Original Article

"A Decade Of Big Data Research In Indonesia. A Bibliometric Analysis"

Mr. Abhijit Vhatkar[1], Dr. Vilis Pawar [2*], Mr. Soumyakant Dash[3]

[1]Assistant Professor, Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth, Pune (Orcid Id: 0000-0001-5308-463x)
[2*]Assistant Professor, Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth, Pune (Orcid Id: 0000-0003-3455-5349)
[3]Assistant Professor, Global Business School and Research Centre, Dr. D. Y. Patil Vidyapeeth, Pune (Orcid Id: 0000-0002-9297-7710)

***Corresponding Author: -** Dr. Vilis Pawar

*Assistant Professor, Global Business School And Research Centre, Dr. D. Y. Patil Vidyapeeth, Pune (Orcid Id: 0000-0003-3455-5349)
DOI: 10.47750/pnr.2023.14.S01.69

Abstract

A bibliometric examination of Big Data research is presented here. This study is conducted on research output from Indonesia. This bibliometric analysis is used to determine the status of current Big Data research as well as its future course. This investigation focuses on Big Data research that was done in Indonesia between 2012 and 2021. 1538 research papers were published during this decade. Data from these research papers were analyzed with the help of MS Excel & VOSviewer software. Steady growth is observed in the research publication output throughout the decade. The top ten cited research articles have been identified. The primary focus of this study is on seven clusters, which are comparable to seven Big Data research areas. These clusters provide Big Data research with a path for the future.

Keywords: Big Data, Bibliometric analysis of Big Data, Research trends in Big Data, Big data research, Vosviewer.

1.INTRODUCTION

Bibliometric evaluation offers a path to dealing with scholarly articles available in various databases (Nasir et al., 2020). The raw data is extracted from databases and analyzed with analytical software (Narin et al., 2020; Firdaus et al., 2019) and meaningful insights are derived from it.

Bibliometric analysis is a scientific method of research (Sanchez et al., 2017). It is not a new terminology. It is there for the last few decades. The availability of various data analysis and visualization tools has resulted in a higher number of bibliometric studies (Donthu et al., 2021).

VOSviewer is among the topmost popular data visualization tools used for this study. It is used for data processing and data visualization. VOSviewer software gives insights into the data and highlights research themes, focused keywords, the current status of the research, the future trend of the research, etc. (Donthu et al., 2021). VOSviewer is used by various researchers for bibliometric analysis on various topics like materials research (Nandiyanto et al., 2021), blockchain technology (Kuzior and Sira, 2022), management bioenergy (Soegoto et al., 2022), Covid-19 (Hamidah et al., 2020), techno-economic education (Ragadhita and Nandiyanto, 2022), machine learning (Oyewola and Dada, 2022), vocational schools (Al Husaeni and Nandiyanto, 2023), social sustainability (Contreras and Abid, 2022), engineering research (Nandiyanto and Al Husaeni, 2022), Briquette Research (Al Husaeni, 2022), low carbon education (Hudha et al., 2020), tourism (Koseoglu et al., 2016), safety culture (Nunen et al., 2018), etc.

Data is the information present in the form of quantity, characters, and symbols. It can be stored or transmitted through media. Big Data is considered as a data with vast variety, huge volume, and major velocity (Villars et al., 2011). It is data in a very large volume (Kitchin and McArdle, 2016). This data is collected from various sources. The nature of Big Data is so vast that the normal data processing tools become insignificant for Big Data processing (Mauro et al., 2015).

Increased use of social media, e-commerce, gaming, websites, etc. resulted in huge data generation. This data is collected, segregated, stored, and analyzed to understand more about customers. Big Data is beneficial in analyzing customer sentiments, the behavior of the customer, customer segmentation, predicting the future, customer retention, personalization, support functions, etc. (Bizer et al., 2012). It shows the importance of Big Data (Pence, 2014).

This study purposes to investigate various studies done on Big Data. It provides a future direction for Big Data research. The study aims to give a clear picture of Big Data research to researchers. This bibliometric study will help decide future research areas in Big Data.

2. METHOD

Structured data is a base for any bibliometric study. The researchers have used the Scopus database for data collection in this study. Scopus database provides us with the data of all research articles which are indexed in Scopus. Research papers indexed by Scopus are generally considered to be of high quality. Hence, the data collected from Scopus is the data of high-quality research.

