Research Trend of Big Data in Education During the Last 10 Years

https://doi.org/10.3991/ijet.v18i10.38453

Binar Kurnia Prahani[1](✉), Irgy Redityo Dawana[1], Budi Jatmiko[1], Tan Amelia[2]
[1]Universitas Negeri Surabaya, Surabaya, Indonesia
[2]Universitas Dinamika, Surabaya, Indonesia
binarprahani@unesa.ac.id

Abstract—Big Data research has rapidly developed in the last 10 years, making it an interesting topic to investigate to understand the trends and developments of Big Data in Education. This study aims to analyze source documents and state contributions, and visualize research trends on the subject of Big Data in education during the last 10 years and identify potential research topics related to Big Data in Education in the future. This study uses a literature review and Meta-Analyses (PRISMA) method associated with bibliometric analysis using the Scopus and VOSviewer databases, through which 1,076 documents were obtained. The results of the bibliometric analysis show that Big Data research in education has increased significantly in the last 10 years. However, in the last year, it has decreased, this is the next challenge, and an opportunity for future research. The most common types of documents are conference papers, the source of most documents is conference proceedings, and the country that contributes the most is China. English is the most widely spoken language, and the authors with the highest contribution are Daniel, B. K. Further research related to Big Data can be used and implemented in the business world for Big Data analysis in education. Big Data can also be integrated with Science, technology, engineering, and mathematics (STEM) which can be a further research opportunity that can be applied to education because analyzing Big Data requires STEM learning skills. Thus, this research recommends finding updates in the study of Big Data in education by integrating STEM in education because analyzing Big Data again requires STEM learning expertise. This study has limitations in using only one database, namely Scopus, to obtain research data. Therefore, it is recommended that Big Data in Education research be conducted using other databases besides Scopus to obtain more extensive data.

Keywords—bibliometric, Big Data, education, learning, technology

1 Introduction

Currently, the advancement and the emergence of technological innovations have a significant impact so that they can change the world quickly [1]–[5]. The rapidly changing world is also marked by borderless information systems based on computing and Big Data [6]. The changes in question are technological, social, macroeconomic,

educational and other changes based on the 4.0 industrial revolution [7]. The industrial revolution 4.0 is marked by the Internet of things, Big Data and all these new developments that have disrupted various fields of human life, including the education sector, which has a relatively significant impact [8], [9]. However, low digital culture, lack of knowledge and training are also challenges for Industry 4.0, so it is hoped that humans can better adapt to the development of the industrial revolution in every era [10].

Now technological tools are being used by many individuals [11]. A considerable amount of data will be generated through technology tools [12]. Technology and platforms that are currently being developed are needed to meet big data. This technology helps store data sets and analysis [13]. Now, Big Data is an exciting topic to study. Big Data is a new trend covering a wide area, especially in the technology world [14]. Big Data refers to technology that involves data that is diverse, changing rapidly, very large and challenging to process, so technology needs to be able to handle it effectively [15]–[19].

The greater use of technology makes Big Data a fundamental thing and makes it used by every individual. Especially now that we have entered the era of Big Data, where there is an exponential growth in the amount of data and very rapid technological developments [20]–[23]. Big Data guarantees the processing of data solutions with new ones and provides real benefits, especially for education [24], [25]. Education is one of the main pillars of the progress of the nation's next-generation, it also needs to keep up with the times so that it does not lag developed countries in the field of educational technology and learning curricula [26]–[28]. However, during the COVID-19 pandemic, a new problem emerged, becoming a challenge in the world of education [29]. To address this, the government implements work from home, studies at home and online learning. One way to keep education running during a pandemic is by utilizing technology with e-learning or online learning [30]. However, e-learning requires correct and fast data, and the online learning process requires data from teachers, teachers, student data, and even student learning outcomes [31]. To store data and find the data you need more efficiently, we use Big Data. Big Data is a storage medium that can accommodate all data in the learning process and all data required by educational institutions [32].

According to [33], all data originating from the field of education between teachers and students is currently being processed to improve the quality and experience of the learning process. Sources of data obtained from educational platforms are used to develop services that are adapted to the characteristics of modern education conditioned by the development of educational technology [34]. But based, [35] argues that data mining in education so far has not been completely successful, so not all valuable information can be extracted. This is because Big Data processing requires collaboration between teachers and experts to obtain relevant information from the data reported through educational digital tools/media and resources [36]. Big Data in education analytics will help facilitate new and better educational experiences [37]. A lot needs to be analyzed when applying Big Data concepts in education. Considering that the world of education has diverse data and different variable data, Big Data in education is needed

to manage educational information [38]. Based on research [39], [40] processing and analyzing this large and complex data requires expertise in STEM (Science, Technology, Engineering, and Mathematics). In processing and analyzing big data, expertise in programming, statistics, mathematics, machine learning, and data mining is required, all of which are related to STEM [41], [42]. In addition, the use of technology in education also requires STEM skills. The utilization of technology in education such as e-learning, augmented reality, virtual reality, and game-based learning requires technology and programming skills [43]–[45]. Therefore, it is important for educators and students to have STEM expertise in order to utilize technology and big data in education effectively and efficiently [46], [47].

Improving learning with technology and utilizing big data is very important because it can improve the quality and effectiveness of learning [48].To find out the extent of utilization and processing of Big Data in education. Then a bibliometric analysis is needed to find results and map research trends related to the contribution of Big Data in Education. By conducting a bibliometric analysis, it is possible to identify new and unearthed findings related to the topic of interest that could become potentially important findings in the future [49], [50]. Bibliometric analysis can also provide valid quality reviews and can visualize data comprehensively [51] so that it can increase understanding of the distribution of research topics and provide guidance in future research [52]–[54]. Therefore, it is necessary to map the analysis of Big Data in Education. The purpose of this study is to conduct a literature review related to the topic of Big Data in Education from 2012 to 2022 using bibliometrics analysis to determine the future opportunities of Big Data in education. As for some of the questions asked in the research problem, among others:

1. What is the trend of publishing Big Data in education articles during the last 10 years?
2. How is the visualisation of mapping Big Data research trends in education over the last 10 years?
3. What document sources, types and countries have contributed the most to Big Data in education research over the last 10 years?
4. What are the top source titles, and research citations on Big Data in education research over the last 10 years?
5. What is the author, top affiliates and the language of Big Data in education research over the last 10 years?
6. How do the opportunities for publication of Big Data in education research?

Based on studies that have been conducted on Big Data in Education, this article provides a detailed description of the literature using the Preferred Reporting Items for Review and Meta-Analyses (PRISMA) approach and bibliometric research using the VOSviewer software. After that, you will find out the results and article information detected for analysis and will answer about the potential for important and future findings on the topic of Big Data in Education.

2 Methodology

This study uses a systematic review study with bibliometric analysis. Combining these two methodologies will be able to show the evolution of scientific studies quantitatively through the use of bibliometrics and qualitatively using in-depth studies on topics and content through systematic reviews [55]. Bibliometric analysis is a quantitative study using the Scopus database literature study [56]–[58]. The bibliometric analysis method was carried out to see the visualization of the publication trend mapping to evaluate the contribution of articles to the development of knowledge of various literary sources using statistical methods [59], [60]. With bibliometric analysis, novelty and trends in research will be obtained [61]. This research was conducted by online search on 25 January 2023. Researchers conducted an online search by entering 'Big Data Education' in the title, and keywords from 2012 to 2022. The steps in this research are illustrated in Figure 1 [62]–[64].

Fig. 1. Five steps in bibliometric analysis

2.1 Data search and identification

Bibliometric studies use the Scopus database as a data source. Scopus is one of the optimal databases that can be used for bibliometric analysis because it can provide and cover the results of more sources of published journals [65]–[67]. In this study, a systematic review was also carried out following the PRISMA method to resolve key factors in the literature search, papers used, data extraction, and summaries. The database is looking for the "Big Data Education" category using the following string:

"(TITLE (big AND data) AND TITLE (education) OR TITLE (educational))"

Following a search using these keywords, 218,886 documents were discovered. However, the number of documents decreased to 11,255 when the latest 10 years, namely 2012 to 2022, were found. The results were reduced to 1,076 documents after it was once more filtered using the article title filter as a research document. The data is then exported in the formats (.ris) and (.csv), which will subsequently be examined to produce graphs and tables. Furthermore, these records were processed in a different program for bibliometric analysis using MS Excel and VOSviewer [54]. In the eligibility process, the article will be evaluated to check if it meets the characteristics desired by the researcher. Explain the steps used in the PRISMA method and visualize its interpretation in this review can be seen in Figure 2.

