

Received February 21, 2022, accepted March 10, 2022, date of publication March 17, 2022, date of current version March 24, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3160457

Educational Data Mining: A Bibliometric Analysis of an Emerging Field

CLARE BAEK^{ID1} AND TENZIN DOLECK^{ID2}

¹Rossier School of Education, University of Southern California, Los Angeles, CA 90007, USA

²Canada Research Chair, Simon Fraser University, Burnaby, BC V5A 1S6, Canada

Corresponding author: Tenzin Doleck (tdoleck@sfu.ca)

ABSTRACT We are now able to collect enormous amounts of information at the learner level. Mining educational data to provide data-driven analytics has spurred great interest among researchers and policymakers that continues to grow. This growing research area is called educational data mining (EDM). Yet the growing interest in the topic has also resulted in a fragmented body of literature. This recent growth justifies and renders it important to synthesize the extant body of multidisciplinary research to bring this literature together into a systematic whole and to assess the extent of our current knowledge. To this purpose, this article provides a bibliometric review of the accumulated literature ($N = 194$) on educational data mining during 2015–2019. Findings suggest that interest in educational data mining has increased in recent years. The studies in this stream of research mainly focus on using state-of-the-art EDM techniques to optimize prediction models to accurately predict learners' academic performance and to detect behaviors of learners for timely intervention. In addition, our findings show that EDM literature contains publications of researchers from diverse countries. Most studies were a result of collaborations between multiple authors, and most authors collaborated with authors from the same country. The United States, China, and Spain are the countries with the most prolific publications in EDM literature. For future research, EDM researchers should increase discussions on connecting theories with EDM techniques, ethics and privacy issues, and international collaboration.

INDEX TERMS Analytics, computational intelligence, data mining, data processing.

I. INTRODUCTION

Technology has fundamentally transformed how information is captured. Today, data is collected, analyzed, and interpreted to shape practices, processes, and decision-making in various fields [1]; and the field of education is no exception [2], [3]. Indeed, data logged by educational technologies is the subject of growing literature in recent years [4] because of the potential to drive educational innovation [5] and render obsolete some traditional ways to do education [6]. At the same time, tools to analyze educational data to gain and deliver educational insights have become readily available [7]. This has moved educators to rethink the way learning and teaching are considered.

On the research front, educational data has been used to: describe student life, predict behaviors and outcomes, inform the delivery of various forms of intervention, personalize

The associate editor coordinating the review of this manuscript and approving it for publication was John Mitchell^{ID}.

learning, customize teaching, among others [6], [8], [9]. Given the interest in how learner-system interaction data [10] can contribute to providing answers to salient educational questions, a specific stream of research has experienced growth in recent years: educational data mining (EDM). A cursory examination finds a strongly increasing number of publications and numerous educational analytics initiatives reflecting the relevancy of this research stream. It is important to note that although no clear-cut definition of educational data mining exists, we adopt the definition proffered by Baker and Yacef: "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings, which they learn in" [11, p. 4]. A substantial amount of research and practical information is now available apropos educational data mining [12]. A key premise of this literature is to promote data-driven innovations in education, with some suggesting that "most

EDM research is considered pedagogically and educational theory-neutral” [13, p. 3].

Although important progress in this field has been made, the state of knowledge remains relatively fragmented due in part to research on EDM being multidisciplinary in nature [14]. As such, this is an opportune time to take stock of the research in this stream of research to improve our understanding of the full range of relevant research and to better characterize the knowledge structure of this field. Moreover, such an effort is also key to unearthing outstanding research questions and needs. In this work, we conduct a bibliometric analysis [15] of the literature on educational data mining to consolidate and synthesize this stream of research. Bibliometric analysis, which is a “set of mathematical and statistical methods to display up-to-date and ongoing knowledge of a research field” [16, p. 3], provides a useful method to unearth the intellectual structure of a research field or topic. In fact, bibliometric analysis is “suitable for science mapping at a time when the emphasis on empirical contributions is producing voluminous, fragmented, and controversial research streams” [17, p. 959].

II. PURPOSE

The present study aims to achieve the following objective: provide an overview of the evolution of educational data mining research and describe the structures characterizing this emerging stream of research. This study explores the most influential sources that the EDM articles cited, the most cited EDM articles, the most frequently used keywords in the EDM articles, the productive countries in EDM research, and the international collaboration structure of the EDM articles. The frequently cited sources will show the extent of the interdisciplinary approach of the field and the leading focus of EDM research [18]. The most cited EDM articles will reveal the significant landmarks of EDM research and the most influential research findings, as well as emerging new ideas [18], [19]. The collaboration trend of EDM will help stakeholders in predicting the future collaboration trends of the field and promote deeper scientific collaborative research [20]. The prolific countries in EDM research will reveal the key players in EDM research as well as suggest ways to promote the spread and sharing of knowledge to all regions [20]. The keywords of the EDM articles will show the trends and themes of EDM literature [21]. In addition, this study aims to reveal the strengths of EDM research and the areas that EDM research can improve. Doing so will suggest the current research trend and future directions for EDM research [22]. We aim to answer the following research questions: 1. What are the prevalent sources cited by the EDM articles during 2015-2019? 2. What are the most cited EDM articles during 2015-2019? 3. What are the most frequently used keywords in the EDM articles published during 2015-2019? 4. What are the prolific countries in EDM literature during 2015-2019? 5. What are the collaboration patterns between countries in EDM literature during 2015-2019?