The Scopus database was searched for the term 'Big Data' on November 22, 2022. The search was done for 'Titles, Keywords, and Abstracts'. More filters like subject area, publication stage, country, year, etc. were applied. Computer science or engineering is considered a subject area under the Scopus database. Final stage research papers are only considered for this study. Research under Indonesia country is selected. Research papers published between the year 2012 to 2021 are considered for this study. The researchers finally received data from 1538 research papers.

The data was downloaded in the form of a 'CSV' file. Data analysis was done with the help of MS Excel software and data visualization was done with the help of VOSviewer software.

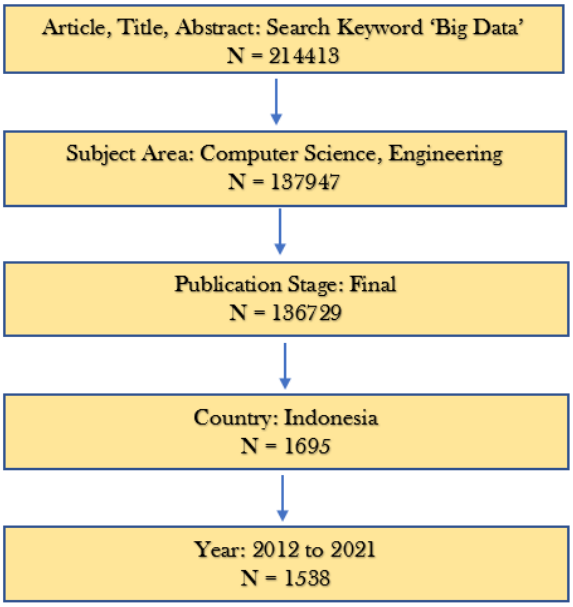

Figure 1. Search Flow for Data Extraction

3. RESULTS & DISCUSSION

Figure 2. indicates the growth of big data research in Indonesia. The research on Big Data has grown steadily and significantly. 8 research papers were published in 2012, 17 papers were published in 2013, 32 papers were published in 2014, 81 papers were published in 2015, 109 papers were published in 2016, 142 papers were published in 2017, 265 papers were published in 2018, 373 papers published in 2019, 301 papers published in 2020, 210 papers published in 2021. These numbers indicate that the research output shows a steady increase from 2012 to 2019. The research output started decreasing after 2019.