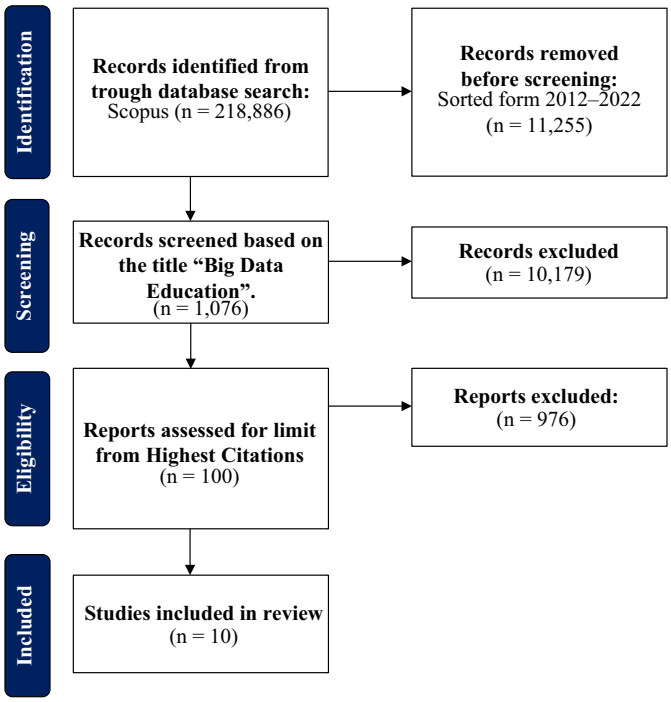


Fig. 2. PRISMA flowchart for the article selection process

2.2 Data analysis

Data analysis in this study was carried out descriptively to determine document source, top affiliation, research citation, country, subject area, top source title, language, authorship, and keywords in 1,076 related documents. Analysis was also carried out by looking at the size of the sphere and the strength of the link relationship based on the results of mapping and visualization using the VOSviewer application. Bibliometric patterns can be created and mapped using the free tool VOSviewer, which uses an in-depth methodology that is simple to understand [68]. In bibliometrics, it will display a network pattern of items through circles whose sizes differ according to the element's significance. While the network connection represents the closeness of the links between items. The position of the circle spacing and the different colours are used to group elements [55]. Bibliometrics can show co-occurrence and collaboration between authors across countries [55], [69]. From this analysis, it is known that some information was obtained, such as the contributions of authors and journals and the materials used according to the selected articles to get an explanation according to the purpose. The final step was to analyze the data by reviewing the top 10 articles cited based on findings and recommendations related to Big Data in Education.

3 Results and discussion

3.1 Research trends, document sources, document types, and countries contribute to Big Data research in education

Figure 3 shows the trend results of Big Data research articles in education over the last 10 years. Based on this figure, it can be shown that developments in the last 10 years in Big Data research in the field of education have increased exponentially. This shows that Big Data research in education is an interesting topic to research today. This is in line with previous research [23], [70], [71], which shows that over the last 10 years, research on big data in education is increasing every year. However, in this study there was a decrease in 2022 as many as 64 publications. This is possible due to the effects of the covid-19 pandemic which has reduced research contributions in 2022 on this topic [72]. Although Big Data research in Education is on the decline, it is an exciting topic to research today and has great potential in education offering extraordinary research possibilities in the future.

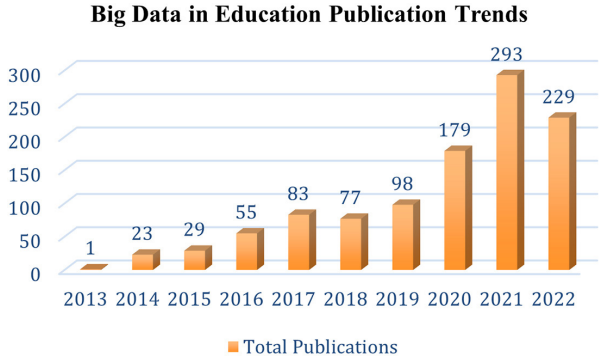


Fig. 3. Year-wise distribution graph on Big Data in education publications

Research in 2018 experienced a decrease in publications by seven publications. However, this decrease did not significantly impact because, in the following years, it experienced a very drastic increase, with the highest increase in the number of publications in 2021 with 290 publications.

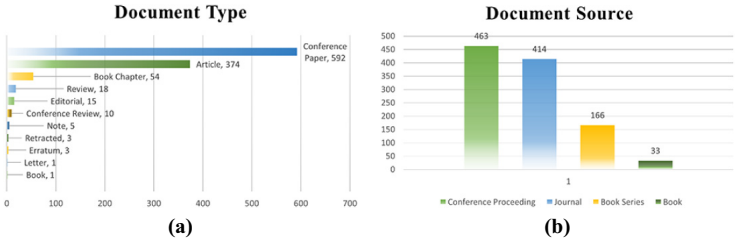


Fig. 4. Document types and sources of Big Data documents in education research

Based on search results by filtering, 1,076 documents were obtained, consisting of 592 conference papers, and 374 articles, which were the most publications in type documents. Moreover, the source of the published document is the conference proceedings of 463 conference proceedings, 414 journals, 166 book series, and 33 book items. The conference proceeding is the most highly published Big Data in education research document. Figure 4 shows the document types and source of Big Data document in education research.

Table 1. Document type from the top 100 publications with the highest citations of Big Data in education

Document Type	Frequency	Cited	Mean	Median	SD
Article	55*	2551*	46.38*	28	55.76*
Editorial	5	207	41.40	17	35.25
Review	9	391	43.44	45*	21.96
Conference Paper	24	609	25.38	21	15.61
Book	1	43	43.00	43	–
Book Chapter	5	146	29.20	27	13.61
Note	1	14	14.00	14	–

Note: *The highest number SD = Standard Deviation.

Table 1 describes the main distribution of document types according to the top 100 documents with the highest citations. The highest document type of publication on this topic is articles (55). The most cited publications are in the form of articles (2,551). So that obtained a high standard deviation of 55.76 in articles. Thus, it can be concluded that researchers are interested in writing the topic of Big Data in education in the form of articles.

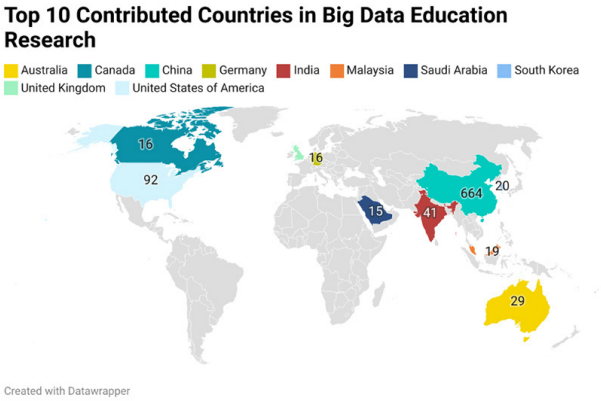


Fig. 5. Top 10 countries' contributions to Big Data research in education during the last 10 years

Figure 5 shows the results of the mapping of the top ten countries that have contributed to Big Data in education research. The country that ranks at the top is China in terms of the highest contribution with 664 articles, then in second place with the United States with 92 articles, then India with 41 articles, followed by others. This is because China has a framework that shows a different interpretation of Big Data from other countries, making China a country with strong Big Data [73]. And this is in line with research [74] that China and the United States are the most contributing and productive countries in Big Data research.

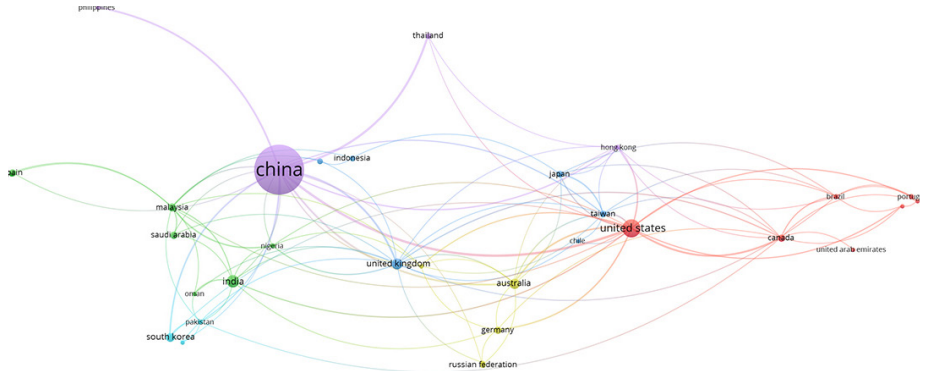


Fig. 6. Mapping by country

Figure 6 is a map of countries based on clusters. There are 31 items with 6 main clusters consisting of China which are interconnected with the Hongkong, Philippines, and Thailand in the purple group. The second cluster consists of seven countries, Brazil, Canada, Israel, Italy, Portugal, United Arab Emirates, and United States in the red cluster. The third cluster includes seven countries: India, Malaysia, Morocco, Nigeria, Oman, Saudi Arabia, and Spain in the green groups. The fourth cluster includes six countries: Chile, Indonesia, Japan, New Zealand, Taiwan, and UK in the dark blue groups. The fifth cluster consist of five countries, Australia, Germany, Netherlands, Russian Federation, and Sweden in the yellow groups. Finally, the sixth cluster consists of tiga countries, Pakistan, South Africa, and South Korea light blue groups. Based on this mapping, the collaboration between countries in Big Data in education research is relatively good.

3.2 Language, top affiliation, and top authorship

Table 2 describes the five main distributions of spoken languages, top affiliate and top authorship, over the last 10 years. Based on the table, it was found that English was the most dominant language with a total of 1,065 articles, followed by Spanish 6 articles, Chinese 3 articles, German 2 articles and Italian 1 article. Based on affiliation, the University of Electronic Science and Technology of China tops the list with 14 published articles. Based on Daniel, B. K. authorship is the most productive contributing writer with five articles. Kim, B. M., Prakash, S., and Ye, L. with four articles, while the others Alam, M with three articles.

Table 2. Top 5 language, affiliates and Big Data authorship in education research during the last 10 years

Top Language		Top Affiliate		Top Authorship	
Language	**Total**	**Affiliation**	**Total**	**Author**	**Total**
English	1,065	University of Electronic Science and Technology of China	14	Daniel, BK	6
Spanish	6	Jilin Engineering Normal University	9	Kim, BM	4
Chinese	3	University of Otago	9	Prakash, S.	4
German	2	Huazhong Normal University	8	Ye, L.	4
Italian	1	Jilin University	7	Alam, M.	3

3.3 Top source title and research citation Big Data in Education research

Table 3 shows source titles, subject areas and authors cited from Big Data research from 2012 to 2022. For example, the title of the highest source, 'ACM International Conference Proceeding Series' is the main source in Big Data research publications with 97 articles. In the highest field of study, the highest discussed was Computer Science (662), followed by Engineering (329), Social Sciences (287), Decision Sciences (133), Physics and Astronomy (128), Mathematics (118), Business, Management and Accounting (50), Environmental Science (42), Medicine (41) and Energy (34).

