



ARTICLE



<https://doi.org/10.1057/s41599-023-02176-x>

OPEN

# Education big data and learning analytics: a bibliometric analysis

Shaza Arissa Samsul <sup>1</sup>, Noraffandy Yahaya<sup>1</sup> & Hassan Abuhassna <sup>1</sup>

The contemporary era's extensive use of data, particularly in education, has provided new insights and benefits. This data is called 'education big data', and the process of learning through such data is called 'learning analytics'. Education in big data and learning analytics are two important processes that produce impactful results and understanding. It is crucial to take advantage of these processes to enhance the current education system. We conduct a bibliometric analysis based on the PRISMA statement template. The publications used for the analysis are based on the years 2012–2021. We examine and analyze a total of 250 publications, mainly sourced from the Scopus database, for insights regarding education big data and learning analytics. All of the publications also undergo filtration according to specific inclusion and exclusion criteria. Based on the bibliometric analysis conducted, we discover the distribution of education big data and learning analytics publications across the years 2012–2021, the most relevant journals and authors, the most significant countries, the primary research keywords, and the most important subject area involved. This study presents the trends and recommendations in education big data and learning analytics. We also offer suggestions for improvement and highlight the potential for enhancement of the education system through the full utilization of education big data and learning analytics.

<sup>1</sup>School of Education, Faculty of Social Sciences and Humanities, Universiti Teknologi Malaysia, 81310 UTM Johor Bahru, Johor, Malaysia.  
email: [sarissa2@graduate.utm.my](mailto:sarissa2@graduate.utm.my)

## Introduction

Big data in education has become a trend in recent years (Wang, 2016). The current era involves the creation and use of an enormous volume of data. Big data is a result of the inclusion of data in several industries, including banking, economics, and education. The transformation of technology into digital operations creates a massive digital treasure trove of data, especially in education (Michalik et al., 2014). Big data in education can be valuable and can be converted into insight using learning analytics. The possession of big data can definitely produce new knowledge and intuition in the education sector (Wang, 2016). The concept of big data is underpinned by the massive increase in the volume, structure, and speed with which data is generated (Daniel, 2017). Educators can analyze and improve the traditional educational system through the usage of big data (Drigas and Leliopoulos, 2014). The key accomplishment of learning analytics in recent years may be identified as the growth of digital learning, which has improved the quality and accessibility of educational data (Sghir et al., 2023).

Industry evolution 4.0 demands that higher education be upgraded in terms of programs and courses to prepare students for a highly computerized learning environment (Mkrttchian et al., 2021). Moreover, technological advancements in big data are unquestionably accelerators for boosting analytics in higher education (Mkrttchian et al., 2021). The structure of learning environments may be changed and enhanced as a result of input from learning analytics data (Talan and Demirbilek, 2023). Hence, education big data and learning analytics are major facilitators in the process of enhancing structure of learning environment. The objective of this study is to analyze the trends and recommendations of education big data and learning analytics publications using the bibliometric analysis method. This study also presents a visualization of the current trend in education big data and learning analytics in different topics. Data is taken from the Scopus database to answer the following research questions:

1. What is the distribution of education big data and learning analytics publications in the years 2012–2021?
2. What are the most relevant journals and authors in education big data and learning analytics research?
3. What are the most significant countries in the education big data and learning analytics research area?
4. What are the primary research keywords for education big data and learning analytics within the last decade?
5. What is the most important subject area involving education big data and learning analytics?

## Materials and methods

The bibliometric analysis and meta-analyses method was used in conducting this study's systematic literature review (SLR). The research process and procedure used in this study are based on the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) statement template. Based on comprehensive reporting through the PRISMA template, readers may evaluate the applicability of the methodologies and consequently, the veracity of the study's conclusion (Page et al., 2021). There are four processes involved in using the PRISMA template, namely identification, screening, and eligibility, and ultimately establishing which studies were included in the review. The details of this processes are explained further in Fig. 1.

The chosen topics that were used in this SLR were education big data and learning analytics. The Scopus database was utilized to source studies for review. As shown in Fig. 1, the SLR process involved using the PRISMA statement template for data selection.

The first step in the identification phase using the PRISMA statement template was to identify records in the database using “Education Big Data” and “Learning Analytics” keywords. The total number of documents found based on this search was 885.

These results then underwent a screening process, which left 252 documents remaining. The screening process excluded studies published in the year 2022 and some subject areas that were irrelevant to this study, such as business, management, and accounting. Conference papers, reviews, and editorial documents were also excluded from the analysis. The process continued with eligibility screening, which reduced the number of documents to 250 after removing full articles that were not in English. A software tool called VOSviewer was used to conduct data analysis and visualization. As explained by Soegoto et al. (2022), VOSviewer can effectively analyze and visualize bibliometric data analysis.

## Results

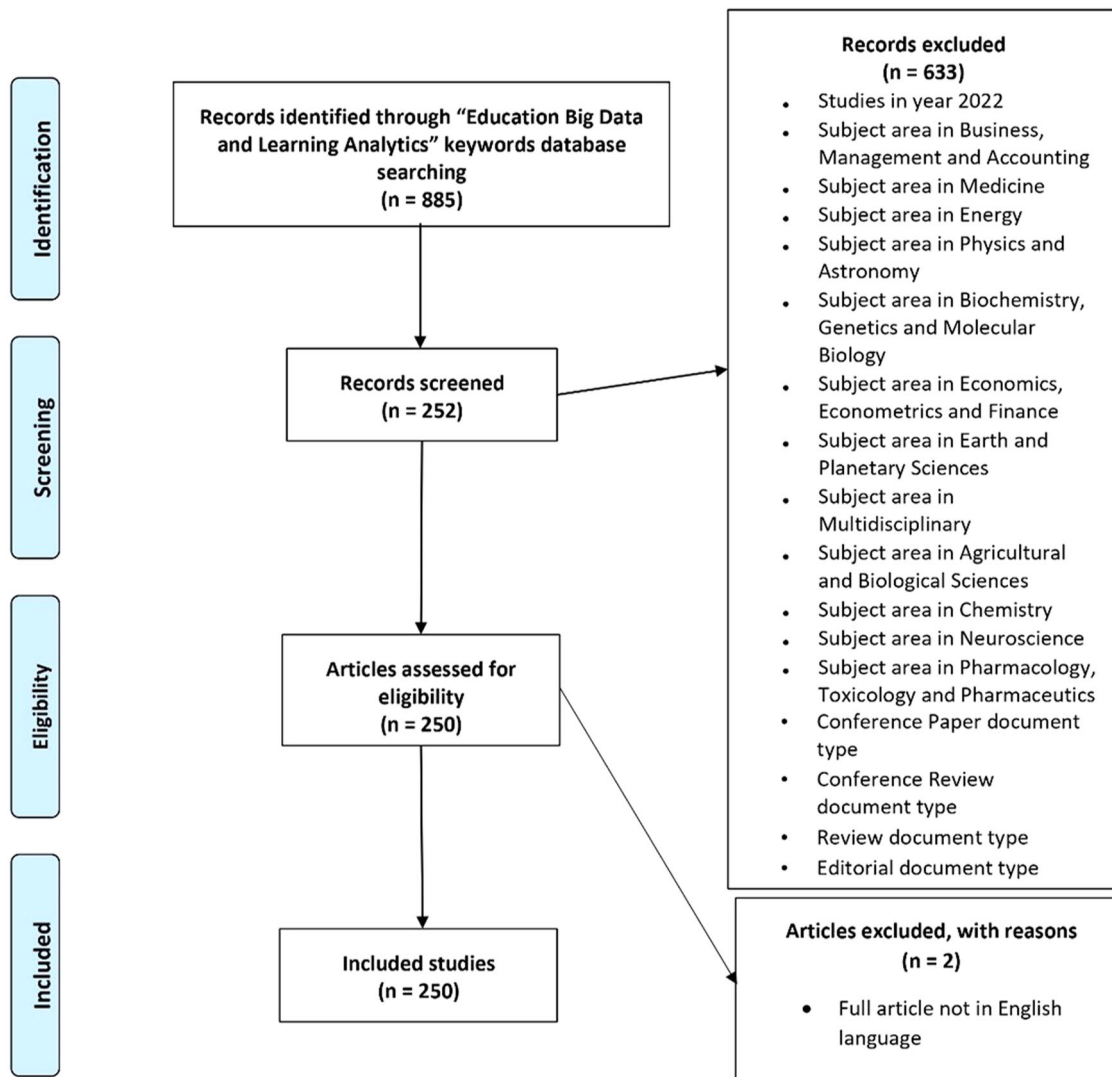
**Research question 1.** This study sought to examine education big data and learning analytics. The first finding answers the first research question, which is about the distribution of education big data and learning analytics publications between 2012 and 2021. As shown in Fig. 2, there was an increase in the number of documents produced during this 10-year period. Just one document on this subject was published in 2012, whereas 54 documents were produced in 2021. There was a decrement between 2017 and 2018, from 31 documents to 26, but this was followed by a further increase from 2019 onwards.