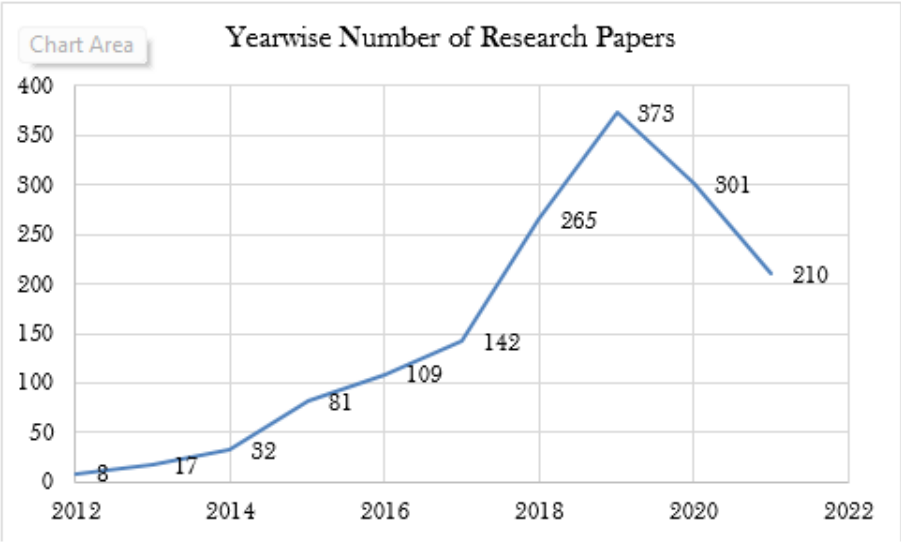

Figure 2. Year-wise research papers on Big Data

Table 1. Top 10 cited articles on Big Data

Sr.	Title	Citations	Source	Year
1	"Big data analytics in supply chain management between 2010 and 2016: Insights to industries". (Tiwari et al., 2018)	286	Computers and Industrial Engineering	2018
2	"Big data emerging technology: Insights into innovative environment for online learning resources". (Huda et al., 2018)	165	International Journal of Emerging Technologies in Learning	2018
3	"Big data clustering: A review". (Shirkhorshidi et al., 2014)	164	Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)	2014
4	"Understanding Modern Learning Environment (MLE) in big data era". (Huda et al., 2018)	96	International Journal of Emerging Technologies in Learning	2018
5	"Convergent communication, sensing and localization in 6g systems: An overview of technologies, opportunities and challenges". (De Lima et al., 2021)	91	IEEE Access	2021
6	"Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis". (Budiharto and Meiliana, 2018).	80	Journal of Big Data	2018
7	"Factors influencing effective use of big data: A research framework". (Surbakti et al., 2020)	67	Information and Management	2020
8	"Traffic big data prediction and visualization using Fast Incremental Model Trees-Drift Detection (FIMT-DD)". (Wibisono et al., 2016)	63	Knowledge-Based Systems	2016
9	"Data Warehouse with Big Data Technology for Higher Education". (Santoso, 2017)	59	Procedia Computer Science	2017
10	"Detection and classification of vehicles for traffic video analytics". (Arinaldi et al., 2018)	54	Procedia Computer Science	2018
	"Cloud incident handling and forensic-by-design: cloud storage as a case study". (Ab Rahman et al., 2017)	54	Concurrency and Computation: Practice and Experience	2017

Table 1. Shows the most cited research papers on big data. The topmost research paper is cited 286 times. Two papers are in the tenth position with 54 citations each.

The keyword data was analyzed with the help of VOSviewer. Minimum occurrences of the keyword are considered as ten. 186 keywords met the criteria and were considered for the analysis.

Figure 3. shows the linking between various keywords. The keyword data is further divided into seven clusters. It is shown with different colours. **Figure 4.** shows the year-wise trend of research output. It highlights the year-wise research from 2012 to 2021. **Figure 5.** Indicates keyword density. The area shown in yellow colour shows the density of the keyword. The dense keyword is the keyword used maximum times and can be considered popular among researchers.

