
www.ijte.net

# Global Trends of Educational Data Mining in Online Learning

**Nie Hui Ling**
Universiti Malaysia Sarawak, Sarawak, Malaysia

**Chwen Jen Chen**
Universiti Malaysia Sarawak, Sarawak, Malaysia

**Chee Siong Teh**
Universiti Malaysia Sarawak, Sarawak, Malaysia

**Dexter Sigan John**
Universiti Malaysia Sarawak, Sarawak, Malaysia

**Looi Chin Ch'ng**
Universiti Teknologi MARA, Sarawak, Malaysia

**Yoon Fah Lay**
Universiti Malaysia Sabah, Sabah, Malaysia

**To cite this article:**

Ling, N.H., Chen, C.J., Teh, C.S., John, D.S., Ch'ng, L.C., & Lay, Y.F. (2023). Global trends of educational data mining in online learning. *International Journal of Technology in Education (IJTE), 6*(4), 656-680. https://doi.org/10.46328/ijte.558

International Journal of Technology in Education
2023, Vol. 6, No. 4, 656-680 https://doi.org/10.46328/ijte.558

# Global Trends of Educational Data Mining in Online Learning

**Nie Hui Ling, Chwen Jen Chen, Chee Siong Teh, Dexter Sigan John, Looi Chin Ch'ng, Yoon Fah Lay**

## Article Info

***Article History***

Received:
12 March 2023
Accepted:
16 September 2023

***Keywords***

Educational data mining
Online learning
Bibliometric analysis
Global trends

## Abstract

Educational data mining (EDM) in online learning involves data mining techniques to analyze data from online environments to gain insights into student behavior, performance, and engagement. This study explored EDM in online learning publication trends and focuses. It involved a bibliometric analysis of 615 scholarly works related to EDM in online learning as recorded in Scopus, the largest peer-reviewed citation database, on February 1, 2023. The study examined EDM in online learning publications regarding its evolution and distribution, key focus areas, impact and performance, and prominent authors and collaborations in the last decade, in which the timespan is the period from 2012 to 2022. This bibliometric analysis shows that EDM in online learning is a dynamic area of scientific research as related publications grow steadily throughout the years and involve worldwide collaborations. The study reveals current research trends, offering valuable insights for future researchers to guide their investigations in this field.

## Introduction

Over the last decade, the rapid advancement of information and communication technology has led to significant growth in online learning or e-learning. Online learning became crucial when COVID-19 struck in March 2020. The pandemic has caused educational institutions to operate remotely, which has resulted in a massive increase in online learners. The proliferation of various online learning environments and student information systems has produced tremendous educational data (Bakhshinategh et al., 2018; Romero & Ventura, 2020). A typical online learning system tracks and records many educational activities that capture the continuous interaction of teaching and learning in databases and log files (Estacio & Raga, 2017). Similarly, a typical student information system records vast volumes of data such as student enrolment, student demographics, attendance records, examination results, and so forth (Dutt et al., 2017). Bienkowski et al. (2012) highlight the challenge of extracting knowledge and patterns from such huge repositories of educational data to benefit stakeholders, improve learning outcomes, and support relevant decision-making. Educational data mining (EDM) emerged as a research field in the 1990s (Romero & Ventura, 2007). It uncovers meaningful insights, hidden patterns, and relationships among a large amount of educational data (Huebner, 2013). It also aids in solving various educational problems and affords the implementation of more interactive, adaptive, and personalized educational environments (Papamitsiou & Economides, 2014). Clustering, classification, sequential patterns, machine learning models, and association rule analysis are some of the most prevalent techniques employed in EDM (Salloum et al., 2020).

Based on some of the latest scientific publications on EDM in online learning, EDM models that aim to enhance learning design and activities in online environments are the focus of recent publications. Real et al. (2020) employed Educational Process Mining techniques to track the learning paths of 73 undergraduate students in an introductory programming course. The study identified differences in the paths taken by students who passed and failed the course and provided insights for teachers to observe students' behavior patterns and progress. Using text-mining techniques, Lee and Recker (2022) analyzed the impact of student and instructor participation in online discussions in introductory online mathematics/statistics courses. They found that online listening behaviors and posts that build on other contributions or apply new knowledge are positively linked to course performance and the instructors' use of open-ended prompts. Realinho et al. (2022) used demographic, socioeconomic, macroeconomic, and academic data on enrollment and past academic performance to develop machine learning models to predict academic performance and dropout risks. Tawfeeq et al. (2022) used data mining techniques to predict students' attitudes toward e-learning and to determine the impact of e-learning on students.