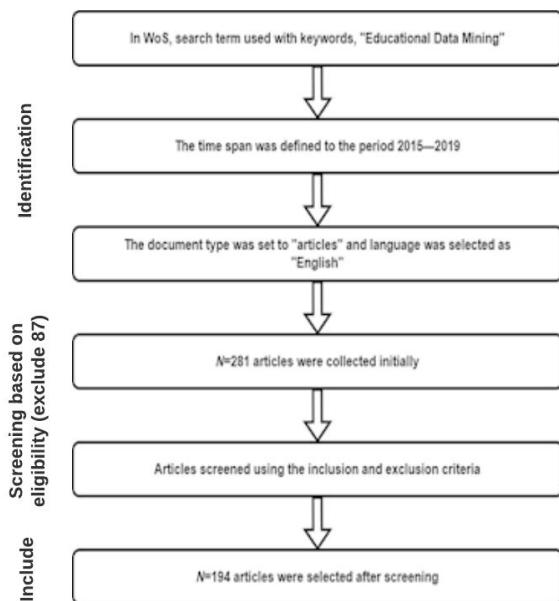


FIGURE 1. Overall screening procedure.

III. METHODOLOGY

A. DATA SOURCE AND RETRIEVAL

We conducted our search in the Web of Science (WoS), as this provides access to high-quality peer-reviewed articles. Moreover, WoS allows us to search across several disciplines. Importantly, WoS has been commonly used and recommended for literature reviews on education-related topics [23]. The data sourcing took place in several stages in November 2019 by following the PRISMA framework [24] (Figure 1).

In the first stage, the Web of Science database (Web of Science Core Collection) was searched by setting the filter to be 2015–2019. We chose this specific period since EDM enjoyed considerable growth during this period. To retrieve and collate relevant research, a search was performed on “educational data mining.” In the second stage, we performed a search on (TOPIC: (“Educational Data Mining”)) AND DOCUMENT TYPES: (Article)), which resulted in the initial hit of 281 sources. Similarly, we collected the titles and abstracts for all the sources. During the third stage, we sorted out the irrelevant articles by reviewing the abstracts following the inclusion and exclusion criteria. Our inclusion and exclusion criteria are presented in Table 1. The final yield was 194 articles.

B. DATA ANALYSIS

A total of 194 publications was included for final review. According to Donthu *et al.* “given the limitations of classical reviews involving large volumes of literature, bibliometrics is a preferred study method to more objectively assess academic progress” [25, p. 233]. There are numerous tools for conducting a bibliometric analysis. For our study, the final bibliometric data were exported and analyzed using the

TABLE 1. Inclusion and exclusion criteria.

	Inclusion	Exclusion
Timeliness	Published during the period 2015-2019	Published before 2015
Type of document	Scholarly journal articles	Not peer-reviewed
	Peer-reviewed	
Availability of Access Language	Full-text access available Written in English	Full-text not available Written in other languages than English
Relevance	Empirical papers (including pilot studies, preliminary findings, testing results with samples)	Theoretical papers, Literature review
	All grade level (e.g., k-12, higher education, adults)	Pertaining to other environments than learning
	Pertaining to learning environments (e.g., elementary school classroom, MOOC for higher education)	
	The primary purpose is exploring educational context (e.g., training teachers)	The primary purpose is not exploring educational context (e.g., developing machines not related to learning)

Bibliometrix R-package and Biblioshiny app [17], a free and open-source tool for science mapping.

IV. FINDINGS

A. MAIN INFORMATION ABOUT THE ARTICLES

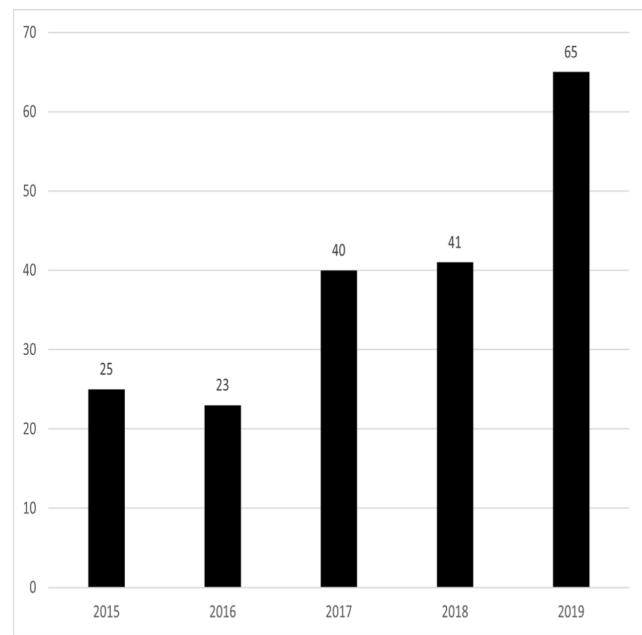
This section presents the study results. Table 2 provides a summary of the 194 articles. The 194 articles had 617 authors in total. Only 14 articles were published by a single author, which suggests that most articles were published by multiple authors. Indeed, the average number of co-authors per document is more than three authors (3.64). The average number of documents published per author is less than one. The number of author appearances is greater than the total number of authors, which reveals that some authors have multiple publications while most authors have one publication. Figure 2 shows that the number of publications has generally increased over the five-year period. The number of publications in 2019 is more than twice the number of publications in 2015.