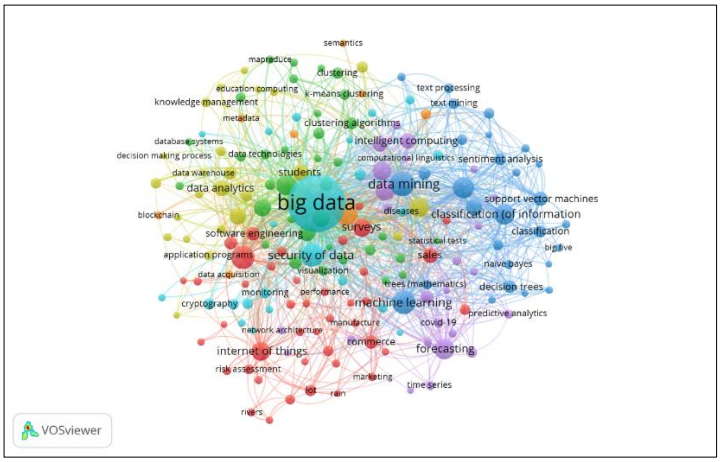

Figure 3. Network visualization for Big Data

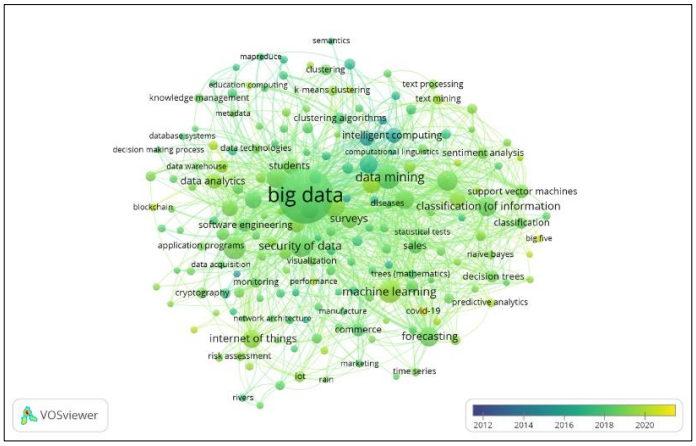

Figure 4. Big Data Overlay visualization

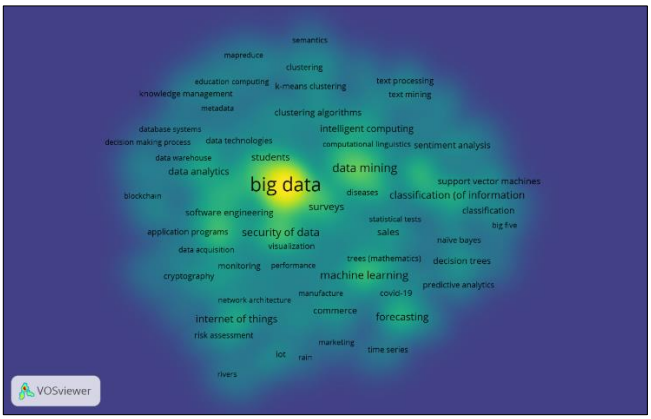

Figure 5. Big Data Density Visualization

Data has indicated 7 clusters in the Big Data topic. This includes Cluster 1. of 42 items includes big data applications, cloud computing, commerce, competition, controllers, costs, customer satisfaction, data communication systems, data transfer, internet, disasters, economics, electronic commerce, energy utilization, engineering research, floods, industrial resolutions, industry 4.0, information management, information technology, internet of things, investments, IoT, literature reviews, maintenance, manufacture, marketing, mobile applications, performance, planning, quality control, rain, regression analysis, internet of things (IoT), risk assessment, risk management, rivers, sales, smart city, software engineering, surveys, sustainable development.

Cluster 2 of 35 items includes agriculture, big data analysis, cluster analysis, clustering, clustering algorithms, data handling, data technologies, data visualization, database systems, digital storage, forestry, geographic information, gis, Hadoop, information analysis, information systems, information use, k-means, k-means clustering, location, map-reduce, mapping, mapreduce, population statistics, query processing, remote sensing, rural areas, search engines, smartphones, statistical tests, telecommunication services, traffic congestion, visualization, web services, websites.

Cluster 3 of 28 items includes big five, classification, learning algorithms, classification (of information), classification methods, classification models, classifiers, data mining, support vector machine, decision tree, decision trees, sentiment analysis, feature extraction, svm, feature selection, machine learning, naïve bayes, nearest neighbour search, personality traits, random forests, social media, support vector machines, natural language processing systems, text mining, text processing, trees (mathematics), social networking (online), Twitter.

Cluster 4 of 25 items consists of advanced analytics, application programs, artificial intelligence, big data analytics, computer aided instruction, computer software, data analytics, data warehouse, data warehouses, decision making, e-learning, education, education computing, engineering education, higher education, information and communication technologies, knowledge based systems, knowledge management, decision making process, learning systems, pattern recognition, performance analysis, decision support systems, students, systematic literature review.
Cluster 5 of 25 items consists of algorithms, artificial neural network, back propagation, computational linguistics, covid-19, crime, data science, deep learning, errors, finance, forecasting, genetic algorithms, intelligent computing, intelligent systems, long short-term memory, mean square error, mobile telecommunication systems, network architecture, neural networks, optimization, prediction, predictive analytics, soft computing, time series, time series analysis.

Cluster 6 of 26 items consists of behavioural research, big data, carbon dioxide, computation theory, convolutional neural network, cryptography, diagnosis, diseases, economic and social effects, extraction, health care, image enhancement, image processing, image segmentation, information security, medical computing, monitoring, network security, object detection, quality of service, security of data, sensor nodes, wireless sensor networks.
Cluster 7 of 8 items includes blockchain, data acquisition, data collection, Indonesia, Jakarta, metadata, semantics, statistics.

Table 2. indicates the total strengths and occurrence of the Big Data.

Figure 6. shows a visualization of the keyword network for cluster 1. Information management is at the centre of the cluster which was commonly used in research with other keywords. Information management is connected to knowledge management, systematic literature review, big data analytics, data analytics, advance analytics, data acquisition, engineering education, wireless sensor networks, internet of things, risk assessment, forecasting, machine learning, classification of information, data mining, big data, information use, data handling, students, mapping, etc.

In cluster 2, data handling was at the centre and connected to other relevant keywords. **Figure 7.** Shows that data handling was mainly connected with big data, the internet of things, data analytics, wireless sensor networks, information use, security data, websites, trees (mathematics), machine learning, forecasting, decision trees, classification of information,

data mining, sentiment analysis, cloud computing, Twitter, text mining, statistics, mapping, big data analysis, Hadoop, etc.