Table 3. Top 10 source titles, subject areas, and author citations in Big Data in education research during the last 10 years

Top Source Title		Top Subject Area		Top Cited Author		
Source Title	**Total**	**Subject Area**	**Total**	**Author**	**Cited**	**Total Documents**
ACM International Conference Proceeding Series	97	Computer Science	662	Daniel, B.	317	71
Journal of Physics Conference Series	95	Engineering	329	Picciano, AG	248	42
Advances In Intelligent Systems And Computing	43	Social Sciences	287	Gillborn, D., Warmington, P., Demack, S.	188	63
Lecture Notes in Electrical Engineering	33	Decision Sciences	133	Fischer et al.	110	39
Wireless Communications and Mobile Computing	23	Physics and Astronomy	128	Ellaway et al.	106	190
Lecture Notes of The Institute for Computer Sciences Social Informatics and Telecommunications Engineering	22	Mathematics	118	Williamson, B.	90	64

(Continued)

Table 3. Top 10 source titles, subject areas, and author citations in Big Data in education research during the last 10 years *(Continued)*

Top Source Title		Top Subject Area		Top Cited Author		
Source Title	**Total**	**Subject Area**	**Total**	**Author**	**Cited**	**Total Documents**
Mobile Information Systems	19	Business, Management and Accounting	50	Daniel, B.K.	89	71
Lecture Notes in Computer Science Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics	15	Environmental Science	42	Gupta, B., Goul, M., Dinter, B.	83	25
Lecture Notes on Data Engineering and Communications Technologies	14	Medicine	41	Gao, P., Li, J., Liu, S.	80	2
Boletin Tecnico/Technical Bulletin	12	Energy	34	Eynon, R.	80	52

3.4 The distribution of year-wise top 100 highest cited publications

Table 4 lists the top 100 publications on Big Data in education, along with the number of papers published each year, highest citation year, average number of citations per paper and average number of citations per paper annually. For example, the highest publication on Big Data in education occurred in 2018 many as 23. The highest citation occurred in 2018, with as many as 817.

Table 4. The distribution of year-wise from the top 100 highest cited publications

Year	Cited	Paper	ACPP	ACPPY	Citable Years
2012	248	1	248.0*	22.5	11
2013	80	1	80.0	8.0	10
2014	331	8	41.4	4.6	9
2015	645	9	71.7	9.0	8
2016	441	15	29.4	4.2	7
2017	555	20	27.8	4.6	6
2018	817*	23*	35.5	7.1	5
2019	314	10	31.4	7.9	4
2020	400	10	40.0	13.3	3
2021	100	2	50.0	25.0	2
2022	30	1	30.0	30.0*	1
Total	**3961**	**100**	**685.11**	**136.21**	–

Note: ACPP = Average Citation Per Paper, ACPPY = Average Citation Per Paper Per Year, * = the highest number.

3.5 Mapping visualization of research trends in Big Data in education

The most keyword discoveries in Big Data trend mapping in education research over the last 10 years are shown in Table 5. The keyword that most often appears in this research is Big Data.

Table 5. Top 10 keywords of all Big Data in education research during last 10 years

All Big Data in Education Research					
Keywords	**Occurrence**	**Total Link Strength**	**Keywords**	**Occurrence**	**Total Link Strength**
Big Data	918	6573	Engineering education	127	1192
Students	317	2886	Colleges and universities	139	1191
Education	174	1309	Information management	126	1175
Education computing	139	1283	Teaching	115	1169
Data mining	147	1275	E-learning	101	912

There were 918 keywords and 6,573 total link strengths. This indicates that Big Data is the main keyword Big Data in education research. They were then followed by students, education, education computing, data mining, Engineering education, colleges and universities, information management, Teaching, e-learning. Based on the mapping, it can be found that the research trends of Big Data in education in the last 10 years are: 1) Data sources for student education; 2) subjects in the field of technical education; 3) collection of important information in a data; 4) technology learning. This is in accordance with research [23], [70], [75] The most used keyword searches are Big Data, education, technology and education computing.

Fig. 7. The most relevant keywords on Big Data in education

The keywords with the highest number of occurrences and the highest link strength on this topic are Big Data, students, education, data mining and so on. Figure 7 shows a graphic of the most used keywords with a larger font size to represent the frequency of the most used keywords. The occurrence rate and the total strength of the links indicate the researcher's interest in the topic [76]. The mapping results of Big Data research trends in education over the last 10 years are shown in Figure 8. The mapping aims to find the latest research-based on mapping trends. The mapping results show eight clusters on Big Data in education research. The first cluster in red (n = 428) focuses on learning methods such as Artificial intelligent, teaching, curriculum, education technology, mobile learning, and innovative teaching. Next, clusters with green colour are obtained (n = 84) focusing on Big Data education such as Big data analyze, data science, higher education, and student learning, etc. Next, clusters in dark blue are obtained (n = 77), focusing on education computing such as colleges and universities, college students, Curriculum systems, etc. Next, the cluster is yellow (n = 53) focusing on engineering education such as e-learning, learning systems, etc. Clusters with purple colour (n = 49), focusing on education such as data analysis, etc. Finally, the cluster in light blue (n = 32) focuses on data technology, such as education system, application, and science and technology, etc., and the other clusters have more occurrences of minor keywords and research focus.

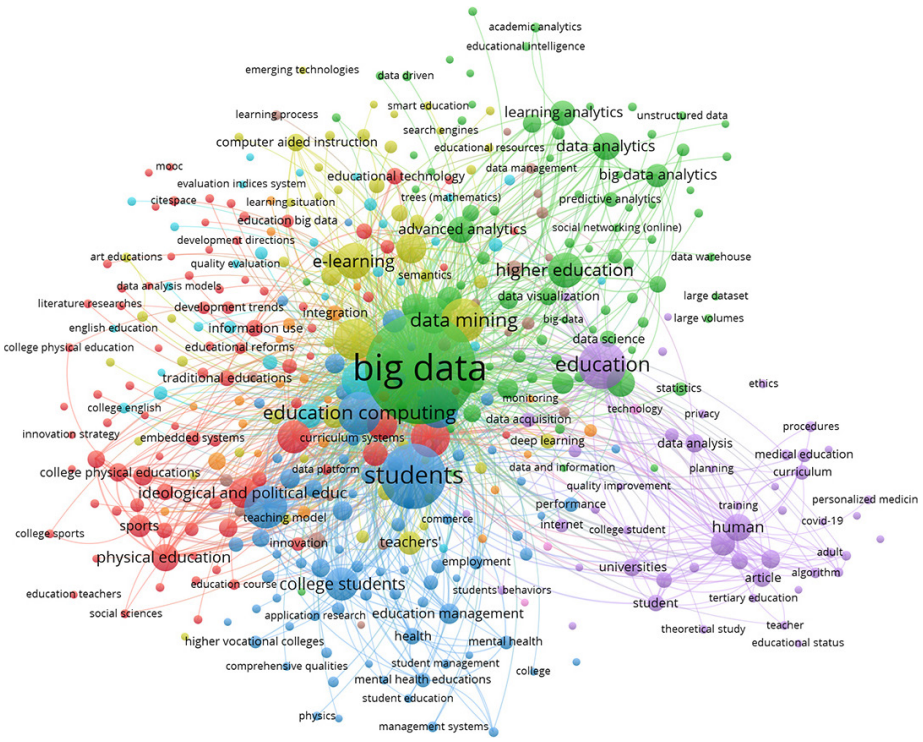
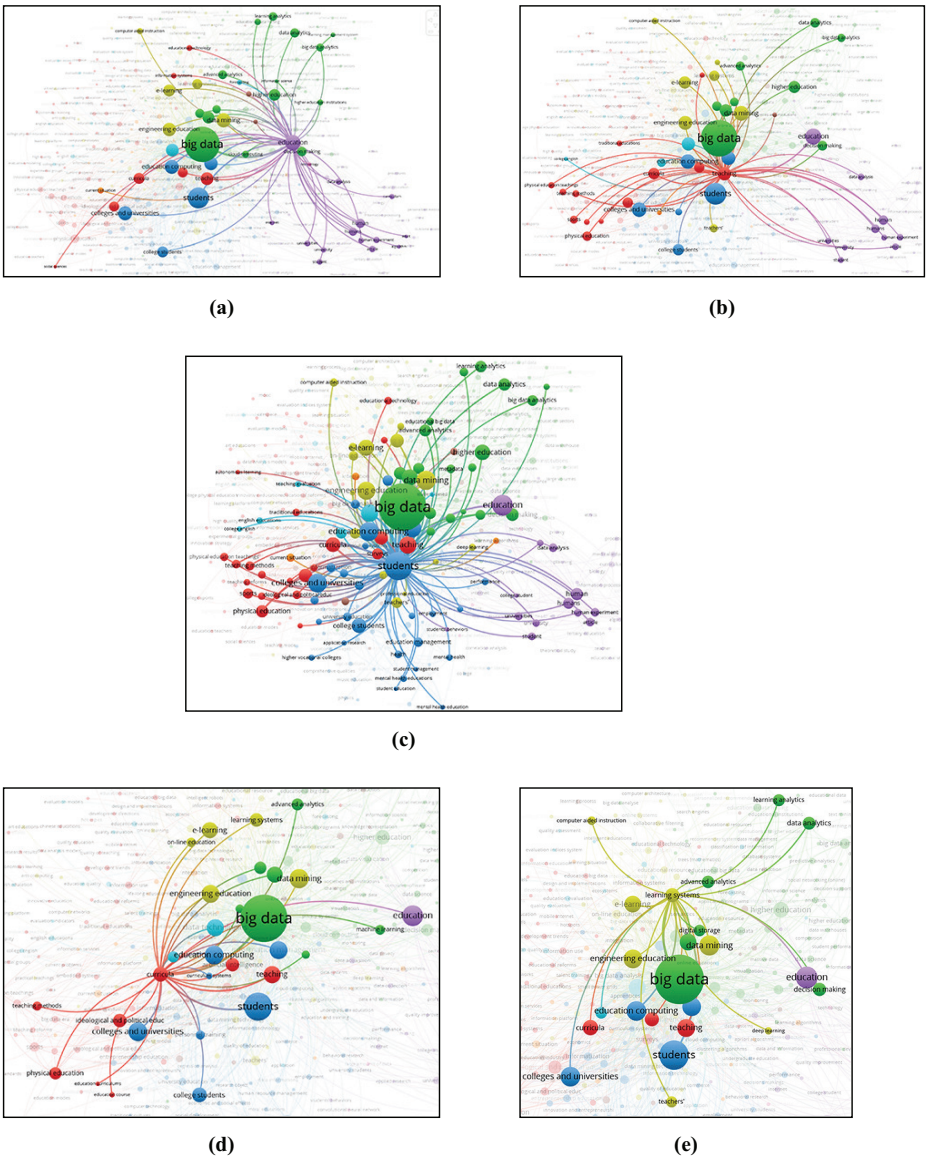