B. SOURCES

This section investigates research question 1, the sources cited by the EDM articles published during 2015-2019 (Table 3). Computers & Education is cited more ($N = 336$) than all the other top 20 frequently cited sources. The aim and focus of all the frequently cited sources pertain to technology, computational methods, and data analysis. For example, Expert Systems with Applications is the second most cited source, and it is a journal that publishes papers

TABLE 2. Summary of bibliographic data.

	Description	Results
Main Information About Data		
Documents		194
Average years from publication		3.49
Average citations per documents		13.53
Average citations per year per doc		3.02
References		6590
Main Information About Data Document Contents		
Keywords Plus (ID)		259
Author's Keywords (DE)		570
Authors		
Authors		617
Author Appearances		706
Authors of single-authored documents		14
Authors of multi-authored documents		603
Authors Collaboration		
Single-authored documents		15
Documents per Author		0.314
Authors per Document		3.18
Co-Authors per Documents		3.64
Collaboration Index		3.37

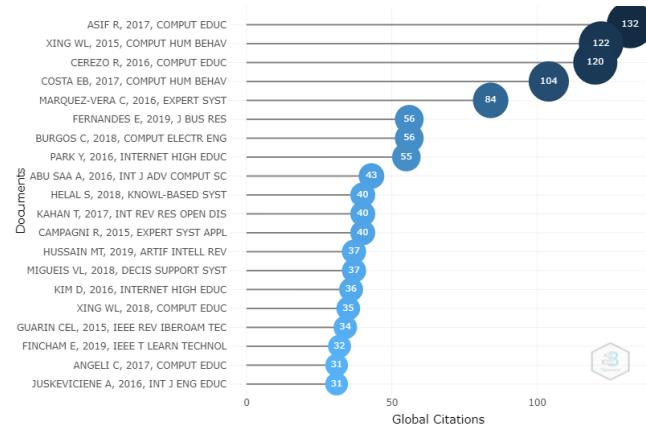
**FIGURE 2.** Publication distribution of educational data mining articles
(*Note: The collection of the articles took place in November 2019).

on intelligent systems applied in various settings including universities. Machine Learning is another frequently cited source which is a journal that publishes research on computational approaches to (machine) learning.

Multiple sources that were frequently cited by the EDM articles focus on research that connects pedagogy to technology and connects theories with computers. Specifically, *Computers & Education* focuses on the pedagogical use of digital technology and implementation of technology that is grounded in the context of use for teaching and learning. *Computers in Human Behavior* focuses on research

TABLE 3. Most local cited sources (from reference lists).

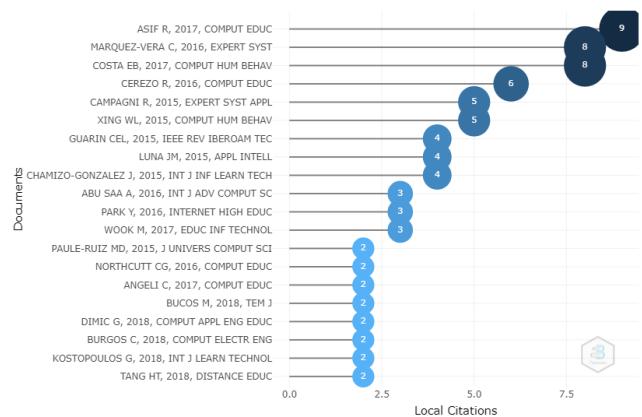
Sources	Articles
COMPUT EDUC	336
EXPERT SYST APPL	132
LECT NOTES COMPUT SC	119
COMPUT HUM BEHAV	95
J ED DATA MINING	70
IEEE T SYST MAN CY C	69
MACH LEARN	69
INTERNET HIGH EDUC	68
EDUC TECHNOL SOC	66
THESIS	59
USER MODEL USER-ADAP	56
BRIT J EDUC TECHNOL	54
LECT NOTES ARTIF INT	49
P 2 INT C LEARN AN K	46
WIRES DATA MIN KNOWL	36
IEEE T LEARN TECHNOL	33
MOR KAUF D	33
PROCEDIA COMPUT SCI	33
REV EDUC RES	32
ETR&D-EDUC TECH RES	31

**FIGURE 3.** Most global cited documents.

that explores the use and effects of computers from a psychological perspective. *Journal of Educational Data Mining* emphasizes research that integrates data mining with pedagogical theories.

C. DOCUMENTS

In this section, we explore the most global cited documents and the most local cited documents by the EDM articles during 2015-2019 to answer our research question 2. Most global cited documents (Figure 3) exhibit the number of citations an EDM article in our bibliometric collection received from articles in the entire database. Most local cited documents (Figure 4) exhibit the number of citations an EDM article in our bibliometric collection received from other EDM articles in our bibliometric collection. As shown in Figures 3 and 4, many of the most global cited and local cited documents overlap.