These latest EDM publications indicate that EDM is crucial for improving online learning. EDM in online learning has developed into an active field of research and led to the rapid growth of publications in recent years. This study aims to offer a holistic view of the current trends for EDM in online learning publications with suggestions for future studies by conducting a bibliometric analysis of publications in the last decade, from 2012 to 2022. Specifically, it seeks answers to the following research questions (RQs):

RQ1: How has EDM in online learning publications evolved and been distributed?

RQ2: What are the key areas discussed in EDM in online learning publications?

RQ3: What is the impact and performance of EDM in online learning publications?

RQ4: Who are the prominent authors of EDM in online learning publications, and how have they collaborated?

This study is valuable because it provides a comprehensive mapping of the fast-developing field of EDM, which has become increasingly important in the digital era. Through bibliometric analysis of a sizable corpus of scholarly articles, this research not only charts the trajectory of the subject over the last ten years but also reveals its key focus areas, impact, and prominent contributors. By emphasizing current research trends and exposing gaps in the literature, the insights gained serve as a useful resource for academics, educators, and policymakers. This can influence educational practices and direct future research and collaborations, supporting the improvement of online learning environments for better student outcomes.

## Methodology

Bibliometric analysis is a research methodology that employs numerical and statistical analysis of academic publications. To uncover important trends, authors, and research gaps, this approach aids in evaluating the impact, relevance, and growth of scholarly publications within a certain topic (Ahmi, 2021). This study employed bibliometric analysis for all four RQs. RQ1 focuses on the number of publications over the examined years, source types, source titles, and languages used. RQ2 identifies the main subject areas and topics discussed, RQ3 examines

the citation metrics, and RQ4 determines the countries and institutions with the most contributions, collaboration patterns, impact, and performance of publications.

**Data Source**

The data source was publications retrieved from the Scopus academic database. The Scopus database offers access to peer-reviewed literature, scientific journals, books, and conference proceedings (Adeniji, 2019). Moreover, Elsevier maintains Scopus and integrates titles from over 5,000 publishers, with over 20,000 serial titles, 150,000 books, and more than 70 million items (Dormezil et al., 2019).

**Data Collection**

The study used the following search query on the title, abstract, and keyword of Scopus publications to identify the closest matching publications. (TITLE-ABS-KEY ("data mining" AND "learning management system*" OR "e-learning system*" OR "online learning system*") OR TITLE-ABS-KEY ("educational data mining" AND "learning management system*" OR "e-learning system*" OR "online learning system*")) AND PUBYEAR > 2011 AND PUBYEAR < 2023. The search retrieved 615 publications.

**Data Analysis**

The study employed various data analysis methods to answer the proposed research questions. It retrieved some results directly from Scopus, while some other Scopus data were inserted manually into an Excel file to derive some descriptive statistics. Publish or Perish was used to perform citation metrics to evaluate the impact and performance of the publications. It also employed VOSviewer (version 1.6.17) to perform keyword co-occurrence and authorship analysis by generating relevant visualizations.

## Results and Discussion

This section describes results retrieved from bibliometric analysis following the four research questions:

RQ1: How has EDM in online learning publications evolved and been distributed?
RQ2: What are the key areas discussed in EDM in online learning publications?
RQ3: What is the impact and performance of EDM in online learning publications?
RQ4: Who are the prominent authors of EDM in online learning publications, and how have they collaborated?

**RQ1: How has EDM in online learning publications evolved and been distributed?**

**Evolution and Distribution of Publications on EDM in Online Learning**

The data source was publications retrieved from the Scopus academic database. The Scopus database offers access to peer-reviewed literature, scientific journals, books, and conference proceedings (Adeniji, 2019).