Fig. 8. Co-occurrence keyword visualisation mapping across all Big Data in education research during the last 10 years

In finding the updates of the mapping results, that is, by looking at the relationship between fewer or smaller keywords [77]. The biggest Big Data implementations focus on students with the most keywords and have been heavily researched in the last 10 years. Meanwhile, there is still little research on learning science and technology on Big Data because the keywords obtained are relatively few. This can be a potential for Big Data research now and in the future.



(a) **(b)** **(c)** **(d)** **(e)**

Fig. 9. More specific keyword mapping results with topics a) education, b) teaching, c) students, d) curriculum and e) leaning systems

Figure 9 is an example of more specific keyword mapping results on education, teaching, learners, curriculum and learning systems. The top trends of Big Data research in education over the past 10 years shown in Figure 9a–c have great potential for further exploration as the topic has a wide range and is flexible in its application. This is because Big Data can contribute to education through many aspects. Then, Figure 9d–e have a pattern of many connection relationships, but not as extensive and less for research trends. However, this topic can also be explored further as Big Data also contributes and focuses on the curriculum and learning system. For example, by utilizing Big Data in Education in the STEM field to improve algorithm skills to process data effectively and efficiently [78], [79].

According to [80] The growing utilization of information technology has provided opportunities to obtain and collect data on a large scale in various fields, including in the field of education and curriculum. After mapping research trends, it has not been seen or there are still few who conduct research related to big data in STEM learning, even though according to [47], [81] In improving learning with technology and utilizing big data, collaboration between educators, STEM experts, and education experts is needed to develop appropriate and relevant learning approaches by utilizing technology and big data optimally. This is a potential for further research because big data in STEM learning has a suitable relationship. because Big Data and STEM are interrelated in realizing contributions in the world of education.

3.6 Contribution of Big Data to education and STEM education

Based on Table A1 Top 10 highest cited papers showed the contribution of Big Data in Education and STEM Education in the Appendix, it can be seen that research related to Big Data has great potential and opportunities, but there are challenges in the field of education, so it is necessary to improve learning Big Data analytics and platforms to support education infrastructure [82]–[85]. In analysing Big Data, it is also necessary to develop learning analytics and Big Data business analysis to support and improve success that can be implemented in education [86]. So that by utilising Big Data can help understand various learning problems [87]–[90]. Due to the current shortage of human resources in STEM, Big data analysis is necessary to increase student enthusiasm in STEM learning and raise the standard of STEM higher education.

Big Data in education also has a significant role in STEM education [91]–[94]. STEM education can emphasize the cognitive development of computational, creative and metacognitive thinking [95]–[98]. STEM can influence the development of non-cognitive characteristics of learners so that they can improve critical thinking and can solve real-world problems [99]–[102]. STEM education is very important to be taught from an early age. STEM should be practiced from an early age, to determine the success of STEM in early education. STEM requires new patterns of learning and knowledge that will involve critical thinking as well as curiosity and analysis [103], [104]. With the application of STEM in children, it can affect children's attitude towards science. STEM education can support children's learning outcomes with learning from teachers in doing STEM activities, but lacking science and math skills and computing education, teachers may struggle in providing STEM education content

knowledge [105]–[107]. Therefore teachers are required for science and math learning courses to provide quality STEM education so that teachers can develop and provide content knowledge in STEM education [105].

The application of Big Data to predict future development trends, especially in education. Like research from [93], [108], Big data helps understand the role of virtual reality technology in STEAM education, integrate it into the teaching curriculum, assist teachers in learning activities and make the learning process more colourful. The use of Big Data analysts helps in the STEM learning process, but the lack of Big Data skills, especially in Indonesia, makes it difficult to understand Big Data processing [109]. To overcome this, a STEM education approach can improve critical thinking skills to process Big Data and face challenges in the industrial revolution 4.0 [110].

Big Data has a significant contribution to STEM education and education, and STEM can also be used to improve Big Data processing and analysis. In education, the use of Big Data can help educators to develop more personalized and effective learning approaches, by understanding students' learning needs, developing more relevant curriculum, and improving teaching effectiveness and efficiency [111]. In STEM education, Big Data can also help improve learning by analysing student learning data and developing more effective and personalized learning strategies [112]–[114]. STEM can be used to improve Big Data processing and analysis in education [115]. STEM expertise in programming, statistics, mathematics, and technology can help in developing algorithms and applications to process data effectively and efficiently [116]. In addition, STEM can also be used to develop technologies and platforms to collect, store, and analyse data on a large scale [117], [118].

Big Data has significant contributions to education and STEM education. By utilizing Big Data technology, educational institutions can improve teaching effectiveness, enhance student experience, develop curriculum, assist decision-making, and drive innovation in education and STEM education. From the results of the Big Data in Education research tend, there is little discussion related to Big data in Education for STEM so this is a recommendation for future research. Suggestions for future research related to Big data in Education need to be developed further because this topic is very interesting and has potential in the field of education, especially in the STEM field. In addition, for future bibliometric research, it is necessary to consider using more diverse database sources other than just relying on Scopus. This will allow for more extensive and detailed data collection.

4 Conclusion

Based on the results of bibliometric analysis and literature review, the following findings were obtained: The trend of publication topics for Big Data in education is exponentially increasing from 2012 to 2022. However, in 2022 it has decreased, this is a challenge for future research, whether there is an increase in research topics, because this topic has the potential to be researched in the future. Based on research mapping, it is known that the research trends of Big Data in education in the last 10 years are: 1) application to teachers and students in education, schools and universities; 2) technology-based education by integrating STEM; 3) business application for

Big Data analytics in education. The most widely published documents from Big Data in education research are conference papers (592), and conference proceedings source journal documents (463). In contrast, the most productive country in Big Data in education research is China (664). In the top source title, "ACM International Conference Proceeding Series", is the primary source in Big Data in education research publications (97). The most studied fields of study are Computer Science (662), Engineering (329) and Social Sciences (287). And the top author citation is Daniel, as the author with the most citations (317). The most prolific writer is Daniel, B. K. (6). The top affiliate researching Big Data in education the most is the University of Electronic Science and Technology of China (14), and English's most widely spoken language (1,065).

The implication of this research is that it found some examples of novelty in on the topic of Big Data in Education so that this research can be used as a reference for future research on the topic of Big Data in Education. The type of research presented makes it possible to determine the profile of document types that can be presented so that the research path to be pursued can be more directed. This research can also find the most relevant articles related to Big Data in Education on the Scopus database and the authors who have the most significant relationships from each journal. Therefore, this study also limits future research trends that can develop on this research topic. The research recommends finding updates in the study of Big Data in education by integrating STEM in education because analyzing Big Data requires STEM learning skills. The existence of this research can be used as a reference for further research on Big Data in education. For further research related to Big Data in Education, it is recommended to develop by utilizing other database sources besides Scopus.

5 Acknowledgement

The author's gratitude goes to Direktorat Jenderal Pendidikan Tinggi, Riset, dan Teknologi; Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi, Indonesia who has supported funding in Penelitian Dasar Unggulan Perguruan Tinggi (PDUPT) – DRTPM 2023.