Data mining was at the centre of cluster 3. **Figure 8.** Shows that data mining is connected with internet of things, knowledge management, metadata, big data analytics, advanced analytics, data analytics, decision making, big data applications, security of data, big data, commerce, machine learning forecasting, trees (mathematics), decision trees, sales, nearest neighbour search, classification, classification of information, systematic literature review, sentiment analysis, intelligent computing, Twitter, text mining, text processing, statics, clustering algorithms, k-means clustering, clustering, mapreduce, etc.

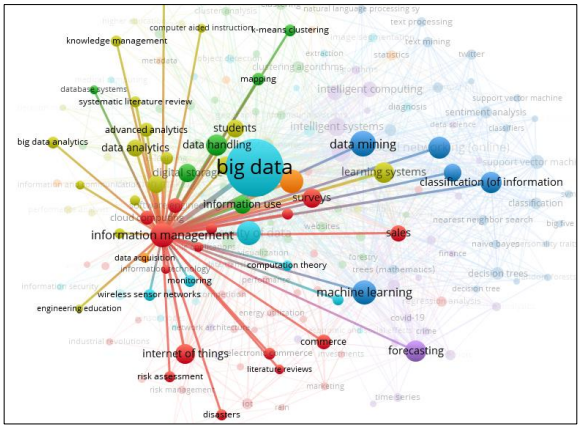

Figure 6. Network visualization for cluster 1

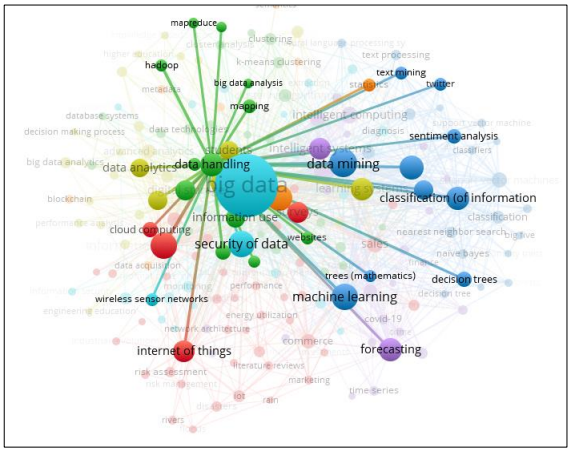

Figure 7. Network visualization for cluster 2

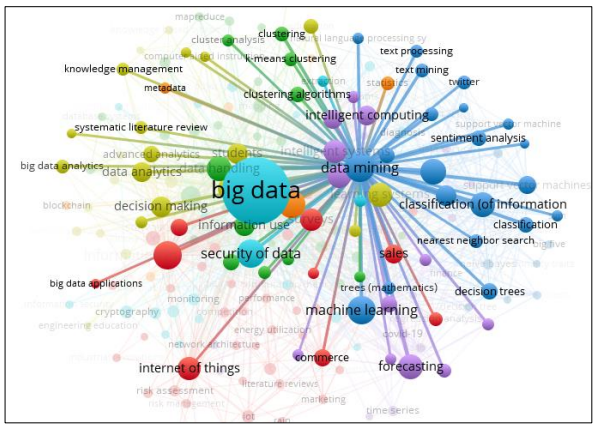

Figure 8. Network visualization for cluster 3

Figure 9. has a visualization of the keyword network for cluster 4. It has learning systems at the centre of the cluster. Learning systems are connected with computer aided instruction, data analytics, big data, information use, security of data, engineering education, risk assessment, machine learning, covid-19, forecasting, sales, decision trees, nearest neighbour search, classification, support vector machine, text mining, text processing, education, computer aided instruction, etc.

Figure 10. shows network visualization of cluster 5. It has forecasting at the centre of the cluster. Forecasting is connected with students, data handling, big data, decision making, security of data, risk assessment, disasters, rain, commerce, time series, decision trees, classification, classification of information, sentiment analysis, statistical tests, sales, trees (mathematics), machine learning, data mining, etc. Big data was at the centre of cluster 6. **Figure 11.** Shows the huge connections of big data with other keywords. Some of the keywords connected to big data are Sentiments, education, k-means clustering, clustering algorithms, intelligent computing, text processing, twitter, diagnosis, sentiment analysis, classifiers, data mining, classification of information, nearest neighbour search, sales, naïve bayes, decision trees, trees (mathematics), machine learning, covid-19, forecasting, time series, websites, commerce, marketing, literature reviews, security of data, information use, IoT, internet of things, risk assessment, rivers, network architecture, cryptography, engineering education, data acquisition, blockchain, big data analytics, data analytics, data technologies, knowledge management, metadata, mapreduce, semantics, etc.