**FIGURE 4.** Most local cited documents.

All six of the top local and global cited articles are on prediction models (Table 4). Two studies (Asif *et al.* (2017) and Cerezo *et al.* (2016)) investigated the indicators or variables that well predict students' academic performance. Two studies (Costa *et al.* (2016) and Marguez *et al.* (2016)) compared different EDM techniques on effectively predicting students' failure early and found the techniques that outperformed other techniques. Another study (Campagni *et al.* (2015)) introduced how an EDM technique can be used to reveal the ideal academic paths that students can take for success. Another article (Xing *et al.* (2015)) brought a theoretical framework into the center of their research by building a model while being guided by Activity Theory and discussed the importance of connecting a theoretical framework in designing EDM models.

D. KEYWORDS

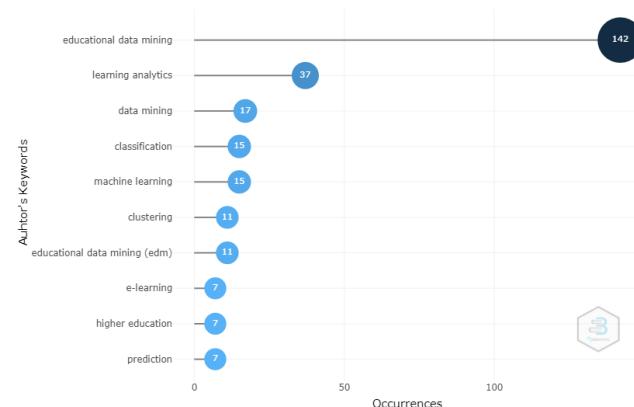
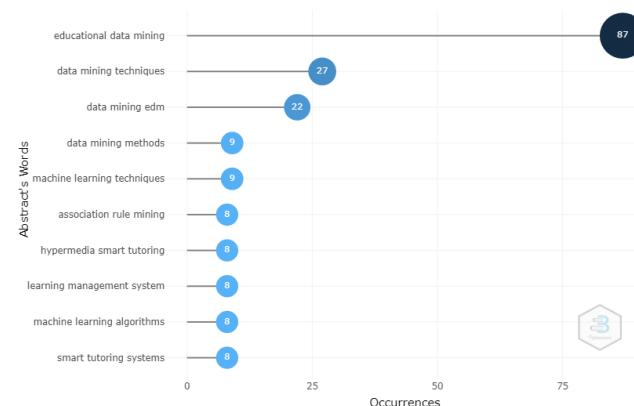
In this section, we explore the most frequently used words in the EDM articles during 2015-2019 to answer our research question 3. We first examine the most frequently used author's keywords. Then, we examine the most frequently occurring trigrams (i.e., three words that occur together) in the articles' abstracts and the trend of these trigrams over the period during 2015-2019. In doing so, we aim to reveal the research topics and changing trends of the research topics in EDM literature. Author's keywords are a list of terms that authors select to represent the overall content of the paper. Figure 5 shows the top ten keywords that were most frequently used during 2015-2019 in the EDM articles. The most frequently used keyword is "Educational Data Mining" as expected. The next frequently used keyword is "Learning Analytics," a field that is closely related to EDM. The EDM technique-related words "classification," "machine learning," "clustering," and "prediction" are also frequently used. These keywords are followed by words related to research platform or setting, "higher education" and "e-learning."

The analysis of trigrams in the abstracts also shows EDM literature's focus on techniques and the diverse applications

TABLE 4. Top cited papers.

Authors	Title	Findings	No. of Local Citations	No. of global Citations
Asif et al., (2017)	Analyzing undergraduate students' performance using educational data mining	A small number of courses can be good indicators for predicting university students' academic performance	9	132
Xing et al., (2015)	Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory	A student prediction model built by connecting a theoretical framework (i.e., Activity Theory) combined with an advanced modeling technique outperformed the traditional model.	5	122
Cerezo et al., (2016)	Students' LMS interaction patterns and their relationship with achievement: A case study in higher education	The time students spent on practical tasks in a LMS is highly related to students' achievement such that this group of students obtain higher marks.	6	120
Costa et al., (2016)	Evaluating the effectiveness of educational data mining techniques for early prediction of students' academic failure in introductory programming courses	The EDM techniques (i.e., Naive Bayes, Support Vector Machine) are effective in early identifying students' failure in an introductory programming course.	8	104
Marguez -Vera et al., (2016)	Early dropout prediction using data mining: a case study with high school students	The authors' proposed algorithm (i.e., ICRM2) outperformed all the traditional classification algorithms used to early predict high school students' dropout.	8	84
Campagni et al., (2015)	Data mining models for student careers	Clustering techniques applied to student database can reveal the ideal academic paths that students can take for successful outcomes.	5	40

of EDM techniques. Figure 6 shows that educational data mining technique-related terms are the most frequently occurring words in the EDM articles' abstracts: "data mining

**FIGURE 5.** Most frequent author's keywords.**FIGURE 6.** Most frequent trigrams in abstracts.

techniques," "data mining methods," "machine learning techniques," "association rule mining," "machine learning algorithms." There are several educational technology platforms in the frequently occurring words: "hypermedia smart tutoring," "learning management system," "smart tutoring systems." The number of times these trigrams occurred in EDM literature each year is shown in Figure 7. The most frequently occurring trigrams show the field's continual focus on data mining techniques over the years as shown by the steady growth of "data mining techniques." All the other words on specific techniques or educational technology platforms have relatively low occurrences and do not show any significant pattern of growth or decline.