6 References

[1] B. (Kevin) Chae, "A general framework for studying the evolution of the digital innovation ecosystem: The case of big data," *Int. J. Inf. Manage.*, vol. 45, 83–94, 2019, https://doi.org/10.1016/j.ijinfomgt.2018.10.023

[2] S. E. Bibri and J. Krogstie, "A novel model for data-driven smart sustainable cities of the future: A strategic roadmap to transformational change in the era of big data," *Futur. Cities Environ.*, vol. 7, no. 1, 1–25, 2021, https://doi.org/10.5334/fce.116

[3] W. Li *et al.*, "A comprehensive survey on machine learning-based big data analytics for IoT-enabled smart healthcare system," *Mob. Networks Appl.*, vol. 26, no. 1, 234–252, 2021, https://doi.org/10.1007/s11036-020-01700-6

[4] M. I. Qureshi, N. Khan, H. Raza, A. Imran, and F. Ismail, "Digital technologies in education 4.0. Does it enhance the effectiveness of learning?," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 4, 31–47, 2021, https://doi.org/10.3991/ijim.v15i04.20291

[5] Z. Lv, R. Lou, J. Li, A. K. Singh, and H. Song, "Big data analytics for 6g-enabled massive internet of things," *IEEE Internet Things J.*, vol. 8, no. 7, 5350–5359, 2021, https://doi.org/10.1109/JIOT.2021.3056128

[6] P. Saeidi, S. P. Saeidi, S. Sofian, S. P. Saeidi, M. Nilashi, and A. Mardani, "The impact of enterprise risk management on competitive advantage by moderating role of information technology," *Comput. Stand. Interfaces*, vol. 63, 67–82, 2019, https://doi.org/10.1016/j.csi.2018.11.009

[7] A. Martinelli, A. Mina, and M. Moggi, "The enabling technologies of industry 4.0: Examining the seeds of the fourth industrial revolution," *Ind. Corp. Chang.*, vol. 30, no. 1, 161–188, 2021, https://doi.org/10.1093/icc/dtaa060

[8] P. Brandtner, F. Darbanian, T. Falatouri, and C. Udokwu, "Impact of COVID-19 on the customer end of retail supply chains: A big data analysis of consumer satisfaction," *Sustain.*, vol. 13, no. 3, 1–18, 2021, https://doi.org/10.3390/su13031464

[9] A. Sharifi, M. Ahmadi, and A. Ala, "The impact of artificial intelligence and digital style on industry and energy post-COVID-19 pandemic," *Environ. Sci. Pollut. Res.*, vol. 28, no. 34, 46964–46984, 2021, https://doi.org/10.1007/s11356-021-15292-5

[10] A. Hariharasudan and S. Kot, "A scoping review on digital English and education 4.0 for industry 4.0," *Soc. Sci.*, vol. 7, no. 11, 2018, https://doi.org/10.3390/socsci7110227

[11] M. Shorfuzzaman, M. S. Hossain, A. Nazir, G. Muhammad, and A. Alamri, "Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment," *Comput. Human Behav.*, vol. 92, 578–588, 2019, https://doi.org/10.1016/j.chb.2018.07.002

[12] M. H. ur Rehman, I. Yaqoob, K. Salah, M. Imran, P. P. Jayaraman, and C. Perera, "The role of big data analytics in industrial internet of things," *Futur. Gener. Comput. Syst.*, vol. 99, 247–259, 2019, https://doi.org/10.1016/j.future.2019.04.020

[13] S. A. Kalaian, R. Kasim, and N. Kasim, *Descriptive and predictive analytical methods for big data*. 2019, https://doi.org/10.4018/978-1-5225-7501-6.ch018

[14] O. Saritas, P. Bakhtin, I. Kuzminov, and E. Khabirova, "Big data augmentated business trend identification: The case of mobile commerce," *Scientometrics*, vol. 126, no. 2, 1553–1579, 2021, https://doi.org/10.1007/s11192-020-03807-9

[15] L. Bote-Curiel, S. Muñoz-Romero, A. Gerrero-Curieses, and J. L. Rojo-álvarez, "Deep learning and big data in healthcare: A double review for critical beginners," *Appl. Sci.*, vol. 9, no. 11, 2019, https://doi.org/10.3390/app9112331

[16] A. Ali, J. Qadir, R. ur Rasool, A. Sathiaseelan, A. Zwitter, and J. Crowcroft, "Big data for development: Applications and techniques," *Big Data Anal.*, vol. 1, no. 1, 2016, https://doi.org/10.1186/s41044-016-0002-4

[17] J. Qadir, A. Ali, R. ur Rasool, A. Zwitter, A. Sathiaseelan, and J. Crowcroft, "Crisis analytics: Big data-driven crisis response," *J. Int. Humanit. Action*, vol. 1, no. 1, 1–21, 2016, https://doi.org/10.1186/s41018-016-0013-9

[18] S. H. Akundi, R. Soujanya, and P. M. Madhuri, "Big data analytics in healthcare using machine learning algorithms: A comparative study," *Int. J. online Biomed. Eng.*, vol. 16, no. 13, 19–32, 2020, https://doi.org/10.3991/ijoe.v16i13.18609

[19] G. Y. Wang, D. D. Cheng, D. Y. Xia, and H. H. Jiang, "Swarm intelligence research: From bio-inspired single-population swarm intelligence to human-machine hybrid swarm intelligence," *Mach. Intell. Res.*, vol. 20, no. 1, 121–144, 2023, https://doi.org/10.1007/s11633-022-1367-7

[20] A. Adadi, *A survey on data-efficient algorithms in big data era*, vol. 8, no. 1. Springer International Publishing, 2021, https://doi.org/10.1186/s40537-021-00419-9

[21] O. Azeroual and R. Fabre, "Processing big data with apache hadoop in the current challenging era of COVID-19," *Big Data Cogn. Comput.*, vol. 5, no. 1, 2021, https://doi.org/10.3390/bdcc5010012

[22] M. Beerkens, "An evolution of performance data in higher education governance: A path towards a 'big data' era?," *Qual. High. Educ.*, vol. 28, no. 1, 29–49, 2022, https://doi.org/10.1080/13538322.2021.1951451

[23] J.-A. Marín-Marín, J. López-Belmonte, J.-M. Fernández-Campoy, and J.-M. Romero-Rodríguez, "Big data in education. A bibliometric review," *Soc. Sci.*, vol. 8, no. 223, 8, 2019, https://doi.org/10.1353/jaas.2009.0001

[24] Z. R. Alashhab, M. Anbar, M. M. Singh, Y. B. Leau, Z. A. Al-Sai, and S. A. Alhayja'a, "Impact of coronavirus pandemic crisis on technologies and cloud computing applications," *J. Electron. Sci. Technol.*, vol. 19, no. 1, 25–40, 2021, https://doi.org/10.1016/j.jnlest.2020.100059

[25] M. Talebkhah, A. Sali, M. Marjani, M. Gordan, S. J. Hashim, and F. Z. Rokhani, "IoT and big data applications in smart cities: Recent advances, challenges, and critical issues," *IEEE Access*, vol. 9, 55465–55484, 2021, https://doi.org/10.1109/ACCESS.2021.3070905

[26] S. Aktan, "Waking up to the dawn of a new era: Reconceptualization of curriculum post Covid-19," *Prospects*, vol. 51, no. 1–3, 205–217, 2021, https://doi.org/10.1007/s11125-020-09529-3

[27] R. Kijima, M. Yang-yoshihara, and M. S. Maekawa, "Using design thinking to cultivate the next generation of female STEAM thinkers," *Int. J. STEM Educ.*, vol. 6, 2021, https://doi.org/10.1186/s40594-021-00271-6

[28] J. Mensah, "Sustainable development: Meaning, history, principles, pillars, and implications for human action: Literature review," *Cogent Soc. Sci.*, vol. 5, no. 1, 2019, https://doi.org/10.1080/23311886.2019.1653531

[29] S. Saikat, J. S. Dhillon, W. F. W. Ahmad, and R. A. Jamaluddin, "A systematic review of the benefits and challenges of mobile learning during the covid-19 pandemic," *Educ. Sci.*, vol. 11, no. 9, 2021, https://doi.org/10.3390/educsci11090459

[30] Z. Fajri, H. Baharun, C. Muali, Shofiatun, L. Farida, and Y. Wahyuningtiyas, "Student's learning motivation and interest; the effectiveness of online learning during COVID-19 pandemic," *J. Phys. Conf. Ser.*, vol. 1899, no. 1, 2021, https://doi.org/10.1088/1742-6596/1899/1/012178

[31] M. A. Khan, M. Khojah, and Vivek, "Artificial intelligence and big data: The advent of new pedagogy in the adaptive e-learning system in the higher educational institutions of Saudi Arabia," *Educ. Res. Int.*, vol. 2022, 1–10, 2022, https://doi.org/10.1155/2022/1263555

[32] M. A. Ashaari, K. S. D. Singh, G. A. Abbasi, A. Amran, and F. J. Liebana-Cabanillas, "Big data analytics capability for improved performance of higher education institutions in the Era of IR 4.0: A multi-analytical SEM & ANN perspective.," *Technol. Forecast. Soc. Change*, vol. 173, 121119, 2021, https://doi.org/10.1016/j.techfore.2021.121119

[33] J. Liang, J. Yang, Y. Wu, C. Li, and L. Zheng, "Big data application in education: Dropout prediction in edx MOOCs," *Proc. – 2016 IEEE 2nd Int. Conf. Multimed. Big Data, BigMM 2016*, 440–443, 2016, https://doi.org/10.1109/BigMM.2016.70

[34] A. Merceron, P. Blikstein, and G. Siemens, "Learning analytics: From big data to meaningful data," *J. Learn. Anal.*, vol. 2, no. 3, 4–8, 2016, https://doi.org/10.18608/jla.2015.23.2

[35] A. Menon, S. Gaglani, M. R. Haynes, and S. Tackett, "Using 'big data' to guide implementation of a web and mobile adaptive learning platform for medical students," *Med. Teach.*, vol. 39, no. 9, 975–980, 2017, https://doi.org/10.1080/0142159X.2017.1324949

[36] M. Huda, A. Maseleno, M. Shahrill, K. A. Jasmi, I. Mustari, and B. Basiron, "Exploring adaptive teaching competencies in big data era," *Int. J. Emerg. Technol. Learn.*, vol. 12, no. 3, 68–83, 2017, https://doi.org/10.3991/ijet.v12i03.6434

[37] J. R. Reidenberg and F. Schaub, "Achieving big data privacy in education," *Theory Res. Educ.*, vol. 16, no. 3, 263–279, 2018, https://doi.org/10.1177/1477878518805308