***Publications by Year***

Table 1 indicates the yearly statistical information of publications for EDM in online learning from 2012 to 2022. This statistical information depicts a general trend of increasing numbers in publications in the past decade. From 2012 to 2022, publications increased gradually from 33 to 63, with 94 being the highest yearly number produced in the past decade.

Table 1. Number of Publications by Year

| Year | Number of Documents | Percentage (%) | Cumulative Percentage (%) |
|---|---|---|---|
| 2012 | 33 | 5.37 | 5.37 |
| 2013 | 33 | 5.37 | 10.74 |
| 2014 | 47 | 7.64 | 18.38 |
| 2015 | 39 | 6.34 | 24.72 |
| 2016 | 50 | 8.13 | 32.85 |
| 2017 | 51 | 8.29 | 41.14 |
| 2018 | 65 | 10.57 | 51.71 |
| 2019 | 94 | 15.28 | 66.99 |
| 2020 | 65 | 10.57 | 77.56 |
| 2021 | 75 | 12.20 | 89.76 |
| 2022 | 63 | 10.24 | 100 |
| Total | 615 | 100 | 100 |

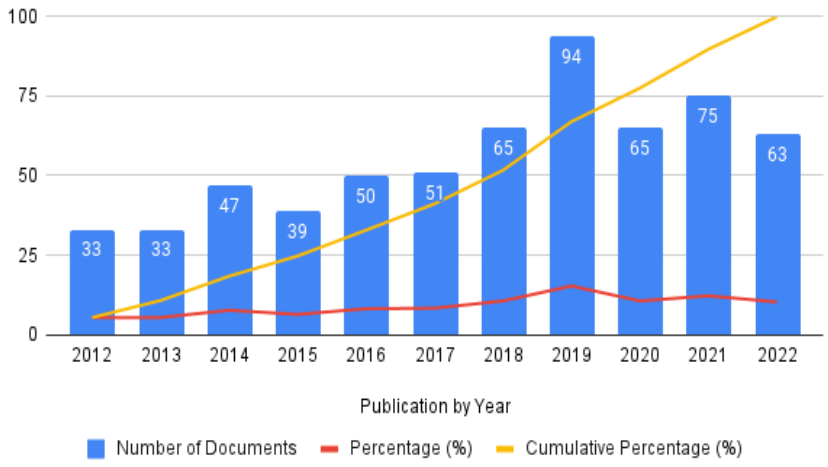


Figure 1. Number of Publications by Year

***Publications Type***

This study uncovers the distribution of EDM in online learning publications by examining its type. The two major document types of publications are conference papers, representing (n=357; 58.0%), followed by journal articles (n=217; 35.3%). In addition, book chapters (n=17; 2.76%) and conference reviews (n=14; 2.28%) are publication

types with more than ten items. Lastly, reviews (n=9; 1.46%) and books (n=1; 0.16%) are the publication types with ten or fewer items.

Table 2. Document Types

| Document Type | Number of Documents | Percentage (%) |
|---|---|---|
| Conference Paper | 357 | 58.0 |
| Journal Article | 217 | 35.3 |
| Book Chapter | 17 | 2.76 |
| Conference Review | 14 | 2.28 |
| Review | 9 | 1.46 |
| Book | 1 | 0.16 |
| Total | 615 | 100 |

The sources for publications of EDM in online learning were mainly conference papers and articles, with more than 83% of the publications. Conferences, accounting for about 60%, still are the preferred source for publications on EDM in online learning. It is expected that the rate of journal publication will increase, accompanying the maturity increase in the area, which is still a recent area of research.

***Source Titles***

Table 3 shows that the most significant number of publications on EDM in online learning is in the conference proceeding series, ACM International Conference Proceeding Series, with 31 publications (5.04%). This is followed closely by the Lecture Notes in Computer Science, including its subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics, which contributed 25 publications (4.07%). Journal series entitled the International Journal of Emerging Technologies in Learning contributed notably, with 12 documents (1.95%), Communications in Computer and Information Science, and Education and Information Technologies with 11 documents each (1.79%). Table 3 comprehensively lists the top 20 publication titles on EDM in online learning.