E. COUNTRIES

This section explores research questions 4 and 5 about the presence of different countries in EDM literature and the structure of international collaboration. First, we examine the most relevant affiliations in EDM literature, which shows the institutions of the authors with at least one article in our collection. Figure 8 shows that Tel Aviv University in Israel has the most articles, followed by San Diego State University in the United States ($N = 8$). Beijing Normal University in China, University of Carlos III Madrid in Spain, and

Word Growth

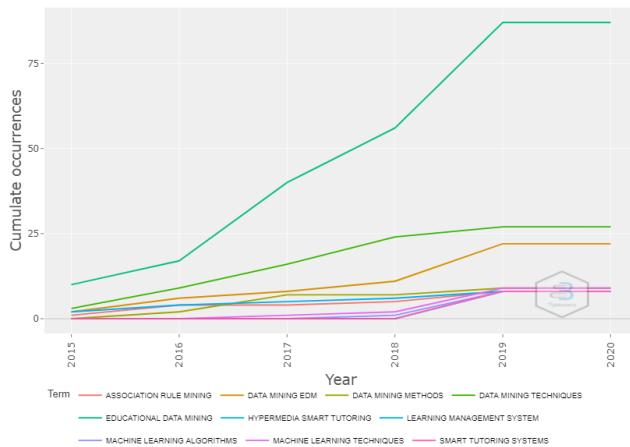


FIGURE 7. Word growth of most frequent trigrams in abstracts (Note. The year 2020 is included since some articles were accessible online earlier but published in a volume/issue in 2020).

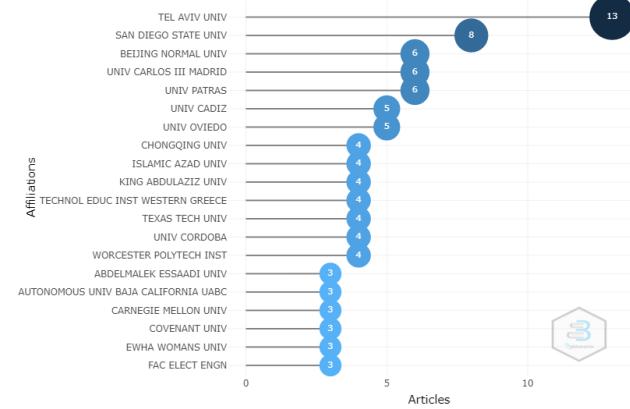


FIGURE 8. Most relevant affiliations.

University of Patras in Greece are the next relevant affiliations with six articles each.

The top 20 countries with the authors who received the greatest number of citations represent countries from a variety of geographic locations (Figure 9). The top 20 countries with the most citations consist of diverse continents including North America, South America, Asia, the Middle East, Australia, Africa, and Europe. The United States and Spain have the highest number of citations with at least more than twice as many citations as the other countries.

The next step is to examine how authors from different countries collaborate with each other. Figure 10 shows the top 20 countries with the most articles published and the patterns of collaboration between countries. The top 10 countries consist of countries from varieties of continents including the Middle East, North America, Asia, South America, and Europe. Authors from the United States have notably more publications than other countries with about 20 more publications than the countries with the second

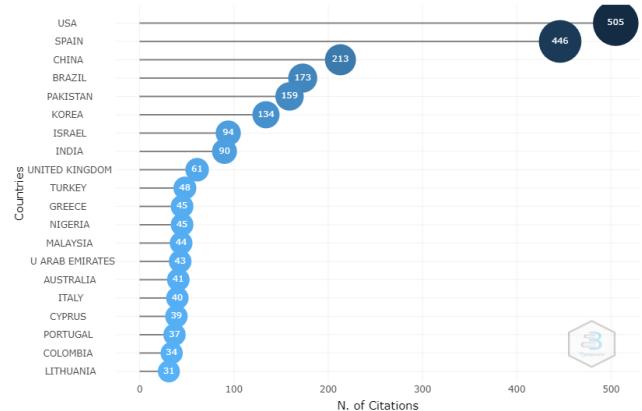


FIGURE 9. Most cited countries.

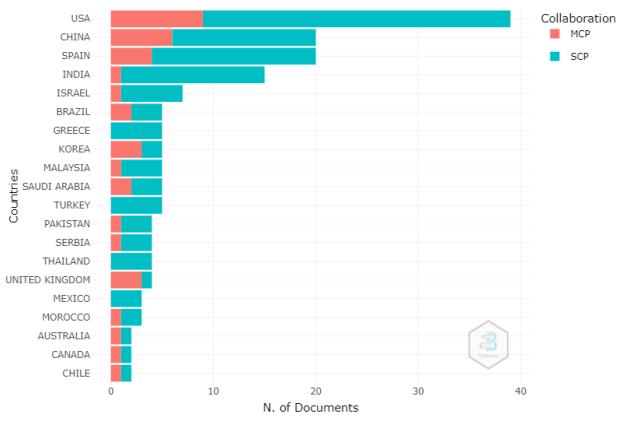


FIGURE 10. Corresponding authors country.

most publications (i.e., China and Spain). Figure 10 also depicts international collaboration. MCP, multiple countries publications, indicates the number of articles with at least one co-author from a different country. SCP, single country publications, indicates the number of articles with all authors from the same country. Korea and the United Kingdom have more MCP than SCP. On the other hand, the top five countries with the most articles have more SCP than MCP.