**Research question 2.** The second research question sought to identify the most relevant journals and authors in education big data and learning analytics research. Figure 3 shows the most relevant journals in education big data and learning analytics research in terms of total publications (TP). The Scopus database was then searched for the top ten frequently cited journals on this topic. The details about the journals, such as TP, Total Citations (TC), Citation Score, Most Cited Article, Times Cited, and Publisher Name, are shown in Table 1.

As shown in Table 1, the most relevant journal in education big data and learning analytics research, with a total of 8 publications and 63 citations, was “IEEE Access” published by IEEE. This journal was followed by “Lecture Notes in Educational Technology” published by Springer Nature, which had a total of 8 publications and 19 citations. The most cited article in this journal was “Big Data Learning Analytics: A new perspective”, which examined the significance of education big data and learning analytics. The “Educational Technology and Society” journal, with a total of 3 publications and 126 citations, was also highly relevant.

Research question two also determined the most productive authors in the area of education big data and learning analytics. The top fifteen authors were searched in the Scopus database. A list of data about the most productive authors in education big data and learning analytics research, based on TC, is presented in Fig. 4. A summary of the authors, including Author Name, Year of First Publication, TP, h-Index, TC, Current Affiliation, and Country, is illustrated in Table 2.

Table 2 provides a summary of the most productive authors in education big data and learning analytics research. According to the Scopus database, the most productive author was Ben Williamson from the University of Edinburgh, UK, whose first publication in this area was in 2007, and who, at the time of investigation, had a total of 60 publications, 1700 citations, an h-index of 25. The second most productive author was Hiroaki Ogata, from Japan, with a total of 371 publications, 3155 citations,



**Fig. 1** The systematic literature review process using the PRISMA statement template based on four phases that are identification, screening, eligibility and included.

and an h-index of 27, followed by Lynne D. Roberts from Australia, with 107 total publication, 1883 TC, and an h-index of 23. Of the top fifteen authors, Ryan Shaun Joazeiro de Baker from Columbia University, New York, United States, had the highest total number of citations, at 7752, and TP, at 278. A list of the most productive authors in education big data and learning analytics research is provided in Table 2.

**Research question 3.** The third research question sought to identify the countries with the most significant contributions to research on education big data and learning analytics. Figure 5 is a map that visualizes the most significant countries in this regard in terms of TP, according to the Scopus database. Table 3 presents summary of the most significant countries in the education big data and learning analytics research area. The criteria listed for the analysis are Rank, Country, TP, and Most Significant Academic Institution.

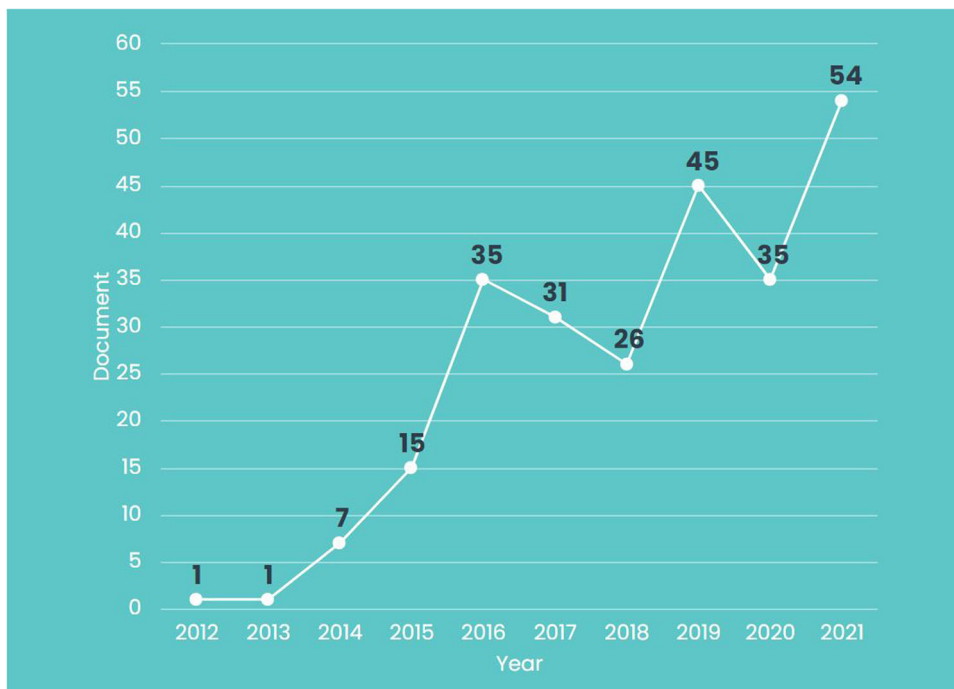
As shown in Table 3, the most significant country in the education big data and learning analytics research area was the United States, with a total of 59 publications, with City College of New York being the most significant research institution in this area. This was followed by the United Kingdom, with a total of 35 publications, with the University of Aberdeen being the most

significant institution. China was ranked third, with a total of 22 publications, and Capital University of Economics and Business was its most significant institution in this area. The other countries that were most productive in this research area are set out in Table 3.