Table 3. Top 20 Source Titles

| Source Title | TD | % | Publisher | CS | SJR | SNIP |
|---|---|---|---|---|---|---|
| ACM International Conference Proceeding Series | 31 | 5.04 | Association for Computing Machinery (ACM) | 1.0 | 0.232 | 0.310 |
| Lecture Notes In Computer Science Including Subseries Lecture Notes In Artificial Intelligence And Lecture Notes In Bioinformatics | 25 | 4.07 | Springer Nature | 2.1 | 0.407 | 0.534 |

| Source Title | TD | % | Publisher | CS | SJR | SNIP |
|---|---|---|---|---|---|---|
| International Journal of Emerging Technologies In Learning | 12 | 1.95 | International Association of Online Engineering | 3.8 | 0.632 | 1.414 |
| Communications In Computer And Information Science | 11 | 1.79 | Springer Nature | 0.9 | 0.209 | 0.286 |
| Education And Information Technologies | 11 | 1.79 | Springer Nature | 6.6 | 1.055 | 2.121 |
| Ceur Workshop Proceedings | 10 | 1.63 | CEUR Workshop | 1.1 | 0.228 | 0.317 |
| IEEE Access | 9 | 1.46 | IEEE | 6.7 | 0.927 | 1.326 |
| Journal Of Physics: Conference Series | 7 | 1.14 | - | 0.8 | 0.21 | 0.395 |
| IEEE Transactions on Learning Technologies | 6 | 0.98 | IEEE | 7.4 | 1.288 | 2.044 |
| Procedia Computer Science | 6 | 0.98 | - | 3.6 | 0.569 | 1.062 |
| International Journal of Advanced Computer Science and Applications | 5 | 0.81 | Science and Information Organization | 1.8 | 0.284 | 0.528 |
| International Journal of Distance Education Technologies | 5 | 0.81 | IGI Global Publishing | 3.4 | 0.458 | 0.790 |
| Smart Innovation Systems and Technologies | 5 | 0.81 | Springer Nature | 1.1 | 0.224 | 0.398 |
| Frontiers In Artificial Intelligence and Applications | 4 | 0.65 | IOS Press | 0.9 | 0.264 | 0.382 |
| Proceedings Frontiers in Education (FIE) Conference | 4 | 0.65 | - | 0.7 | 0.221 | 0.395 |
| Springerbriefs in Applied Sciences and Technology | 4 | 0.65 | Springer Nature | 0.6 | 0.104 | 0.0 |
| Lecture Notes in Electrical Engineering | 4 | 0.65 | Springer Nature | 0.6 | 0.148 | 0.145 |
| Computer Applications in Engineering Education | 3 | 0.49 | Wiley-Blackwell | 4.2 | 0.594 | 1.28 |
| Computers in Human Behavior | 3 | 0.49 | Elsevier | 14.9 | 2.174 | 3.234 |
| Educational Technology and Society | 3 | 0.49 | International Forum of | 7.2 | 1.311 | 2.445 |

| Source Title | TD | % | Publisher | CS | SJR | SNIP |
|---|---|---|---|---|---|---|
| | | | Educational Technology, National Taiwan Normal University and Society | | | |

Note: TD=total documents; %= percentage; CS= CiteScore counts the citations received in 2012-2022 to articles, reviews, conference papers, book chapters, and data papers published in 2012-2022 and divides this by the number of publications published in 2012-2022; SJR= SCImago Journal Rank measures weighted citations received by the serial. Citation weighting depends on the subject field and prestige (SJR) of the citing serial; SNIP= Source Normalized Impact per Paper measures actual citations received relative to citations expected for the serial's subject field.

Although the source title indicates there are already several publication sources focused only on EDM in educational technologies, there are still many publications in sources dedicated to more general focuses, such as computer science and computational intelligence systems. The analysis of the subject categories of the publications reveals that EDM in online learning is a multidisciplinary field of research, focusing mainly on the several sub-areas of computer science and education. Major educational technology journals that publish work on EDM in online learning include the International Journal of Emerging Technologies in Learning, Education and Information Technologies, IEEE Transactions on Learning Technologies, the International Journal of Distance Education Technologies, and Educational Technology and Society.