V. DISCUSSION

Our findings show that the volume of publications has increased over the years with the most recent year, 2019, having at least twice as many publications as the previous four years. This increasing trend of the number of publications reflects the increase in attention paid to EDM research. This growth and evolution can be due to new emerging technologies and techniques developed as well as researchers from other disciplines joining EDM-related research. Indeed, over the several years leading up to 2019, more free datasets and tools have become readily available and the interest in new applications and educational environments has increased [10].

Our analysis of the sources that the EDM articles cited imply several trends of the field's research. First, EDM research appears to be relying heavily on one particular source, *Computers & Education*, as the number of citations this journal received is more than twice as many citations as the second most cited sources. This type of citation structure resembles the Pareto principle of power law where a few sources garner the majority of the citations [19]. Future EDM research should consider referencing from more diverse sources as this can mitigate the potential bias in perspectives and enrich the discourse.

Another trend of the source structure implies EDM research's core focus on developing and improving techniques to analyze data using innovative technologies as well as EDM research's involvement of pedagogy and learning contexts. Some of the most frequently cited sources focus on data mining techniques (e.g., *Machine Learning*) whereas some other sources also focus on publishing research that connects pedagogy and theories to technology (e.g., *Computers in Human Behavior*). The trend of EDM research on examining innovative techniques and connecting pedagogy to technology is reflected in our analysis of the most cited documents as well. The top six of the most cited EDM articles in our collection are all on optimizing prediction models through new methods and techniques. Particularly, these studies introduced new approaches to improve the accuracy of predicting learners' performance from the extant models.

It is notable that the second most cited article (i.e., Xing *et al.* (2015)) was on building a prediction model that outperforms the traditional models by integrating a theoretical framework with an advanced modeling technique. The frequent citations of an article that addresses theories align with the aforementioned trend of the source structure in which the EDM articles cited sources that focus on research connecting pedagogy and theories to technology. One of the key and unique strengths of EDM research is its potential to refine, test, and develop educational theories via empirical evidence provided by EDM methods which would then lead to improved learning and teaching outcomes [11], [26]. Thus, the prominent presence of theories in citing sources and articles exhibits EDM research's active engagement with involving theories to connect to EDM applications.

On the contrary, our analysis of keywords implies that the application of theories into the actual practice of EDM research has been limited. Our analyses of the author's keywords, trigrams in abstracts, and the growth of these trigrams over the 2015–2019 period collectively show that discussions of theories are limited in the EDM articles. The frequently used keywords all pertain to data, methods, techniques, and technology platforms. Classification, clustering, association rule mining, and machine learning algorithms are the techniques that frequently occurred in EDM literature. Learning management system and smart tutoring system are the two educational technology platforms that were studied frequently. The nonexistence of theory-related key terms can suggest the limited engagement with theories in actual

research. This limitation of the application of theories into EDM research aligns with the findings of previous studies that although EDM research has developed sophisticated methods for analyses, its impact on practice and theory has been limited [10], [27]. EDM research that is framed without educational theories would not lead to the fulfillment of one of the main goals of EDM, which is to improve learning and teaching [26]. Thus, EDM researchers should ensure that their studies are framed around theories and that the power of EDM techniques is integrated with theories to yield practical outcomes.

Our analysis of most relevant affiliations, most cited countries, and the authors' country of the EDM articles depict an eclectic EDM research community with authors from diverse locations. Countries and institutions from WEIRD countries (western, educated, industrialized, rich, and demographic countries) as well as from non-WEIRD countries are actively engaged in EDM research. The trend of EDM research that represents diverse countries is encouraging and corroborates how EDM research can serve as powerful tools that accommodate diverse learners. Furthermore, EDM research can serve as powerful tools that could serve as supplemental resources which would provide personalized learning to underrepresented communities and developing countries [2]. Providing personalized learning via innovative technologies to students from diverse communities can promote inclusive, equitable, and quality education for all, as one of the sustainable development goals of the United Nations aspires [28]. To continue to promote the inclusive research community, EDM researchers should expand international collaboration in the future. Especially, researchers from the leading countries in publishing EDM articles such as the United States should extend their collaboration beyond other authors in the same country.

This study has some limitations. This study only includes the EDM articles published during 2015–2019. Also, since we used the Web of Science database to collect the articles, the limitations to this database may apply to this study as well. This study should be replicated in the future with articles published in more recent years. It would be interesting to compare this study with the bibliometric results with articles published in more recent years to examine whether there has been a difference in the trend of the field. Also, this study should be replicated by collecting EDM studies from different databases to examine if it yields different results.