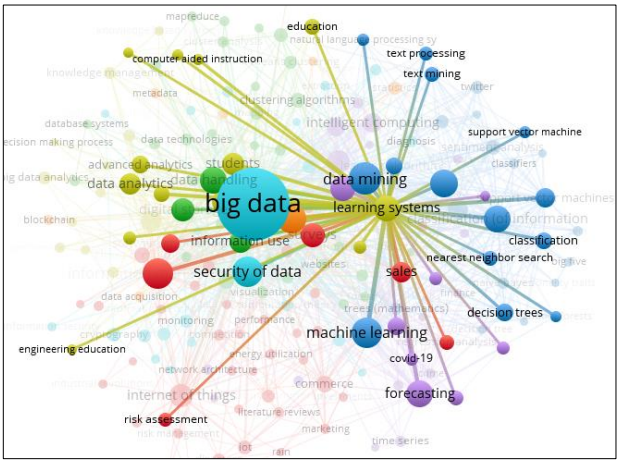

Figure 9. Network visualization of cluster 4

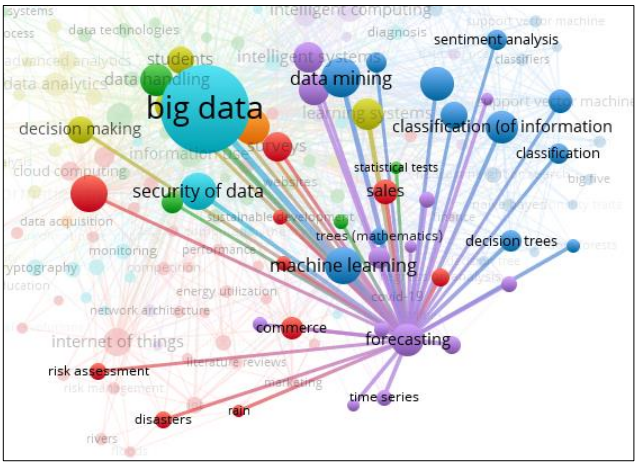

Figure 10. Visualization of network for cluster 5

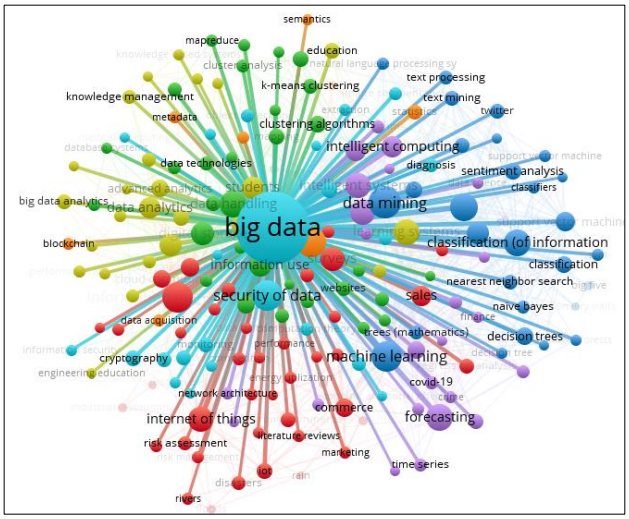

Figure 11. Visualization of network for cluster 6

Figure 12. shows network visualization of cluster 7. It has Indonesia at the centre of the cluster. Indonesia is connected with other keywords. Jakarta, mapping, big data analytics, data analytics, blockchain, data acquisition, security of data, risk assessment, time series, forecasting, covid-19, machine learning, sales, classification of information, sentiment analysis, data mining, twitter, statistics, intelligent computing, mapping, etc.