***Languages Used***

Referring to Table 4, 98.4% of the 615 publications use English. Spanish is the second most common language (n=5; 0.81%). Subsequently, Portuguese, Turkish, and Chinese contributed less than five publications, which accounted for 0.49%, 0.16%, and 0.16% of the total publications, respectively.

Table 4. Language Used

| Language | Number of Documents | Percentage (%) |
|---|---|---|
| English | 606 | 98.4 |
| Spanish | 5 | 0.81 |
| Turkish | 3 | 0.49 |
| Portuguese | 1 | 0.16 |
| Serbian | 1 | 0.16 |
| Book | 1 | 0.16 |
| Total | 615 | 100 |

Unsurprisingly, 98% of the publications were written in the English Language. This is due to the substantial domination of the English language in the publication sources accepted by the scientific community as having high-quality standards. This way, most researchers tend to publish in these sources expecting to be recognized by

the scientific community.

**RQ2: What are the key areas discussed in EDM in online learning publications?**

**Key Areas of EDM in Online Learning**

This study examined two aspects, namely (i) subject area, (ii) keywords occurrence, and (iii) keywords co-occurrence analysis, to identify the key areas discussed in EDM in online learning publications.

***Subject Area***

Table 5 shows the distribution of EDM in online learning publications by subject area. Twenty-three subject areas were identified, with the most publications in computer science (41.05%, n=477), followed by social sciences 198 (17.04%) and engineering 193 (16.61%). Mathematics and decision sciences followed with 88 (7.57%) and 57 (4.9%) publications. Other subject areas, including physics and astronomy, business, management and accounting, materials sciences, energy, medicine, psychology, arts and humanities, chemical engineering, environmental science, biochemistry, genetics and molecular biology, multidisciplinary, earth and planetary sciences, health professions, neurosciences, chemistry, economics, econometrics and finance, nursing, and pharmacology, toxicology, and pharmaceutics, accounted for less than 30 publications each.

Table 5. Subject Areas

| Subject Area | Number of Documents | Percentage (%) |
|---|---|---|
| Computer Sciences | 477 | 41.05 |
| Social Sciences | 198 | 17.04 |
| Engineering | 193 | 16.61 |
| Mathematics | 88 | 7.57 |
| Decision Sciences | 57 | 4.90 |
| Physics and Astronomy | 22 | 1.89 |
| Business, Management, and Accounting | 21 | 1.81 |
| Material Sciences | 21 | 1.81 |
| Energy | 20 | 1.72 |
| Medicine | 15 | 1.29 |
| Psychology | 12 | 1.03 |
| Arts and Humanities | 8 | 0.69 |
| Chemical Engineering | 6 | 0.52 |
| Environmental Science | 6 | 0.52 |
| Biochemistry, Genetics and Molecular Biology | 4 | 0.34 |
| Multidisciplinary | 4 | 0.34 |
| Earth and Planetary Sciences | 2 | 0.17 |
| Health Professions | 2 | 0.17 |
| Neuroscience | 2 | 0.17 |

| Subject Area | Number of Documents | Percentage (%) |
|---|---|---|
| Chemistry | 1 | 0.086 |
| Economics, Econometrics and Finance | 1 | 0.086 |
| Nursing | 1 | 0.086 |
| Pharmacology, Toxicology and Pharmaceutics | 1 | 0.086 |

Findings suggest that EDM in online learning publications is confined mainly within the domain of computer sciences, social sciences, and engineering areas, particularly as these relate to education.

***Keyword Occurrence***

Table 6 summarizes the keywords that occur most in EDM in online learning publications. Based on the findings, three core keywords related to the search query were data mining (n=533), learning management system (LMS) (n=438), and e-Learning (n=360). Apart from the keywords relevant to the search query, student is the keyword most associated with EDM in e-learning studies (n=277). Furthermore, learning analytics (n=108), teaching and learning (n=94), machine learning (n=60), clustering (n=56), and classification (n=53) were the keywords with 50 and above occurrences.

