Review Article

Bibliometric insights into data mining in education research: A decade in review

Yessane Shrrie Nagendhra Rao [1]

0009-0002-9283-2922

Chwen Jen Chen [1*]

0000-0001-5175-7060

[1] Faculty of Cognitive Sciences and Human Development, Universiti Malaysia Sarawak, Sarawak, MALAYSIA
[*] Corresponding author: cjchen@unimas.my

ARTICLE INFO	ABSTRACT
	This bibliometric study on data mining in education synonymous with big educational data utilizes VOSviewer and Harzing's Publish and Perish to analyze the metadata of 1,439 journal articles found in Scopus from 2010 to 2022. As bibliometric analyses in this field are lacking, this study aims to provide a comprehensive outlook on the current developments and impact of research in this field. This study employs descriptive and trends analysis, co-authorship analysis, co-citation analysis, co-occurrences of keywords, terms map analysis, and analysis of the impact and performance of publications. It also partially replicates a similar study conducted by Wang et al. (2022), who used the Web of Science (WoS) database. The study is reported in an article entitled 'Big data and data mining in education: A bibliometrics study from 2010 to 2022'. Results show that data mining in education is a growing research field. There is also a significant difference between the publications in Scopus and WoS. The study found several research areas and topics, such as student academic performance prediction, e-learning, machine learning, and innovative data mining techniques, to be the core basis for collaborating and continuing current research in this field. These results highlight the importance of continuing research on data mining in education, guiding future research in tackling educational challenges. **Keywords:** educational data mining, big data, education, bibliometric analysis, Scopus

INTRODUCTION

Technology is no longer considered new and unconventional (Rodrigues et al., 2018). The fast-growing tendency of technology deliberately leads to more systematic and sophisticated ways of capturing users' data. These data are procured in large sets that cannot be interpreted thoroughly through conventional methods and are currently referred to as big data (Marín-Marín et al., 2019; Sin & Muthu, 2015). Nevertheless, various data mining techniques have been used to analyze and interpret the data collected as much as possible to influence practices, procedures, and decision-making in diverse fields, including the field of education (Baek & Doleck, 2022; Menon et al., 2017; Rodrigues et al., 2018).

Data Mining & Big Data in Education

According to International Educational Data Mining Society (2011), data mining in education or educational data mining (EDM) is defined as "an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to understand students better, and the settings, which they learn in." Data mining in education has been a growing research area among many researchers (Baek & Doleck, 2022; Romero & Ventura, 2010). Researchers are interested in using data mining techniques and algorithms on big data to analyze and derive relevant and useful educational insights (Romero & Ventura, 2010).

Romero and Ventura (2010) have identified several contexts in which EDM has and can be used to "convert the raw data coming from the educational systems into useful information" (p. 601). The contexts include the traditional classroom setting, where educators and learners are seated in a face-to-face learning environment, blended learning environments, e-learning environments, learning management systems (LMS), intelligent tutoring systems (ITS), massive open online course (MOOC), and adaptive educational hypermedia systems (Romero & Ventura, 2010, 2017). Apart from data mining, Romero and Ventura (2010) highlighted that web mining techniques are also used in the context of online systems by applying them to students' data, which is stored in the underlying databases and log files of the users.

According to Fischer et al. (2020), the emergence of the digital era has brought about the emergence of big data, where students' data in a traditional setting have now come to be digitized through the student information system) and also procured through the increased implementation of LMS. Masood and Mokmin (2017) have also highlighted the use of ITS with big data, which can provide a positive impact on the process of teaching and learning (T&L) with the help of the Internet of things (IoT) technologies. A variety of data mining techniques can be applied to the said data, which can generally be categorized into prediction methods, including inferential techniques that bring upon a paradigm shift in transforming the dynamic information obtained through these data into adaptive knowledge, structure discovery algorithms, with an emphasis on exploring the edifices of content and competencies in an educational setting, as well as the internal structures of learners' social networks; relationship mining, including correlation and sequential pattern mining; and visual analytics (Baker & Siemens, 2014; Fischer et al., 2020).

Bibliometric Analysis & Previous Related Studies

This study employed the bibliometrics analysis (BA) method to explore and examine research development over the past decade in the context of data mining in education. BA is a quantitative method in research used to analyze the properties of the metadata of academic publications, which are relevant to the scope of study of a researcher while also describing the development trends of the said research field (Ahmi, 2021; Wang et al., 2022). It is a popular research method in many fields. However, the publications on BA in data mining in education are lacking or are more or less combined with systematic literature reviews (SLR). With EDM being a multidisciplinary field, fragmented into different research areas, BA research method is useful for analyzing the common areas, where EDM studies are conducted (Baek & Doleck, 2022; Hermaliani et al., 2022). BA allows for the fundamental understanding of EDM (Baek & Doleck, 2022). On the other hand, with big data becoming a widely researched concept, BA can be used to acknowledge further the impact of big data in education through the scholarly articles and publications related to it (Marín-Marín et al., 2019).

Marín-Marín et al. (2019) conducted a bibliometric study encompassing four databases, which were Web of Science (WoS), Scopus, Education Research Information Center (ERIC), and PsycINFO, spanning from 2010 to 2018. Marín-Marín et al. (2019) found that there has been increased publication on big data in education in 2017, offering valuable insights into referent authors, countries, and high statistical value and mapping. Marín-Marín et al. (2019) also found that publications on big data in education started in 2010, underlining the trending and promising nature of big data research and its relevance in evaluating T&L processes. Another study on BA conducted by Baek and Doleck (2022) using the Scopus database from 2015 to 2019 shows that the primary goals of current EDM research are to create and enhance methods for data analysis using cutting-edge technologies and incorporate pedagogy and learning contexts (Baek & Doleck, 2022). Wang et al. (2022) discovered that a major contribution to the literature on EDM and big data in their BA were from the educational technology and computer education fields. They used the WoS database to retrieve the selected articles. The results of BA show rapid growth in research on data mining and big data over the past decade. (Wang et al., 2022).

BA method has also been used relatively more with SLR. Among 291 publications derived from Scopus with approximate keywords, bibliometric analysis, SLR, EDM, and big data in education, 288 of the publications utilized the review methodology. Only three publications independently adopted BA method as of December 22, 2022, which are the earlier described Baek and Doleck (2022), Marín-Marín et al. (2019), and Wang et al. (2022). One review, however, did include BA as a secondary method to gauge development trends. Hermaliani et al. (2022) conducted a systematic review and bibliometric study on how EDM has evolved in predicting student performances from 2015 to 2021 and which data mining techniques were used using the Scopus