VI. CONCLUSION

This bibliometric study revealed the recent trend of the evolving field, Educational Data Mining. As our analyses show, EDM research has been increasingly active in using EDM techniques to yield new methods to improve educational outcomes, particularly in building prediction models to detect students' performances. Our analyses imply that although the EDM articles have been citing references related to theories, theories have not been integrated into the center of research. Also, our analyses imply a diverse

and inclusive community of EDM researchers but relatively limited international collaborations.

REFERENCES

- [1] A. K. Kar and Y. K. Dwivedi, "Theory building with big data-driven research—Moving away from the 'what' towards the 'why,'" *Int. J. Inf. Manage.*, vol. 54, Oct. 2020, Art. no. 102205, doi: [10.1016/j.ijinfomgt.2020.102205](https://doi.org/10.1016/j.ijinfomgt.2020.102205).
- [2] C. Baek and T. Doleck, "A bibliometric analysis of the papers published in the journal of artificial intelligence in education from 2015–2019," *Int. J. Learn. Anal. Artif. Intell. Educ.*, vol. 2, no. 1, p. 67, May 2020, doi: [10.3991/ijai.v2i1.14481](https://doi.org/10.3991/ijai.v2i1.14481).
- [3] A. Charitopoulos, M. Rangoussi, and D. Koulouriotis, "On the use of soft computing methods in educational data mining and learning analytics research: A review of years 2010–2018," *Int. J. Artif. Intell. Educ.*, vol. 30, no. 3, pp. 371–430, Oct. 2020, doi: [10.1007/s40593-020-00200-8](https://doi.org/10.1007/s40593-020-00200-8).
- [4] K. L.-M. Ang, F. L. Ge, and K. P. Seng, "Big educational data & analytics: Survey, architecture and challenges," *IEEE Access*, vol. 8, pp. 116392–116414, 2020, doi: [10.1109/ACCESS.2020.2994561](https://doi.org/10.1109/ACCESS.2020.2994561).
- [5] T. Doleck, D. J. Lemay, R. B. Basnet, and P. Bazelaïs, "Predictive analytics in education: A comparison of deep learning frameworks," *Educ. Inf. Technol.*, vol. 25, no. 3, pp. 1951–1963, May 2020, doi: [10.1007/s10639-019-10068-4](https://doi.org/10.1007/s10639-019-10068-4).
- [6] L. Chen, P. Chen, and Z. Lin, "Artificial intelligence in education: A review," *IEEE Access*, vol. 8, pp. 75264–75278, 2020, doi: [10.1109/ACCESS.2020.2988510](https://doi.org/10.1109/ACCESS.2020.2988510).
- [7] S. Slater, S. Joksimović, V. Kovanovic, R. S. Baker, and D. Gasevic, "Tools for educational data mining," *J. Educ. Behav. Statist.*, vol. 42, no. 1, pp. 85–106, Feb. 2017, doi: [10.3102/1076998616666808](https://doi.org/10.3102/1076998616666808).
- [8] K. M. L. Jones, A. Rubel, and E. LeClere, "A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics," *J. Assoc. Inf. Sci. Technol.*, vol. 71, no. 10, pp. 1227–1241, Oct. 2020, doi: [10.1002/asi.24327](https://doi.org/10.1002/asi.24327).
- [9] B. T.-M. Wong and K. C. Li, "A review of learning analytics intervention in higher education (2011–2018)," *J. Comput. Educ.*, vol. 7, no. 1, pp. 7–28, Mar. 2020, doi: [10.1007/s40692-019-00143-7](https://doi.org/10.1007/s40692-019-00143-7).
- [10] C. Romero and S. Ventura, "Educational data mining and learning analytics: An updated survey," *WIREs Data Mining Knowl. Discovery*, vol. 10, no. 3, p. e1355, May 2020, doi: [10.1002/widm.1355](https://doi.org/10.1002/widm.1355).
- [11] R. S. J. D. Baker and K. Yacef, "The state of educational data mining in 2009: A review and future visions," *J. Educ. Data Mining*, vol. 1, no. 1, pp. 3–17, 2009.
- [12] D. J. Lemay, C. Baek, and T. Doleck, "Comparison of learning analytics and educational data mining: A topic modeling approach," *Comput. Educ., Artif. Intell.*, vol. 2, Jan. 2021, Art. no. 100016, doi: [10.1016/j.caai.2021.100016](https://doi.org/10.1016/j.caai.2021.100016).
- [13] B. Rienties, H. K. Simonsen, and C. Herodotou, "Defining the boundaries between artificial intelligence in education, computer-supported collaborative learning, educational data mining, and learning analytics: A need for coherence," *Frontiers Educ.*, vol. 5, p. 128, Jul. 2020, doi: [10.3389/feduc.2020.00128](https://doi.org/10.3389/feduc.2020.00128).
- [14] C. Baek and T. Doleck, "Educational data mining versus learning analytics: A review of publications from 2015 to 2019," *Interact. Learn. Environments*, pp. 1–23, Jun. 2021, doi: [10.1080/10494820.2021.1943689](https://doi.org/10.1080/10494820.2021.1943689).