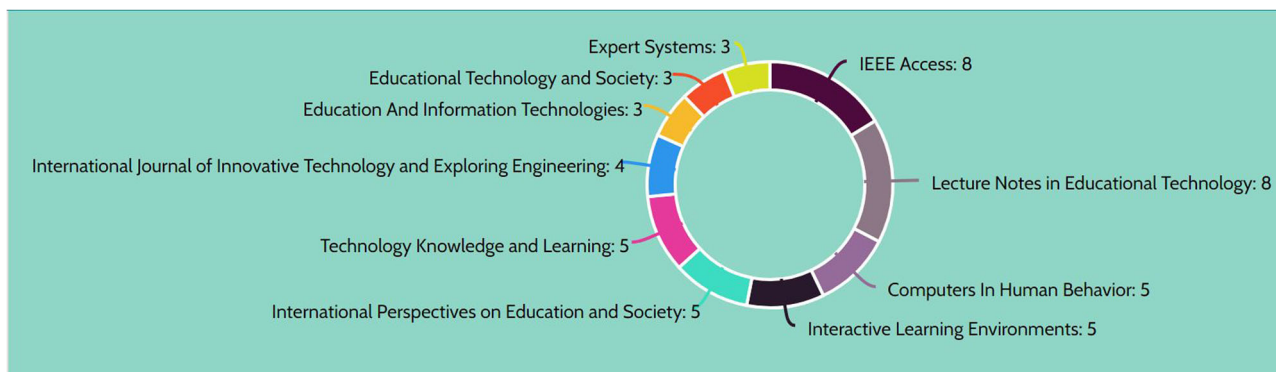
The study then measured the number of documents produced by each country, which are portrayed using a bar chart to provide a clearer view. As illustrated in Fig. 6, the highest number of documents produced was in the United States, followed by the United Kingdom, China, and India. Malaysia is also included in the top ten countries, being ranked ninth.

Next, this study examined the relationship of co-authorship with countries related to education big data and learning analytic research using VOSviewer software. The highest total number of link strengths of co-authorship and country was in the United States, with 14 links involving 59 documents and 1544 TC. As shown in Fig. 7, the country with the second highest link strength was the United Kingdom, which also had 14 links with other countries, involving 35 documents and 752 TC. The map also shows other countries' co-authorship relationships.

**Research question 4.** The fourth research question was about the primary research keywords used in education big data and



**Fig. 2** The line graph showing number of publications produced over 10 years. “Document” in y-axis is represented as number of publication and “Year” in x-axis represents the year observed.



**Fig. 3** The donut chart showing most relevant journals in education big data and learning analytics research according to total publications. The indicator (text: number), text used represents journal name and number represents number of total publications.

learning analytics research within the last decade. This study examined the co-occurrence of all keywords in the data associated with education big data and learning analytics. Figure 8 is a map based on the co-occurrence relationships of all keywords in the Scopus database. The keyword that had the highest co-occurrence (Oc) of 126 and link strength of 485 was “Big Data”, followed by “Learning Analytics” (Oc = 89). Other keywords with high co-occurrence included “Learning Systems” (Oc = 28), “Machine Learning” (Oc = 38), “Data Analytics” (Oc = 36), “Data Mining” (Oc = 30), and “Education” (Oc = 36).

The co-occurrence of author keywords was also analyzed, as mapped in Fig. 9. The keyword with the highest occurrence was “Big Data”, with 90 occurrences and 150 total links with other keywords, followed by “Learning Analytics” (Oc = 88). Other keywords, such as “Machine Learning” (Oc = 33), “Higher Education” (Oc = 32), “Data Analytics” (Oc = 14), and “Educational Data Mining” (Oc = 14), are also included in the map.

**Research Question 5.** To answer the fifth research question, which sought to identify the most important subject area

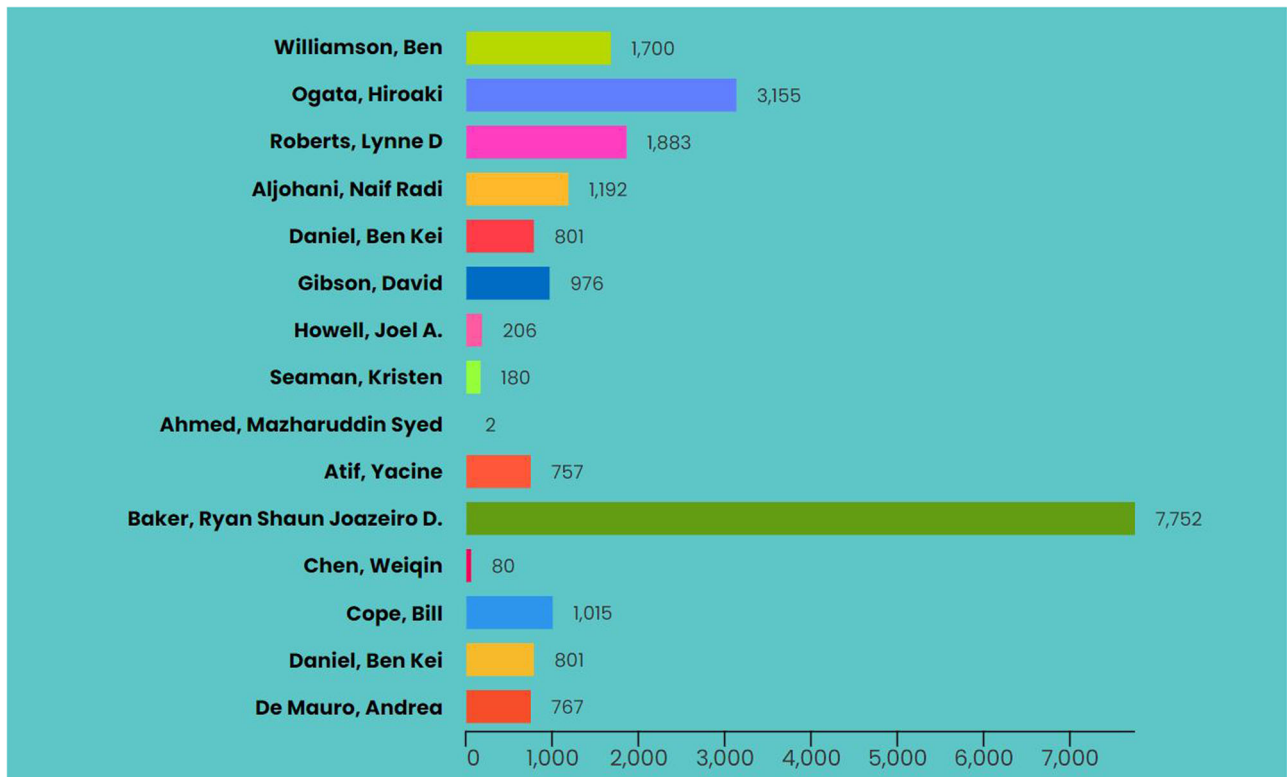
involving education big data and learning analytics, the data were analyzed according to the subject area. As demonstrated in Fig. 10, the highest percentage (34.6%), comprising 152 of the documents published, was in the Computer Science area. Computer Science thus appears to be the most relevant subject area regarding education big data and learning analytics. This was followed by 145 (33%) documents from the Social Sciences area. The fewest publications, at three documents, were in the Chemical Engineering area. The remaining data from several publications regarding education big data and learning analytics according to the subject area are presented in Fig. 10.