[38] Z. Yang and M. Talha, "A coordinated and optimized mechanism of artificial intelligence for student management by college counselors based on big data," *Comput. Math. Methods Med.*, vol. 2021, 2021, https://doi.org/10.1155/2021/1725490

[39] K. F. Kee, B. Le, and K. Jitkajornwanich, "If you build it, promote it, and they trust you, then they will come: Diffusion strategies for science gateways and cyberinfrastructure adoption to harness big data in the science, technology, engineering, and mathematics (STEM) community," *Concurr. Comput. Pract. Exp.*, vol. 33, no. 19, e6192, 2021, https://doi.org/10.1002/cpe.6192

[40] G. Giacomello and O. Preka, "The 'social' side of big data: Teaching BD analytics to political science students," *Big Data Cogn. Comput.*, vol. 4, no. 2, 13, 2020, https://doi.org/10.3390/bdcc4020013

[41] E. P. Xing *et al.*, "Petuum: A new platform for distributed machine learning on big data," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2015, 1335–1344, https://doi.org/10.1145/2783258.2783323

[42] R. S. Robeva, J. R. Jungck, and L. J. Gross, "Changing the nature of quantitative biology education: Data science as a driver," *Bull. Math. Biol.*, vol. 82, 1–30, 2020, https://doi.org/10.1007/s11538-020-00785-0

[43] N. Pellas, P. Fotaris, I. Kazanidis, and D. Wells, "Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning," *Virtual Real.*, vol. 23, no. 4, 329–346, 2019, https://doi.org/10.1007/s10055-018-0347-2

[44] N. M. Alzahrani, "Augmented reality: A systematic review of its benefits and challenges in e-learning contexts," *Appl. Sci.*, vol. 10, no. 16, 5660, 2020, https://doi.org/10.3390/app10165660

[45] S. Nuanmeesri and L. Poomhiran, "Perspective electrical circuit simulation with virtual reality.," *Int. J. Online Biomed. Eng.*, vol. 15, no. 5, 2019, https://doi.org/10.3991/ijoe.v15i05.9653

[46] Z.-T. Zhu, M.-H. Yu, and P. Riezebos, "A research framework of smart education," *Smart Learn. Environ.*, vol. 3, 1–17, 2016, https://doi.org/10.1186/s40561-016-0026-2

[47] D. J. Shernoff, S. Sinha, D. M. Bressler, and L. Ginsburg, "Assessing teacher education and professional development needs for the implementation of integrated approaches to STEM education," *Int. J. STEM Educ.*, vol. 4, 1–16, 2017, https://doi.org/10.1186/s40594-017-0068-1

[48] M. Shorfuzzaman, M. S. Hossain, A. Nazir, G. Muhammad, and A. Alamri, "Harnessing the power of big data analytics in the cloud to support learning analytics in mobile learning environment," *Comput. Human Behav.*, vol. 92, 578–588, 2019, https://doi.org/10.1016/j.chb.2018.07.002

[49] N. T. Van, A. F. Abbas, H. Abuhassna, F. Awae, and D. Dike, "Digital readiness for social educators in health care and online learning during COVID-19 pandemic: A bibliometric analysis," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 18, 104–115, 2021, https://doi.org/10.3991/ijim.v15i18.25529

[50] S. Parmar, A. K. Siwach, and A. Kumar, "Fifty years research output in oral submucous fibrosis: A bibliometric analysis of publications from 1967 to 2016," *DESIDOC J. Libr. Inf. Technol.*, vol. 40, no. 2, 470–478, 2020, https://doi.org/10.14429/djlit.40.02.14727

[51] P. K. Priyan, W. I. Nyabakora, and G. Rwezimula, "A bibliometric review of the knowledge base on financial inclusion," *SN Bus. Econ.*, vol. 3, no. 2, 1–25, 2023, https://doi.org/10.1007/s43546-023-00441-4

[52] J. M. Merigó and J. B. Yang, "A bibliometric analysis of operations research and management science," *Omega (United Kingdom)*, vol. 73, 37–48, 2017, https://doi.org/10.1016/j.omega.2016.12.004

[53] I. Zupic and T. Čater, "Bibliometric methods in management and organization," *Organ. Res. methods*, vol. 18, no. 3, 429–472, 2015, https://doi.org/10.1177/1094428114562629
[54] K. F. Hashim, A. Rashid, and S. Atalla, "Social media for teaching and learning within higher education institution: A bibliometric analysis of the literature (2008–2018)," *Int. J. Interact. Mob. Technol.*, vol. 12, no. 7, 4–19, 2018, https://doi.org/10.3991/ijim.v12i7.9634
[55] S. Pizzi, A. Caputo, A. Corvino, and A. Venturelli, "Management research and the UN sustainable development goals (SDGs): A bibliometric investigation and systematic review," *J. Clean. Prod.*, vol. 276, 124033, 2020, https://doi.org/10.1016/j.jclepro.2020.124033
[56] S. Admoko, N. R. J. Mukhayyarotin, and E. H. Madlazim, "Bibliometric profile of science education research on argumentation and the contribution of Indonesia," vol. 209, IJCSE, 502–509, 2021.
[57] V. Chhapola, S. Tiwari, B. Deepthi, and S. K. Kanwal, "Citation classics in pediatrics: A bibliometric analysis," *World J. Pediatr.*, vol. 14, no. 6, 607–614, 2018, https://doi.org/10.1007/s12519-018-0146-6
[58] M. Mishra *et al.*, "An overview of research on natural resources and indigenous communities: A bibliometric analysis based on Scopus database (1979–2020)," *Environ. Monit. Assess.*, vol. 193, no. 2, 2021, https://doi.org/10.1007/s10661-020-08793-2
[59] M. Karakus, A. Ersozlu, and A. C. Clark, "Augmented reality research in education: A bibliometric study," *Eurasia J. Math. Sci. Technol. Educ.*, vol. 15, no. 10, 2019, https://doi.org/10.29333/ejmste/103904
[60] Y. M. Guo, Z. L. Huang, J. Guo, H. Li, X. R. Guo, and M. J. Nkeli, "Bibliometric analysis on smart cities research," *Sustain.*, vol. 11, no. 13, 2019, https://doi.org/10.3390/su11133606
[61] N. Suprapto *et al.*, "The comparison of Scimago Institutions Rankings (SIR), Scopus, and SINTA profile: A case of the top Indonesian Institutions," *Libr. Philos. Pract.*, vol. 5788, 1–11, 2021.
[62] I. Setyaningsih, N. Indarti, and F. Jie, "Bibliometric analysis of the term 'green manufacturing,'" *Int. J. Manag. Concepts Philos.*, vol. 11, no. 3, 315, 2018, https://doi.org/10.1504/IJMCP.2018.093500
[63] S. Haryandi, S. Suyidno, M. Misbah, D. Dewantara, S. Mahtari, and M. A. Ibrahim, "Scientific creativity: A bibliometric review and analysis," *Momentum Phys. Educ. J.*, vol. 5, no. 1, 10–20, 2021, https://doi.org/10.21067/mpej.v5i1.5002
[64] P. N. Ari Masitoh, S. Latifah, A. Saregar, A. Aziz, Suharto, and W. Jamaluddin, "Bibliometric analysis of physics problem solving," *IOP Conf. Ser. Earth Environ. Sci.*, vol. 1796, no. 1, 2021, https://doi.org/10.1088/1742-6596/1796/1/012009
[65] T. Karakose, T. Tülübaş, and S. Papadakis, "Revealing the intellectual structure and evolution of digital addiction research: An integrated bibliometric and science mapping approach," *Int. J. Environ. Res. Public Health*, vol. 19, no. 22, 2022, https://doi.org/10.3390/ijerph192214883
[66] P. Mongeon and A. Paul-Hus, "The journal coverage of web of science and scopus: A comparative analysis," *Scientometrics*, vol. 106, 213–228, 2016, https://doi.org/10.1007/s11192-015-1765-5
[67] R. Yirci, T. Karakose, I. Kocabas, T. Tülübaş, and S. Papadakis, "A bibliometric review of the knowledge base on mentoring for the professional development of school administrators," *Sustainability*, vol. 15, no. 4, 3027, 2023, https://doi.org/10.3390/su15043027
[68] R. A. M. Boloy *et al.*, "Waste-to-energy technologies towards circular economy: A systematic literature review and bibliometric analysis," *Water. Air. Soil Pollut.*, vol. 232, no. 7, 2021, https://doi.org/10.1007/s11270-021-05224-x
[69] I. Mustapha, N. T. Van, M. Shahverdi, M. I. Qureshi, and N. Khan, "Effectiveness of digital technology in education during COVID-19 pandemic. A bibliometric analysis," *Int. J. Interact. Mob. Technol.*, vol. 15, no. 8, 136–154, 2021, https://doi.org/10.3991/ijim.v15i08.20415