- [15] J. Wallin, "Bibliometric methods: Pitfalls and possibilities," *Basic Clin. Pharmacol. Toxicol.*, vol. 97, no. 5, pp. 261–275, 2005, doi: [10.1111/j.1742-7843.2005.pto_139.x](https://doi.org/10.1111/j.1742-7843.2005.pto_139.x).
- [16] J. D. Ampah, A. A. Yusuf, S. Afrane, C. Jin, and H. Liu, "Reviewing two decades of cleaner alternative marine fuels: Towards IMO's decarbonization of the maritime transport sector," *J. Cleaner Prod.*, vol. 320, Oct. 2021, Art. no. 128871, doi: [10.1016/j.jclepro.2021.128871](https://doi.org/10.1016/j.jclepro.2021.128871).
- [17] M. Aria and C. Cuccurullo, "bibliometrix: An R-tool for comprehensive science mapping analysis," *J. Informetrics*, vol. 11, no. 4, pp. 959–975, Nov. 2017, doi: [10.1016/j.joi.2017.08.007](https://doi.org/10.1016/j.joi.2017.08.007).
- [18] B. Peritz and J. Bar-Ilan, "The sources used by bibliometrics-scientometrics as reflected in references," *Scientometrics*, vol. 54, no. 2, pp. 269–284, 2002, doi: [10.1023/a:1016018013096](https://doi.org/10.1023/a:1016018013096).
- [19] V. Garousi and M. V. Mäntylä, "Citations, research topics and active countries in software engineering: A bibliometrics study," *Comput. Sci. Rev.*, vol. 19, pp. 56–77, Feb. 2016, doi: [10.1016/j.cosrev.2015.12.002](https://doi.org/10.1016/j.cosrev.2015.12.002).
- [20] H. Hu, D. Wang, and S. Deng, "Global collaboration in artificial intelligence: Bibliometrics and network analysis from 1985 to 2019," *J. Data Inf. Sci.*, vol. 5, no. 4, pp. 86–115, Nov. 2020, doi: [10.2478/jdis-2020-0027](https://doi.org/10.2478/jdis-2020-0027).
- [21] M. Farrukh, I. A. Shahzad, F. Meng, Y. Wu, and A. Raza, "Three decades of research in the technology analysis & strategic management: A bibliometrics analysis," *Technol. Anal. Strategic Manage.*, vol. 33, no. 9, pp. 989–1005, Sep. 2021, doi: [10.1080/09537325.2020.1862413](https://doi.org/10.1080/09537325.2020.1862413).
- [22] A. Firdaus, M. F. A. Razak, A. Feizollah, I. A. T. Hashem, M. Hazim, and N. B. Anuar, "The rise of 'blockchain': Bibliometric analysis of blockchain study," *Scientometrics*, vol. 120, no. 3, pp. 1289–1331, Sep. 2019, doi: [10.1007/s11192-019-03170-4](https://doi.org/10.1007/s11192-019-03170-4).
- [23] Q.-K. Fu and G.-J. Hwang, "Trends in mobile technology-supported collaborative learning: A systematic review of journal publications from 2007 to 2016," *Comput. Educ.*, vol. 119, pp. 129–143, Apr. 2018, doi: [10.1016/j.compedu.2018.01.004](https://doi.org/10.1016/j.compedu.2018.01.004).
- [24] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, and L. A. Stewart, "Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement," *Systematic Rev.*, vol. 4, no. 1, pp. 1–9, 2015, doi: [10.7326/0003-4819-151-4-200908180-00135](https://doi.org/10.7326/0003-4819-151-4-200908180-00135).
- [25] N. Donthu, W. Reinartz, S. Kumar, and D. Pattnaik, "A retrospective review of the first 35 years of the international journal of research in marketing," *Int. J. Res. Marketing*, vol. 38, no. 1, pp. 232–269, Mar. 2021, doi: [10.1016/j.ijresmar.2020.10.006](https://doi.org/10.1016/j.ijresmar.2020.10.006).
- [26] B. Bakhshinategah, O. R. Zaiane, S. ElAtia, and D. Ipperciel, "Educational data mining applications and tasks: A survey of the last 10 years," *Educ. Inf. Technol.*, vol. 23, no. 1, pp. 537–553, Jan. 2018, doi: [10.1007/s10639-017-9616-z](https://doi.org/10.1007/s10639-017-9616-z).
- [27] S. Dawson, S. Joksimovic, O. Poquet, and G. Siemens, "Increasing the impact of learning analytics," in *Proc. 9th Int. Conf. Learn. Anal. Knowl.*, Mar. 2019, pp. 446–455, doi: [10.1145/3303772.3303784](https://doi.org/10.1145/3303772.3303784).
- [28] R. Nazar, I. S. Chaudhry, S. Ali, and M. Faheem, "Role of quality education for sustainable development goals (SDGS)," *Int. J. Social Sci.*, vol. 4, no. 2, pp. 486–501, Aug. 2018, doi: [10.20319/pjss.2018.42.486501](https://doi.org/10.20319/pjss.2018.42.486501).



CLARE BAEK is currently pursuing the Ph.D. degree with the Rossier School of Education, University of Southern California.

Her research interests include student's self-regulated learning and learning outcomes in technology-integrated learning environments.



TENZIN DOLECK received the Ph.D. degree from McGill University.

He is currently an Assistant Professor and the Canada Research Chair with Simon Fraser University. Prior to joining Simon Fraser University, he was a Provost's Postdoctoral Fellow with the University of Southern California. His research interests include data science, educational data mining, and learning analytics.