**Discussion**

Based on the result of an analysis of documents sourced from the Scopus database, all of the research questions have been answered in sequence. The distribution of publications on education big data and learning analytics between the years 2012 and 2021 shows an increasing pattern. This clearly shows that awareness about the importance of education big data and learning analytics is rising. Research by Şahin and Yurdugül (2020) supports this,

**Table 1 Summary of the most relevant journals in education big data and learning analytics research.**

Journal	TP	TC	Citation Score (2019 & 2020)	Most cited article (Reference)	Times Cited	Publisher
IEEE Access	8	63	4.8	Analyzing Learners' Behavior in MOOCs: An Examination of Performance and Motivation Using a Data-Driven Approach	21	IEEE
Lecture Notes in Educational Technology	8	19	1.0	Big Data Learning Analytics: A New Perspective	7	Springer Nature
Computers in Human Behavior	5	171	13.8	Predicting Academic Performance of Students from VLE Big Data Using Deep Learning ModelsOpen Access	70	Elsevier
Interactive Learning Environments	5	21	5.1	A Predictive Analytics Framework as a Countermeasure for Attrition of Students	9	Taylor & Francis
International Perspectives on Education and Society	5	1	1.3	Learning Analytics for Student Success at University: Trends and Dilemmas	1	Emerald
Technology Knowledge and Learning	5	122	4.0	Give Me a Customizable Dashboard: Personalized Learning Analytics Dashboards in Higher Education	46	Springer Nature
International Journal of Innovative Technology and Exploring Engineering	4	2	0.6	Big Data Analytics Model for the Education Sector	1	Blue Eyes Intelligence Engineering and Sciences Publication
Education and Information Technologies	3	31	5.4	Beyond Engagement Analytics: Which Online Mixed-Data Factors Predict Student Learning Outcomes?	30	Springer Nature
Educational Technology and Society	3	126	7.2	Applying Learning Analytics for the Early Prediction of Students' Academic Performance in Blended Learning	109	International Forum of Educational Technology and Society
Expert Systems	3	38	3.4	Intelligent Heart Disease Prediction in Cloud Environment Through Ensembling	32	Wiley-Blackwell



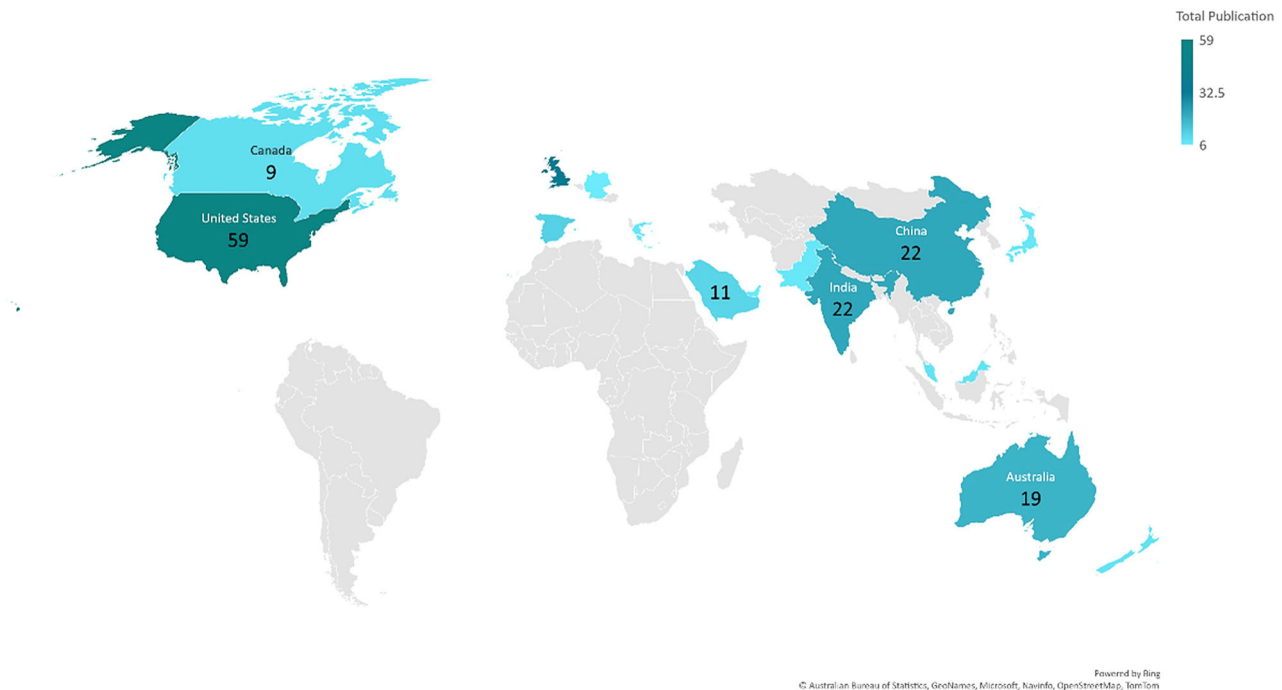
**Fig. 4 The bar graph showing most productive authors in education big data and learning analytics research according to total citation.** The y-axis represents name of author and x-axis represents number of total citation.