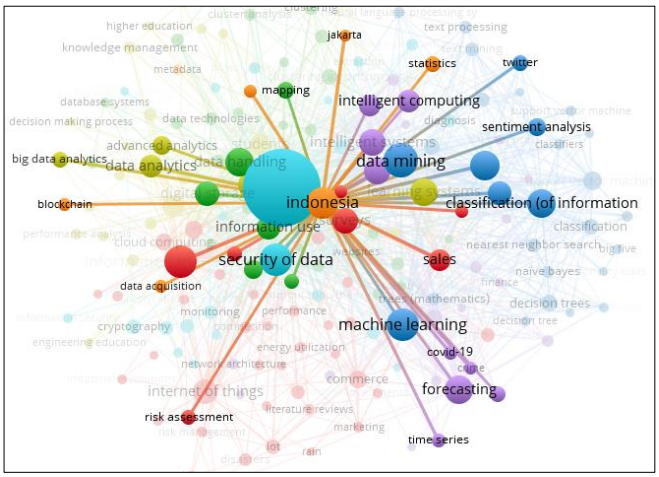

Figure 12. Network visualization of cluster 7

Table 2. Total Strength & Occurrence in Clusters for Big Data

Cluster	Total Strength	Occurrence
1	442	101
2	375	79
3	615	114
4	443	71
5	407	81
6	2379	621
7	334	97

The clusters shown above give information about connections among various keywords. This cluster analysis gives ideas and directions for further research. Big data research has potential with other main keywords like machine learning, data mining, security of data, classification of information, learning systems, information management, forecasting, social networking (online), data handling, Indonesia, etc.

CONCLUSION:-

Research in the Big Data domain is increasing steadily. This bibliographic study gives an idea about past and present research trends in Big Data. The research is focused on Indonesia as a country and it can be considered a representation of Indonesian research output on Big Data. Research on Big Data has steadily increased in this decade but it has seen a fall in the last two years. VOSviewer analysis has given seven important clusters. These clusters focus on seven key areas around which the research of Big Data revolves. Information management, data mining, data handling, learning system, forecasting, big data, and Indonesia are the important keywords among the seven clusters. Each key term with its connections can be used for deciding the future ideas of research in Big Data.

Author's Note:-

The authors of this research declare no direct or indirect conflict of interest in this research. The authors confirm that this research does not have any plagiarism.

REFERENCES

1. Ab Rahman, N. H., Cahyani, N. D. W., & Choo, K. K. R. (2017). Cloud incident handling and forensic-by-design: cloud storage as a case study. Concurrency and Computation: Practice and Experience, 29(14), e3868.
2. Al Husaeni, D. N. (2022). Bibliometric Analysis of Briquette Research Trends During the Covid-19 Pandemic. ASEAN Journal for Science and Engineering in Materials, 1(2), 99-106.
3. Al Husaeni, D. N., & Nandiyanto, A. B. D. A bibliometric analysis of vocational school keywords using vosviewer. ASEAN Journal of Science and Engineering Education, 3(1), 1-10.
4. Arinaldi, A., Pradana, J. A., & Gurusinga, A. A. (2018). Detection and classification of vehicles for traffic video analytics. Procedia computer science, 144, 259-268.
5. Budiharto, W., & Meiliana, M. (2018). Prediction and analysis of Indonesia Presidential election from Twitter using sentiment analysis. Journal of Big data, 5(1), 1-10.
6. Bizer, C., Boncz, P., Brodie, M. L., & Erling, O. (2012). The meaningful use of big data: four perspectives--four challenges. ACM Sigmod Record, 40(4), 56-60.