[70] A. Kalantari *et al.*, "A bibliometric approach to tracking big data research trends," *J. Big Data*, vol. 4, no. 1, 1–18, 2017, https://doi.org/10.1186/s40537-017-0088-1
[71] C. Ye, "The research development of big data in education: A bibliometric analysis based on citespace," *Proc. – 2018 7th Int. Conf. Educ. Innov. through Technol. EITT 2018*, 116–122, 2018, https://doi.org/10.1109/EITT.2018.00031
[72] N. Grammes *et al.*, "Research output and international cooperation among countries during the COVID-19 pandemic: Scientometric analysis," *J. Med. Internet Res.*, vol. 22, no. 12, 1–12, 2020, https://doi.org/10.2196/24514
[73] Y. Shi, Z. Shan, J. Li, and Y. Fang, "How China deals with big data," *Ann. Data Sci.*, vol. 4, no. 4, 433–440, 2017, https://doi.org/10.1007/s40745-017-0129-9
[74] Y. Huang, Y. Zhang, J. Youtie, A. L. Porter, and X. Wang, "How does national scientific funding support emerging interdisciplinary research: A comparison study of big data research in the US and China," *PLoS One*, vol. 11, no. 5, 1–20, 2016, https://doi.org/10.1371/journal.pone.0154509
[75] D. R. Raban and A. Gordon, "The evolution of data science and big data research: A bibliometric analysis," *Scientometrics*, vol. 122, no. 3, 1563–1581, 2020, https://doi.org/10.1007/s11192-020-03371-2
[76] K. Sharma and P. Khurana, "Growth and dynamics of Econophysics: A bibliometric and network analysis," *Scientometrics*, vol. 126, no. 5, 4417–4436, 2021, https://doi.org/10.1007/s11192-021-03884-4
[77] J. Swacha, "State of research on gamification in education: A bibliometric survey," *Educ. Sci.*, vol. 11, no. 2, 1–15, 2021, https://doi.org/10.3390/educsci11020069
[78] V. Mayer-Schönberger and K. Cukier, *Big data: A revolution that will transform how we live, work, and think*. Houghton Mifflin Harcourt, 2013.
[79] M.-L. Wang, W. Shao, X.-K. Hao, and D.-Q. Zhang, "Machine learning for brain imaging genomics methods: A review," *Mach. Intell. Res.*, vol. 20, no. 1, 57–78, 2023, https://doi.org/10.1007/s11633-022-1361-0
[80] X. Li *et al.*, "Curriculum reform in big data education at applied technical colleges and universities in China," *IEEE Access*, vol. 7, 125511–125521, 2019, https://doi.org/10.1109/ACCESS.2019.2939196
[81] K.-Y. Lin, Y.-T. Wu, Y.-T. Hsu, and P. J. Williams, "Effects of infusing the engineering design process into STEM project-based learning to develop preservice technology teachers' engineering design thinking," *Int. J. STEM Educ.*, vol. 8, 1–15, 2021, https://doi.org/10.1186/s40594-020-00258-9
[82] B. Williamson, "The hidden architecture of higher education: Building a big data infrastructure for the 'smarter university,'" *Int. J. Educ. Technol. High. Educ.*, vol. 15, no. 1, 1–26, 2018, https://doi.org/10.1186/s41239-018-0094-1
[83] B. K. Daniel, "Big data and data science: A critical review of issues for educational research," *Br. J. Educ. Technol.*, vol. 50, no. 1, 101–113, 2019, https://doi.org/10.1111/bjet.12595
[84] B. Daniel, "Big data and analytics in higher education: Opportunities and challenges," *Br. J. Educ. Technol.*, vol. 46, no. 5, 904–920, 2015, https://doi.org/10.1111/bjet.12230
[85] A. G. Picciano, "The evolution of big data and learning analytics in American higher education," *J. Asynchronous Learn. Netw.*, vol. 16, no. 3, 9–20, 2012, https://doi.org/10.24059/olj.v16i3.267
[86] S. A. El-Seoud, H. F. El-Sofany, M. Abdelfattah, and R. Mohamed, "Big data and cloud computing: Trends and challenges.," *Int. J. Interact. Mob. Technol.*, vol. 11, no. 2, 2017, https://doi.org/10.3991/ijim.v11i2.6561
[87] R. Eynon, "The rise of Big data: What does it mean for education, technology, and media research?," *Learn. Media Technol.*, vol. 38, no. 3, 237–240, 2013, https://doi.org/10.1080/17439884.2013.771783

[88] J. A. Reyes, "The skinny on big data in education: Learning analytics simplified," *TechTrends*, vol. 59, no. 2, 75–80, 2015, https://doi.org/10.1007/s11528-015-0842-1

[89] B. Gupta, M. Goul, and B. Dinter, "Business intelligence and big data in higher education: Status of a multi-year model curriculum development effort for business school undergraduates, MS graduates, and MBAs," *Commun. Assoc. Inf. Syst.*, vol. 36, 449–476, 2015, https://doi.org/10.17705/1CAIS.03623

[90] R. Dubey and A. Gunasekaran, "Education and training for successful career in big data and business analytics," *Ind. Commer. Train.*, vol. 47, no. 4, 174–181, 2015, https://doi.org/10.1108/ICT-08-2014-0059

[91] R. Vasiu and D. Andone, "An analyze and actions to increase the quality in STEM higher education," *2019 9th IEEE Integr. STEM Educ. Conf. ISEC 2019*, 295–299, 2019, https://doi.org/10.1109/ISECon.2019.8882107

[92] L. Jiang and H. Yuan, "Stable parallel algorithms for interdisciplinary computer-based online education with real problem scenarios for STEM education," *Complexity*, vol. 2021, 2021, https://doi.org/10.1155/2021/5432703

[93] X. Hu, Y. Wang, X. Fan, and Z. Ma, "Construction and application of VR/AR-based STEAM curriculum system in primary and middle schools under big data background," *J. Phys. Conf. Ser.*, vol. 1624, no. 3, 2020, https://doi.org/10.1088/1742-6596/1624/3/032049

[94] S. Xu, "Application research of stem education in big data + math core literacy cultivation," *Proc. – 2019 Int. Conf. Virtual Real. Intell. Syst. ICVRIS 2019*, vol. 67, 81–86, 2019, https://doi.org/10.1109/ICVRIS.2019.00029

[95] S. Papadakis, "The impact of coding apps to support young children in computational thinking and computational fluency. A literature review," *Front. Educ.*, vol. 6, 1–12, 2021, https://doi.org/10.3389/feduc.2021.657895

[96] P. Stamatios, "Can preschoolers learn computational thinking and coding skills with ScratchJr? A systematic literature review," *Int. J. Educ. Reform*, 10567879221076076, 2022, https://doi.org/10.1177/10567879221076077

[97] E. Tzagkaraki, S. Papadakis, and M. Kalogiannakis, "Exploring the use of educational Robotics in primary school and its possible place in the curricula," in *Education in & with Robotics to Foster 21st-Century Skills: Proceedings of EDUROBOTICS 2020*, Springer, 2021, 216–229, https://doi.org/10.1007/978-3-030-77022-8_19

[98] P. Juric, M. B. Bakaric, and M. Matetic, "Cognitive predispositions of students for STEM success and differences in solving problems in the computer game for learning mathematics.," *Int. J. Eng. Pedagog.*, vol. 11, no. 4, 81–95, 2021, https://doi.org/10.3991/ijep.v11i4.20587

[99] S. Papadakis, M. Kalogiannakis, and A. İ. C. Gözüm, "Editorial: STEM, STEAM, computational thinking, and coding: Evidence-based research and practice in children's development," *Front. Psychol.*, vol. 13, 2022, https://doi.org/10.3389/fpsyg.2022.1110476

[100] E. Shchedrina, E. Galkina, I. Petunina, and R. Lushkov, "Integration of mobile learning into complex problem-solving processes during STEM education," *Int. J. Interact. Mob. Technol.*, vol. 14, no. 21, 19–37, 2020, https://doi.org/10.3991/ijim.v14i21.18463

[101] S. Fajrina, L. Lufri, and Y. Ahda, "Science, technology, engineering, and mathematics (STEM) as a learning approach to improve 21st century skills: A review," *Int. J. online Biomed. Eng.*, vol. 16, no. 7, 95–104, 2020, https://doi.org/10.3991/ijoe.v16i07.14101

[102] I. Simonics, "Relationships among economy, industry, vocational education and training and higher engineering education," *Int. J. Eng. Pedagog.*, vol. 10, no. 5, 4–6, 2020, https://doi.org/10.3991/ijep.v10i5.16747

[103] Z. Mercan, S. Papadakis, A. İ. Can Gözüm, and M. Kalogiannakis, "Examination of STEM parent awareness in the transition from preschool to primary school," *Sustainability*, vol. 14, no. 21, 14030, 2022, https://doi.org/10.3390/su142114030

[104] M. Stebbins and T. Goris, "Evaluating STEM education in the U.S. Secondary schools: Pros and cons of the «project lead the way» platform," *Int. J. Eng. Pedagog.*, vol. 9, no. 1, 50–56, 2019, https://doi.org/10.3991/ijep.v9i1.9277

[105] A. İ. C. Gözüm, S. Papadakis, and M. Kalogiannakis, "Preschool teachers' STEM pedagogical content knowledge: A comparative study of teachers in Greece and Turkey," *Front. Psychol.*, vol. 13, 1–19, 2022, https://doi.org/10.3389/fpsyg.2022.996338

[106] S. Papadakis and M. Kalogiannakis, "Evaluating a course for teaching introductory programming with Scratch to pre-service kindergarten teachers," *Int. J. Technol. Enhanc. Learn.*, vol. 11, no. 3, 231–246, 2019, https://doi.org/10.1504/IJTEL.2019.100478

[107] I. Irwanto, A. D. Saputro, W. Widiyanti, M. F. Ramadhan, and I. R. Lukman, "Research trends in STEM education from 2011 to 2020: A systematic review of publications in selected journals," *Int. J. Interact. Mob. Technol.*, vol. 16, no. 5, 19–32, 2022, https://doi.org/10.3991/ijim.v16i05.27003

[108] T. L. Larkin, "Topic order in introductory physics and its impact on the STEM curricular ladder," *Int. J. Eng. Pedagog.*, vol. 7, no. 1, 136, 2017, https://doi.org/10.3991/ijep.v7i1.6528

[109] H. Hakeem, "Layered software patterns for data analysis in big data environment," *Int. J. Autom. Comput.*, vol. 14, no. 6, 650–660, 2017, https://doi.org/10.1007/s11633-016-1043-x

