data)* (pp. 3638-3644). IEEE.