7. Contreras, F., & Abid, G. (2022). Social sustainability studies in the 21st century: A bibliometric mapping analysis using VOSviewer Software. Pakistan Journal of Commerce and Social Sciences (PJCSS), 16(1), 167-203.
8. De Lima, C., Belot, D., Berkvens, R., Bourdoux, A., Dardari, D., Guillaud, M., ... & Wymeersch, H. (2021). Convergent communication, sensing and localization in 6G systems: An overview of technologies, opportunities and challenges. IEEE Access, 9, 26902-26925.
9. De Mauro, A., Greco, M., & Grimaldi, M. (2015, February). What is big data? A consensual definition and a review of key research topics. In AIP conference proceedings (Vol. 1644, No. 1, pp. 97-104). American Institute of Physics.
10. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. Journal of Business Research, 133, 285-296.
11. Firdaus, A., Razak, M.F.A., Feizollah, A. et al. The rise of "blockchain": bibliometric analysis of blockchain study. Scientometrics **120**, 1289–1331 (2019). https://doi.org/10.1007/s11192-019-03170-4
12. Hamidah, I., Sriyono, S., & Hudha, M. N. (2020). A Bibliometric analysis of Covid-19 research using VOSviewer. Indonesian Journal of Science and Technology, 34-41.
13. Hudha, M. N., Hamidah, I., Permanasari, A., Abdullah, A. G., Rachman, I., & Matsumoto, T. (2020). Low Carbon Education: A Review and Bibliometric Analysis. European Journal of Educational Research, 9(1), 319-329.
14. Huda, M., Maseleno, A., Atmotiyoso, P., Siregar, M., Ahmad, R., Jasmi, K., & Muhamad, N. (2018). Big data emerging technology: insights into innovative environment for online learning resources. International Journal of Emerging Technologies in Learning (iJET), 13(1), 23-36.
15. Huda, M., Maseleno, A., Teh, K. S. M., Don, A. G., Basiron, B., Jasmi, K. A., ... & Ahmad, R. (2018). Understanding Modern Learning Environment (MLE) in Big Data Era. International Journal of Emerging Technologies in Learning, 13(5).
16. Kitchin, R., & McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society, 3(1), 2053951716631130.
17. Koseoglu, M. A., Rahimi, R., Okumus, F., & Liu, J. (2016). Bibliometric studies in tourism. Annals of tourism research, 61, 180-198.
18. Kuzior, A., & Sira, M. (2022). A Bibliometric Analysis of Blockchain Technology Research Using VOSviewer. Sustainability, 14(13), 8206.
19. Nandiyanto, A. B. D., & Al Husaeni, D. F. (2021). A bibliometric analysis of materials research in Indonesian journal using VOSviewer. Journal of Engineering Research.
20. Nandiyanto, A. B. D., & Al Husaeni, D. F. (2022). Bibliometric Analysis of Engineering Research Using Vosviewer Indexed by Google Scholar. Journal of Engineering Science and Technology, 17(2), 883-894.
21. Narin, F., Olivastro, D., & Stevens, K. A. (1994). Bibliometrics/Theory, Practice and Problems. Evaluation Review, 18(1), 65–76. https://doi.org/10.1177/0193841X9401800107
22. Nasir, A., Shaukat, K., Khan, K. I., Hameed, I. A., Alam, T. M., & Luo, S. (2020). What is core and what future holds for blockchain technologies and cryptocurrencies: A bibliometric analysis. IEEE Access, 9, 989-1004.
23. Oyewola, D. O., & Dada, E. G. (2022). Exploring machine learning: a scientometrics approach using bibliometrix and VOSviewer. SN Applied Sciences, 4(5), 1-18.
24. Pence, H. E. (2014). What is big data and why is it important? Journal of Educational Technology Systems, 43(2), 159-171.
25. Ragadhita, R., & Nandiyanto, A. B. D. (2022). Computational bibliometric analysis on publication of techno-economic education. Indonesian Journal of Multidiciplinary Research, 2(1), 213-220.
26. Sánchez, A. D., Del Río, M. D. L. C., & García, J. Á. (2017). Bibliometric analysis of publications on wine tourism in the databases Scopus and WoS. European Research on Management and Business Economics, 23(1), 8-15.
27. Santoso, L. W. (2017). Data warehouse with big data technology for higher education. Procedia Computer Science, 124, 93-99.
28. Shirkhorshidi, A. S., Aghabozorgi, S., Wah, T. Y., & Herawan, T. (2014, June). Big data clustering: a review. In International conference on computational science and its applications (pp. 707-720). Springer, Cham.
29. Soegoto, H., Soegoto, E. S., Luckyardi, S., & Rafdhi, A. A. (2022). A bibliometric analysis of management bioenergy research using VOSviewer application. Indonesian Journal of Science and Technology, 7(1), 89-104.
30. Surbakti, F. P. S., Wang, W., Indulska, M., & Sadiq, S. (2020). Factors influencing effective use of big data: A research framework. Information & Management, 57(1), 103146.
31. Tiwari, S., Wee, H. M., & Daryanto, Y. (2018). Big data analytics in supply chain management between 2010 and 2016: Insights to industries. Computers & Industrial Engineering, 115, 319-330.
32. Van Nunen, K., Li, J., Reniers, G., & Ponnet, K. (2018). Bibliometric analysis of safety culture research. Safety science, 108, 248-258.
33. Villars, R. L., Olofson, C. W., & Eastwood, M. (2011). Big data: What it is and why you should care. White paper, IDC, 14, 1-14.
34. Wibisono, A., Jatmiko, W., Wisesa, H. A., Hardjono, B., & Mursanto, P. (2016). Traffic big data prediction and visualization using fast incremental model trees-drift detection (FIMT-DD). Knowledge-Based Systems, 93, 33-46.