**Table 2 Summary of the most productive authors in education big data and learning analytics research.**

Author	Year of 1st publication	TP	h-index	TC	Current affiliation	Country
Williamson, Ben	2007	60	25	1700	The University of Edinburgh, Edinburgh, United Kingdom	United Kingdom
Ogata, Hiroaki	1994	371	27	3155	Kyoto University, Kyoto, Japan	Japan
Roberts, Lynne D.	1998	107	23	1883	Curtin University, Perth, Australia	Australia
Aljohani, Naif Radi	2011	108	19	1192	King Abdulaziz University, Jeddah, Saudi Arabia	Saudi Arabia
Daniel, Ben Kei	2002	67	13	801	University of Otago, Dunedin, New Zealand	New Zealand
Gibson, David	1996	71	14	976	Curtin University, Perth, Australia	Australia
Howell, Joel A.	2016	19	7	206	Curtin University, Perth, Australia	Australia
Seaman, Kristen	2016	8	6	180	Curtin University, Perth, Australia	Australia
Ahmed, Mazharuddin Syed	2016	3	1	2	Ara Institute of Canterbury, Canterbury, New Zealand	New Zealand
Atif, Yacine	1995	81	14	757	Högskolan i Skövde, Skovde, Sweden	Sweden
Baker, Ryan Shaun	2003	278	45	7752	Columbia University, New York, United States	United States
Joazeiro D.						
Chen, Weiqin	2014	34	5	80	OsloMet - storbyuniversitetet, Oslo, Norway	Norway
Cope, Bill	1993	88	17	1015	UNIVERSITY OF ILLINOIS, Urbana-Ghampaign, United States	United States
De Mauro, Andrea	2006	9	6	767	Università degli Studi di Roma "Tor Vergata", Rome, Italy	Italy

**The Most Significant Countries in Education Big Data and Learning Analytics**



**Fig. 5** The map chart for most significant countries in education big data and learning analytics according to total publications.

indicating that education big data and learning analytics are two significant fields that can improve the e-learning environment. Interest in education big data and learning analytics has escalated because of the possibilities of advancement in many sectors. Yu and Couldry (2022) also stated that digital platforms and learning analytics are proliferating in the education industry. This is supported by the increasing pattern in the distribution of education big data and learning analytics publications in the last decade.

Furthermore, the bibliometric analysis shows that the most relevant journals in education big data and learning analytics research were published by IEEE. The most frequently cited

article was about the usage of a data-driven approach for understanding learners' behavior. This article demonstrated that the most well-organized analytical method to make advancements in learning strategies is the learning analytics approach (Al-Shabandar et al., 2018). This study also discovered that decision-making and learning methods can be expanded using big data in education and learning analytics. Big data implementations can fully realize the potential of this personalized learning and improve teaching (Lutfiani and Meria, 2022). The article from the most prolific author in this analysis, Ben Williamson from the University of Edinburgh in the UK, discussed two important learning advancements that can be made using big

data, namely instructing machines and the use of computerized choice systems to influence human judgments (Knox et al., 2020).

In addition, the analysis found that the most significant country in the area of education big data and learning analytics research was the United States, with the highest number of publications. City College of New York was the most significant academic institution in the United States. One study from the United States agrees that major impacts on both educational practices can be seen by fully utilizing educational data mining and learning analytics (Baker and Inventado, 2014). Another study suggests that collaboration between the education system, industry players, and government entities in processing data analytics can facilitate the transition of technology to Industry 4.0 (Qin and Chiang, 2019). According to one of the most cited articles from the United States, employment and admittance screening, financial management, sponsorship tracking, and academic achievement evaluation are just a few of the administration and educational applications that might benefit from big data principles and data analytics (Picciano, 2012). Hence, numerous advancements and improvements can be

achieved with education big data and learning analytics, with collaboration from many sectors.

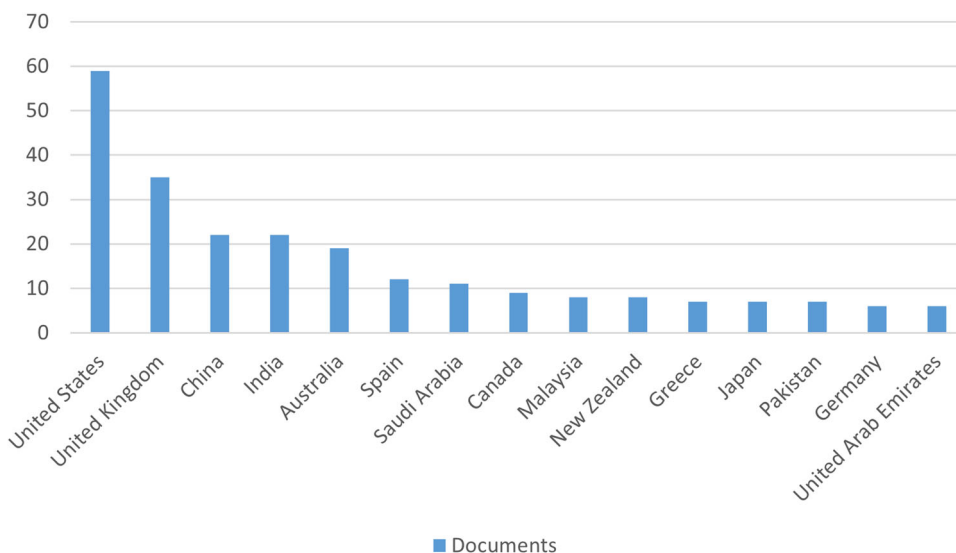
Moreover, the primary research keywords for education big data and learning analytics within the last decade were also analyzed in this study. Based on the results, “Big Data” and “Learning Analytics” were the most frequently used keywords for the education big data and learning analytics research areas. Both of these keywords also had high co-occurrence with all other keywords. Research about potential ways to optimize e-learning agrees that big data and learning analytics play a crucial role in the future of higher education (García and Secades, 2013). Learning analytics has made an important contribution to the education field in producing reliable projections of academic achievements based on assessments of the educational process (Tempelaar et al., 2021). Over the last 5 years, educational big data and learning analytics have been a growing focus on classifying and identifying students’ behavior (Lemay et al., 2021).

This study also identified the most important subject area involving education big data and learning analytics, namely computer science. Computer science, education, statistics, and other social sciences fields are brought together by educational data science to investigate and comprehend theoretical and practical phenomena (Daniel, 2016). Educational data mining also involves the combination of computer science, education, and statistics to better comprehend learning, administration processes, and research issues in higher education (Ray and Saeed, 2018).