[110] R. N. Hafni, T. Herman, E. Nurlaelah, and L. Mustikasari, "The importance of science, technology, engineering, and mathematics (STEM) education to enhance students' critical thinking skill in facing the industry 4.0," *J. Phys. Conf. Ser.*, vol. 1521, no. 4, 0–7, 2020, https://doi.org/10.1088/1742-6596/1521/4/042040

[111] D. Gibson, "Big data in higher education: Research methods and analytics supporting the learning journey," *Technology, Knowledge and Learning*, vol. 22. Springer, 237–241, 2017, https://doi.org/10.1007/s10758-017-9331-2

[112] G. Xu, Y. Li, Y. Hu, and W. Zhou, "Learning analytics in a Web3D based inquiry learning environment," in *Artificial Intelligence in STEM Education*, CRC Press, 2022, 259–276, https://doi.org/10.1201/9781003181187-21

[113] C. Fischer *et al.*, "Mining big data in education: Affordances and challenges," *Rev. Res. Educ.*, vol. 44, no. 1, 130–160, 2020, https://doi.org/10.3102/0091732X20903304

[114] X. Chen, D. Zou, H. Xie, and F. L. Wang, "Past, present, and future of smart learning: A topic-based bibliometric analysis," *Int. J. Educ. Technol. High. Educ.*, vol. 18, no. 1, 2021, https://doi.org/10.1186/s41239-020-00239-6

[115] Y. Samuel, J. George, and J. Samuel, "Beyond stem, how can women engage big data, analytics, robotics and artificial intelligence? An exploratory analysis of confidence and educational factors in the emerging technology waves influencing the role of, and impact upon, women," *arXiv Prepr. arXiv2003.11746*, 2020, https://doi.org/10.2139/ssrn.3735279

[116] D. Weintrop *et al.*, "Defining computational thinking for mathematics and science classrooms," *J. Sci. Educ. Technol.*, vol. 25, 127–147, 2016, https://doi.org/10.1007/s10956-015-9581-5

[117] D. Zhao *et al.*, "An innovative multi-layer gamification framework for improved STEM learning experience," *IEEE Access*, vol. 10, 3879–3889, 2021, https://doi.org/10.1109/ACCESS.2021.3139729

[118] E. Al-Masri, S. Kabu, and P. Dixith, "Emerging hardware prototyping technologies as tools for learning," *IEEE Access*, vol. 8, 80207–80217, 2020, https://doi.org/10.1109/ACCESS.2020.2991014

[119] D. Gillborn, P. Warmington, and S. Demack, "QuantCrit: Education, policy, 'Big Data' and principles for a critical race theory of statistics," *Race Ethn. Educ.*, vol. 21, no. 2, 158–179, 2018, https://doi.org/10.1080/13613324.2017.1377417

[120] R. H. Ellaway, M. V. Pusic, R. M. Galbraith, and T. Cameron, "Developing the role of big data and analytics in health professional education," *Med. Teach.*, vol. 36, no. 3, 216–222, 2014, https://doi.org/10.3109/0142159X.2014.874553

[121] P. Gao, J. Li, and S. Liu, "An introduction to key technology in artificial intelligence and big data driven e-learning and e-education," *Mob. Networks Appl.*, vol. 26, no. 5, 2123–2126, 2021, https://doi.org/10.1007/s11036-021-01777-7

[122] A. Christopoulos, N. Pellas, and M. Laakso, "Education sciences a learning analytics theoretical framework for," *Education Science*, vol. 10, no. 11, 317, 2020.

[123] Z. Kohen and O. Nitzan, "Excellence in mathematics in secondary school and choosing and excelling in STEM professions over significant periods in life," *Int. J. Sci. Math. Educ.*, vol. 20, no. 1, 169–191, 2022, https://doi.org/10.1007/s10763-020-10138-x

[124] M. H. M. Adnan, S. A. Ariffin, H. F. Hanafi, M. S. Husain, and I. Y. Panessai, "A social media analytics framework to increase prospective students' interests in STEM and TVET Educationx," *Asian J. Univ. Educ.*, vol. 16, no. 4, 82–90, 2020, https://doi.org/10.24191/ajue.v16i4.11945

[125] C. J. Ballen and H. T. Holmegaard, "With big data comes big responsibilities for science equity research," *J. Microbiol. Biol. Educ.*, vol. 20, no. 1, 10–13, 2019, https://doi.org/10.1128/jmbe.v20i1.1643

7 Appendix

Table A1. Top 10 highest cited papers showed the contribution of Big Data in education and stem education

Author(s)	Citation; SJR	Recommendations
Daniel, B. [84]	317; 1.07 (Q1)	By increasing institutional transparency clearly and demonstrating the changes achieved by analytics. Analytics in Big Data also can help students and instructors provide insights to support student learning needs, thereby providing opportunities in education.
Gillborn, D., Warmington, P., Demack, S. [119]	188; 1.79 (Q1)	Analyzing big data can be done using Big Data technology. Although the analysis process is too complex for humans to understand, we can trust the results because Big Data is able to identify patterns in big data. By using Big Data, we can change the way we live, work and analyze our data.
Fischer, C., Pardos, ZA, Baker, RS, (...), Baker, R., Warschauer, M. [113]	110; 2.93 (Q1)	Given these limitations, Big Data can provide significant benefits if future research measures different types of big data or forms of quantitative or qualitative analysis. That way, the existence of Big Data will show an increase as well as the encouragement and potential benefits of Big Data mining in the world of education.
Ellaway, RH, Pusic, MV, Galbraith, RM, Cameron, T. [120]	106; 1.07 (Q1)	Therefore, it is important to take into account both the necessity to permit the use of Big Data techniques and the requirement to develop new method approaches to use Big Data for health professional education in the field of health professional education.

(Continued)

Table A1. Top 10 highest cited papers showed the contribution of Big Data in education and stem education *(Continued)*

Author(s)	Citation; SJR	Recommendations
Williamson, B [82]	90; 2.1 (Q1)	Applications, educational analytics, learning platforms, and data and visualization dashboards will facilitate the educational infrastructure.
Daniel, BK [83]	89; 1.07 (Q1)	By increasing the need for educational data sharing in addressing data security issues. And the need for the involvement of educators in collecting various forms of data and the need to explore and identify strategies that support educational researchers. Because the success of implementing Big Data in education depends on the skills and abilities of educational researchers and data science approaches.
Gupta, B., Goul, M., Dinter, B. [89]	83; 0.69 (Q2)	Business analytics that can use Big Data support organizational decision-making and advance curriculum initiatives in education.
Gao, P., Li, J., Liu, S. [121]	80; 0,82 (Q2)	To solve problems in information technology and mobile network-based education, it is necessary for scholars to pursue opportunities to research and pursue theoretical and technological studies in the fields of AI and Big Data-Based Learning and Distance Education. This will provide an opportunity for educators and scholars to publish their work and have real applications in education, especially in the field of e-learning. This system can also be used to produce concept maps and e-learning portals based on big data
Eynon, R. [87]	72; 2.05 (Q1)	To improve efficiency and competitiveness for evaluating the performance of schools and teachers, Big Data is increasingly being promoted in the form of "Technical Fixes" for educational research. So that it can help to understand various problems in technology learning and research.
Christopoulos, A., Pellas, N., Laakso, M.-J. [122]	24; 0.52 (Q2)	Big Data analysis is needed in compiling a Virtual Reality picture regarding the needs that students have in STEM Education
Hafni, RN, Herman, T., Nurlaelah, E., Mustikasari, L. [110]	8; 0.21 (Q4)	By using a STEM education approach, you can improve critical thinking skills to face challenges in the 4.0 industrial revolution, especially Big Data
Kohen, Z., Nitzan, O. [123]	1; 1.15 (Q1)	Big Data analysis can detect changing trends in selecting STEM as a major for study and career because STEM education currently lacks a workforce.
Adnan, MHM, Ariffin, SA, Hanafi, HF, Husain, MS, Panessai, IY[124]	1; 0.37 (Q3)	The need for Big Data analysis with framework components to increase student interest in STEM education
Ballen, CJ, Holmegaard, HT [125]	1; 0.42 (Q2)	The need for Big Data analysis in research to critically consider its limitations and biases.

8 Authors

Dr. Binar Kurnia Prahani, **M.Pd.** is a lecturer at the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya. Apart from teaching, he is also active in conducting research and community service. Several international publications indexed by Scopus so far there are 86 titles (H Index 11), 10 books with ISBNs and 23 copyrights that have been produced by the research team. He focus on field of digital learning, science-physics innovative learning, bibliometric analyze, and educational design research. (email: binarprahani@unesa.ac.id)

Irgy Redityo Dawana is a fourth-year undergraduate student in the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya.

Prof. Dr. Budi Jatmiko, M.Pd. is a Professor at the Department of Physics, Faculty of Mathematics and Natural Sciences, Universitas Negeri Surabaya. Active in writing articles in the journal reputable International (Scopus indexed), good as first and correspondence author has produced 54 articles, with H index: 11. He focus on field of curriculum development, digital learning, 21th century skills, innovative learning, thinking skills and educational design research.

Tan Amelia, S.Kom., M.MT. is a researcher and lecturer in the Department of Information System, Universitas Dinamika. Her research interests include software engineering with a particular interest in requirements prioritization. She graduated from the Universitas Dinamika with bachelor's degrees in Information systems and then completed her Master's degrees in Technology Management from the Institut Teknologi Sepuluh Nopember. She is currently pursuing a Ph.D. degree with the Faculty of Computing, Universiti Malaysia Pahang, Kuantan, Malaysia.

Article submitted 2023-01-29. Resubmitted 2023-03-07. Final acceptance 2023-03-07. Final version published as submitted by the authors.