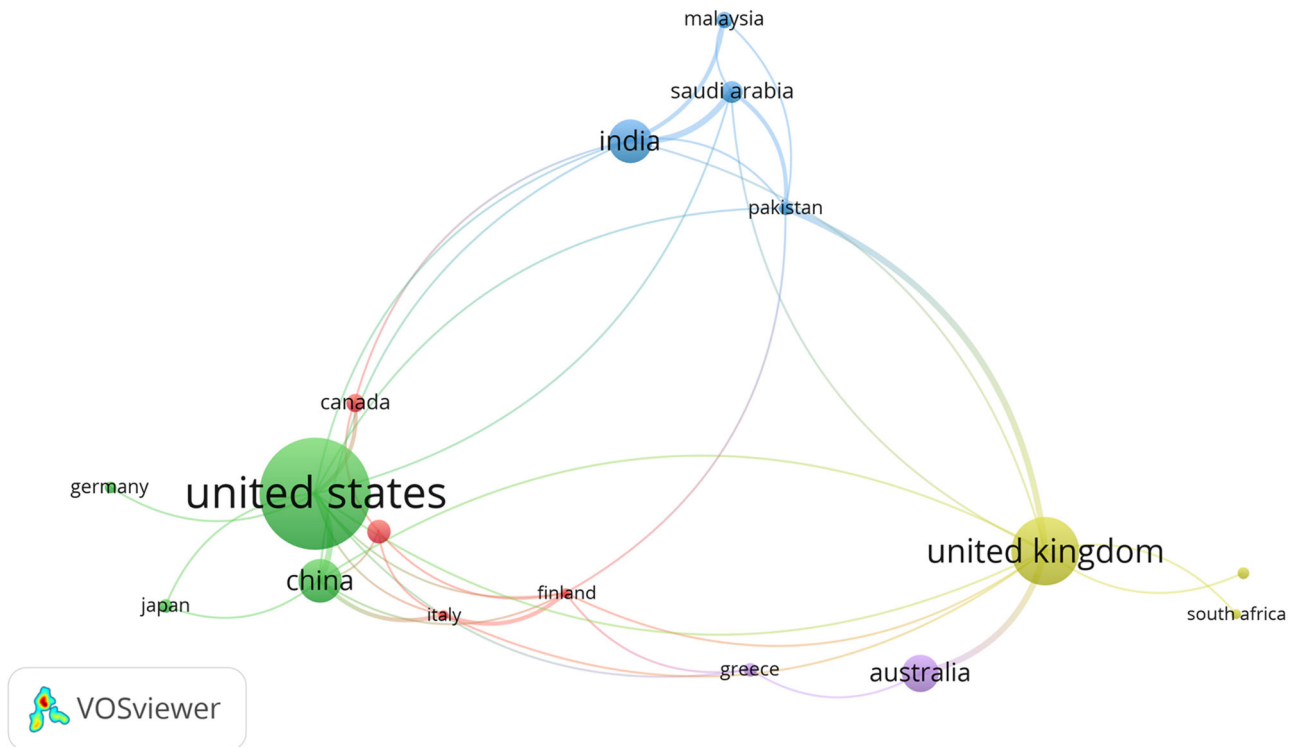
In summary, based on the present study’s analysis, the trends and recommendations of education big data and learning analytics include acting as a system for early detection that recognizes students who are at risk for academic failure or dropout, helping to provide dashboards for learning analytics, enabling amalgamation with Artificial Intelligence (AI) and machine learning, and providing future orientation in education. The increasing awareness of the importance of education big data and learning analytics is beneficial in enabling early detection of declining student performance based on the availability of big data sources. Dashboards and data visualization using learning analytics can greatly help to analyse complex data to produce insights for prevention and measures to be taken for specific problems. The simplification of data visualization from learning analytics is the future trend in the process of enabling amalgamation with AI and machine learning.

**Table 3 Summary of the most significant countries in the education big data and learning analytics research area.**

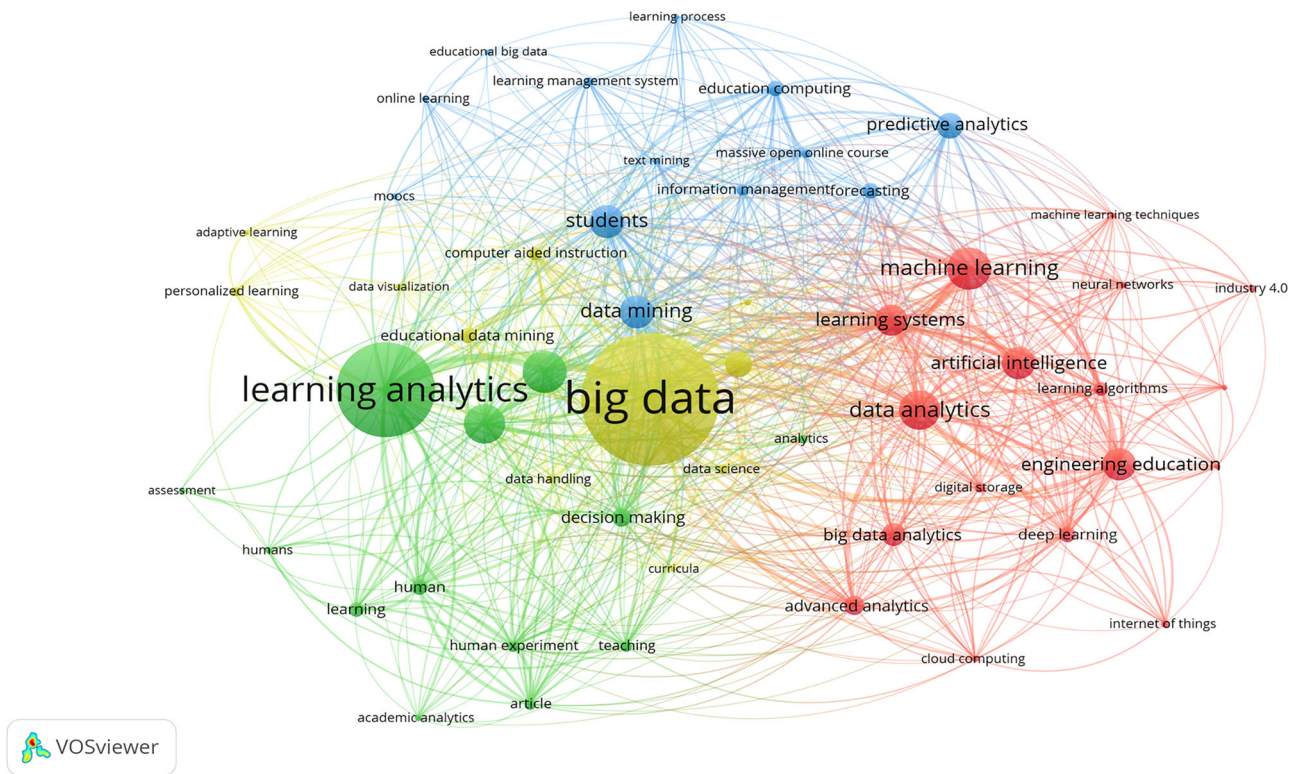
Rank	Country	TP	Most significant academic institution
1	United States	59	City College of New York
2	United Kingdom	35	University of Aberdeen
3	China	22	Capital University of Economics and Business
4	India	22	SRM University
5	Australia	19	University of South Australia
6	Spain	12	Universitat Ramon Llull
7	Saudi Arabia	11	King Abdul Aziz University
8	Canada	9	Carleton University
9	Malaysia	8	Universiti Teknologi Malaysia
10	New Zealand	8	Whitireia Community Polytechnic & Wellington Institute of Technology
11	Greece	7	University of Western Macedonia
12	Japan	7	Kyoto University
13	Pakistan	7	COMSATS University Islamabad (CUI)
14	Germany	6	Federal Police
15	United Arab Emirates	6	Skyline University College



**Fig. 6 The bar graph showing number of publications produced according to country.** The y-axis represents number of publications and x-axis represents country name.



**Fig. 7 A map based on the relationship of co-authorship with countries.** The line linking each country represents the relationship between co-authorship with other countries. The size of the circle shows number of publications.

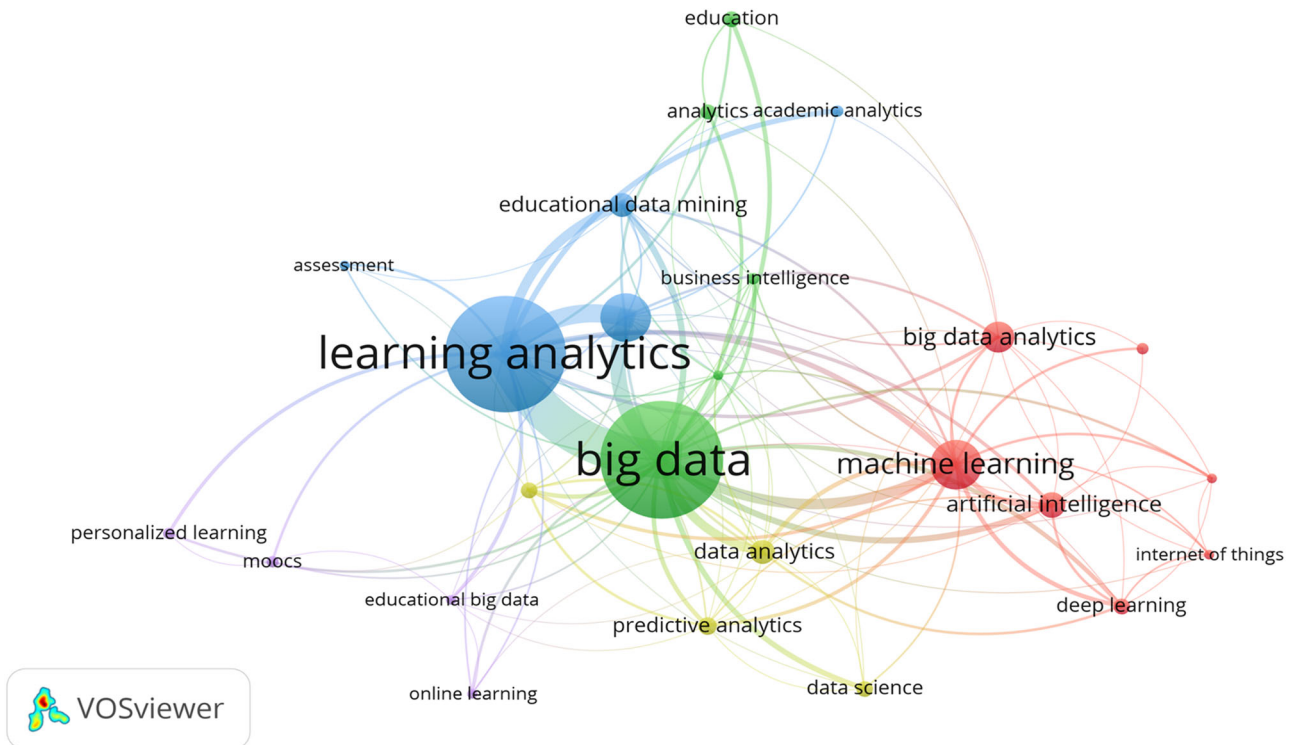


**Fig. 8 A map based on the relationship of co-occurrence with all keywords.** The line linking each keyword represents their co-occurrence with other keywords. The size of the circle shows number of occurrence.

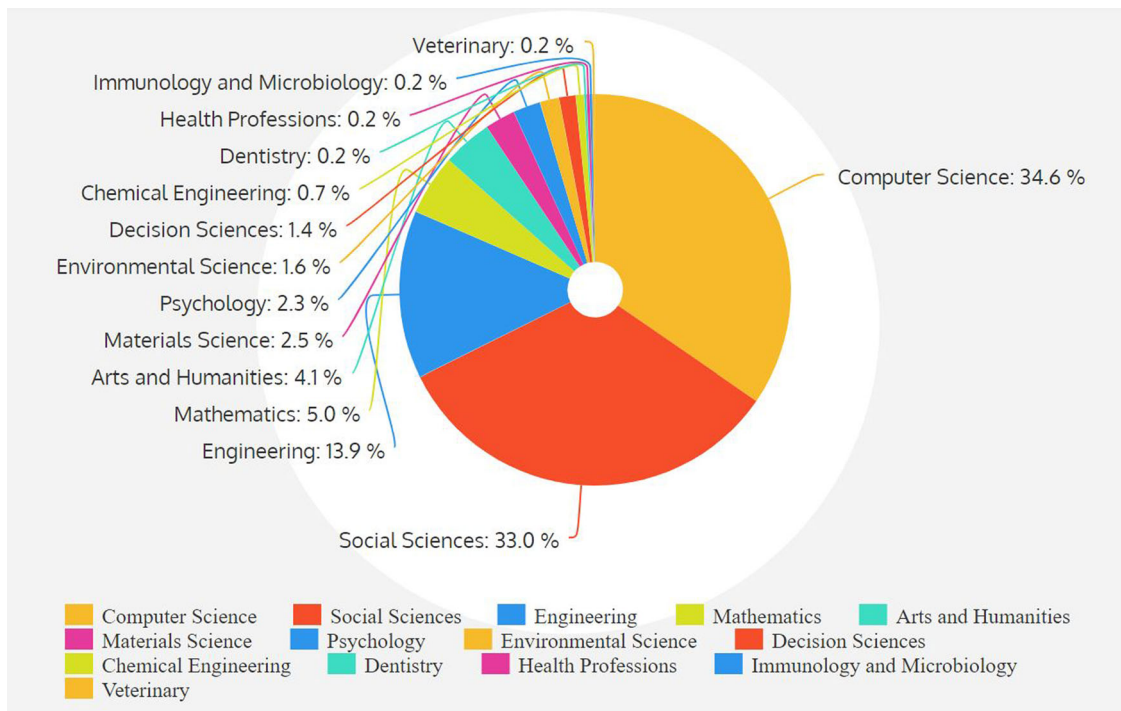
The integration of AI and machine learning in the education system could bring a huge impact. As an example, AI and machine learning can create new opportunities for automated evaluation

and personalized critiques, and enable smart teaching systems. All of these trends in fully utilizing big data and learning analytics are likely to continue and escalate over time.





**Fig. 9** A map based on the relationship of co-occurrence with author keywords. The line linking each keyword represents the relationship between their co-occurrence with other author keywords. The size of the circle shows number of occurrence.



**Fig. 10** The pie chart showing number of publications produced according to the subject area.

**Conclusion and implications**

The new era involves enormous amounts of data, which should be fully utilized for the advancement and enhancement of traditional systems in many sectors, especially in education. This study involved a bibliometric analysis of 250 publications regarding education big data and learning analytics. The

increasing number of publications involving education big data and learning analytics publications during the past 10 years, from 2012 until 2021, implies that the importance of both of these topics has been acknowledged. Most of the relevant journals were published by IEEE, which has made a significant contribution to the field of education big data and learning analytics. Moreover,

the most productive author to date is Ben Williamson, currently affiliated with the University of Edinburgh, UK, with a total of 1700 citations since his first publication in this area in 2007. Moreover, this study also found that the United States, with a total of 59 publications, is the most significant country in the area of education big data and learning analytics research, with its most prolific institution being City College of New York. The United States also had the highest total link strength of co-authorship on education big data and learning analytics. Since the United States is a big country, its exposure of the importance of using education big data and learning analytics worldwide is highly significant. “Big Data” and “Learning Analytics” were the keywords with the highest occurrence in most of the publications analyzed in this study. Most of the journals used these two keywords to explain details about education big data and learning analytics research. Computer Science and Social Sciences appear to be the most important subject areas regarding education big data and learning analytics. Both of these areas play important parts in making full use of the benefits of education big data and learning analytics. Other sectors were also involved, such as engineering, mathematics, art and humanities, psychology, and many more.

The main conclusion that can be drawn from this study is that big data and learning analytics are currently important skills to be maneuvered. Both big data and learning analytics could bring many significant benefits, such as improving the e-learning environment, understanding learners’ behavior, advancement in learning strategies, and many more. Learning techniques can be enhanced to achieve the best structured analytical strategy by employing learning analytics. This future trend of manipulating big data and learning analytics is certainly growing, especially in the field of education. In the future, new teaching and learning impacts gained from education big data and learning analytics could escalate the trend of personalized learning, predictive analytics, and adaptive learning, and enable data-driven decision-making. Throughout the upcoming years, education big data and learning analytics will remain significant in the field of education.

### Limitations

One of the limitations of this study is regarding limited information access, as it only used Scopus to identify publications to undergo bibliometric analysis. Other databases, such as Springer Link, IEEE Xplore Digital Library, or Web of Science, might have provided different insights and produced different results. In addition, the results of this research could also have been narrowed down and thus been more accurate if more specific keywords had been used in the analysis. The keywords used were “Education Big Data” and “Learning Analytics”. The depth of analysis could also have been increased if more keywords regarding education big data and learning analytics had been used, such as “Big Data Analytics”, “Educational Data Mining”, “Deep Learning,” and many more.

### Data availability

All data sets are available upon request.

Received: 10 October 2022; Accepted: 21 September 2023;

Published online: 16 October 2023

### References

- Al-Shabandar R, Hussain AJ, Liatsis P, Keight R (2018) Analyzing learners behavior in MOOCs: An examination of performance and motivation using a data-driven approach. *IEEE Access* 6:73669–73685
- Baker RS, Inventado PS (2014) *Educational Data Mining and Learning Analytics*. In: pp. 61–75. Springer, New York, NY. [https://doi.org/10.1007/978-1-4614-3305-7\\_4](https://doi.org/10.1007/978-1-4614-3305-7_4)

- Daniel BK (2016) *Big data and learning analytics in higher education: current theory and practice*. Springer International Publishing, Switzerland
- Daniel BK (2017) Big data in higher education: the big picture. In *Big data and learning analytics in higher education*, Springer, Cham, p 19–28
- Drigas AS, Leliopoulos P (2014) The use of big data in education. *Int J Comput Sci Issues* 11(5):58
- García OA, Secades VA (2013) Big Data & learning analytics: a potential way to optimize elearning technological tools. In: *International Association for Development of the Information Society International conference e-learning*
- Knox J, Williamson B, Bayne S (2020) Machine behaviourism: future visions of “learnification” and “datafication” across humans and digital technologies. *Learn Media Technol* 45(1):31–45
- Lemay DJ, Baek C, Doleck T (2021) Comparison of learning analytics and educational data mining: a topic modeling approach. *Comput Educ: Artif Intell* 2:100016
- Lutfiani N, Meria L (2022) Utilization of big data in educational technology research. *Int Trans Educ Technol* 1(1):73–83
- Michalik P, Štöfa J, Zolotova I (2014) Concept definition for Big Data architecture in the education system. Paper presented at the 12th International Symposium on Applied Machine Intelligence and Informatics (SAMII), IEEE, pp 331–334
- Mkrtrtchian V, Gamidullaeva L, Finogeev A, Chernyshenko S, Chernyshenko V, Amirov D et al. (2021) Big data and internet of things (IoT) technologies’ influence on higher education: current state and future prospects. *Int J Web-Based Learn Teach Technol* 16(5):137–157
- Page MJ, McKenzie JE, Bossuyt PM, Boutron I, Hoffmann TC, Mulrow CD et al. (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Int J Surg* 88:105906
- Picciano AG (2012) The evolution of big data and learning analytics in American higher education. *J Asynchronous Learn Netw* 16(3):9–20
- Qin SJ, Chiang LH (2019) Advances and opportunities in machine learning for process data analytics. *Comput Chem Engin* 126:465–473
- Ray S, Saeed M (2018) Applications of educational data mining and learning analytics tools in handling big data in higher education. In Alani M, Tawfik H, Saeed M, Anya O (eds) *Applications of big data analytics*. Springer, Cham
- Şahin M, Yurdugül H (2020) Educational data mining and learning analytics: past, present and future. *Bartın Univ J Fac Educ* 9(1):121–131
- Sghir N, Adadi A, Lahmer M (2023) Recent advances in predictive learning analytics: a decade systematic review (2012–2022). *Educ Inform Technol* 28(7):8299–8333
- Soegoto H, Soegoto ES, Luckyardi S, Rafidhi AA (2022) A bibliometric analysis of management bioenergy research using Vosviewer application. *Indones J Sci Technol* 7(1):89–104
- Talan T, Demirbilek M (2023) Bibliometric analysis of research on learning analytics based on web of science database. *Inform Educ* 22(1):161–181
- Tempelaar D, Rienties B, Nguyen Q (2021) The contribution of dispositional learning analytics to precision education. *Educ Technol Soc* 24(1):109–122
- Wang Y (2016) Big opportunities and big concerns of big data in education. *TechTrends* 60(4):381–384
- Yu J, Coudry N (2022) Education as a domain of natural data extraction: analysing corporate discourse about educational tracking. *Inform Commun Soc* 25(1):127–144

### Acknowledgements

This work was funded by the Ministry of Higher Education Malaysia under the Fundamental Research Grant Scheme (FRGS/1/2020/SSI0/UTM/02/8).

### Author contributions

All authors contributed to the research conception and design. The introduction and methodology were done by NY. The methodology and analysis were also planned and performed by HA. The full draft of the manuscript, including all parts was analyzed and written by SAS. All authors read and approved the final manuscript.

### Competing interests

The authors declare no competing interests.

### Ethical approval

This article does not contain any studies with human participants performed by any of the authors.

### Informed consent

This article does not contain any studies with human participants performed by any of the authors.

**Additional information**

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1057/s41599-023-02176-x>.

**Correspondence** and requests for materials should be addressed to Shaza Arissa Samsul.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023