Table 6. Top 20 Keywords

| Keyword | Occurrence | Total Link Strength |
|---|---|---|
| Data mining | 533 | 520 |
| Learning Management Systems (LMS) | 438 | 435 |
| e-Learning | 360 | 355 |
| Student | 277 | 277 |
| Learning analytics | 108 | 106 |
| Teaching and learning | 94 | 93 |
| Machine learning | 60 | 60 |
| Clustering | 56 | 56 |
| Classification | 53 | 53 |
| Algorithm | 49 | 49 |
| Decision Tree | 48 | 48 |
| Prediction | 48 | 48 |
| Higher Education | 36 | 36 |
| Information and Communication Technologies | 32 | 32 |
| Knowledge-Based Systems | 30 | 30 |
| Text Mining | 27 | 27 |
| Decision Making | 27 | 27 |
| Social Networking (Online) | 26 | 26 |
| Educational Environment | 25 | 25 |
| Data Analysis | 24 | 24 |

| Keyword | Occurrence | Total Link Strength |
|---|---|---|
| Decision Tree | 48 | 48 |
| Prediction | 48 | 48 |
| Higher Education | 36 | 36 |

The ten top most occurred keywords show that EDM in online learning concerns educational data mining, e-learning, learning analytics, Moodle, machine learning, clustering, classification, online learning, prediction, and student performance prediction. The most common learning management system used in related studies is Moodle. Moodle LMS is the open-source learning management system used by hundreds of millions of learners worldwide and is claimed to be the heart of the world's most customizable and trusted e-learning solution (Usov et al., 2020).

The unique feature of Moodle LMS is that it provides learning analytics about students' behavior (Usov et al., 2020). Researchers frequently employ these learning analytics to work with data mining techniques to predict student performance, cluster students, minimize dropout rate, and so forth (Abana, 2019; Kabakchieva, 2013; Zhang et al., 2020). Based on the most common associated keywords, it indicates that numerous researchers are working with Moodle LMS to retrieve the logs and conduct data mining.

Research regarding software engineering and implementation of learning analytics in Moodle LMS (Dondorf et al., 2019) is relevant. For example, Dondorf et al. (2019) conducted a study revealing that Moodle LMS has external data storage services to cope with a massive number of Moodle logs and yet only allowed calculations within the Moodle LMS. Therefore, the retrieval and transfer of data can be hindered between the Moodle LMS and third-party services, and extra effort is necessary for the set-up process in the third-party services (Dondorf et al., 2019). In their study, they developed a plugin easily installable in Moodle LMS to work with Moodle logs, as no external application is required. In sum, future work can further advance the functionality and usability of Moodle LMS.

The keywords suggested that machine learning, classification, clustering, and prediction were commonly used data mining techniques. Hence, studies on EDM in online learning can adopt other data mining techniques. For example, time series analysis examines temporal patterns in data from online learning, such as changes in student activity over time. This can reveal patterns in the involvement and performance of students, including trends, seasonality, and recurrent patterns.

Deep learning approaches, such as recurrent neural networks (RNNs) and transformer models, perform natural language processing tasks on text data available in online learning platforms for automated grading of open-ended tasks, text summarizing, and sentiment analysis. Using network analysis, researchers can examine how students in online learning communities interact with one another and share knowledge. Understanding the effects of social networks on learning outcomes, identifying influential learners, and researching collaboration patterns are all aided by this. Additionally, outlier detection can be another data mining technique used to investigate abnormalities in student behaviors and performance.

***Keyword Co-occurrence Analysis***

This study used VOSviewer, a software tool for constructing and visualizing bibliometric networks to map authors' keywords. Based on the system-illustrated diagram, the color, circle size, font size, and thickness of connecting lines indicate the relationships with other keywords (Wahid et al., 2020). The size of nodes indicates the frequency of occurrence. The curves between the nodes show their co-occurrence in the same publication. The shorter the distance between two nodes, the larger the co-occurrences of the two keywords (Yu et al., 2020). The findings of the co-occurrence analysis for EDM in online learning can be viewed at https://tinyurl.com/4vfd3ep7. The network visualization map (Figure 2) of author keywords with at least five occurrences reveals seven clusters:

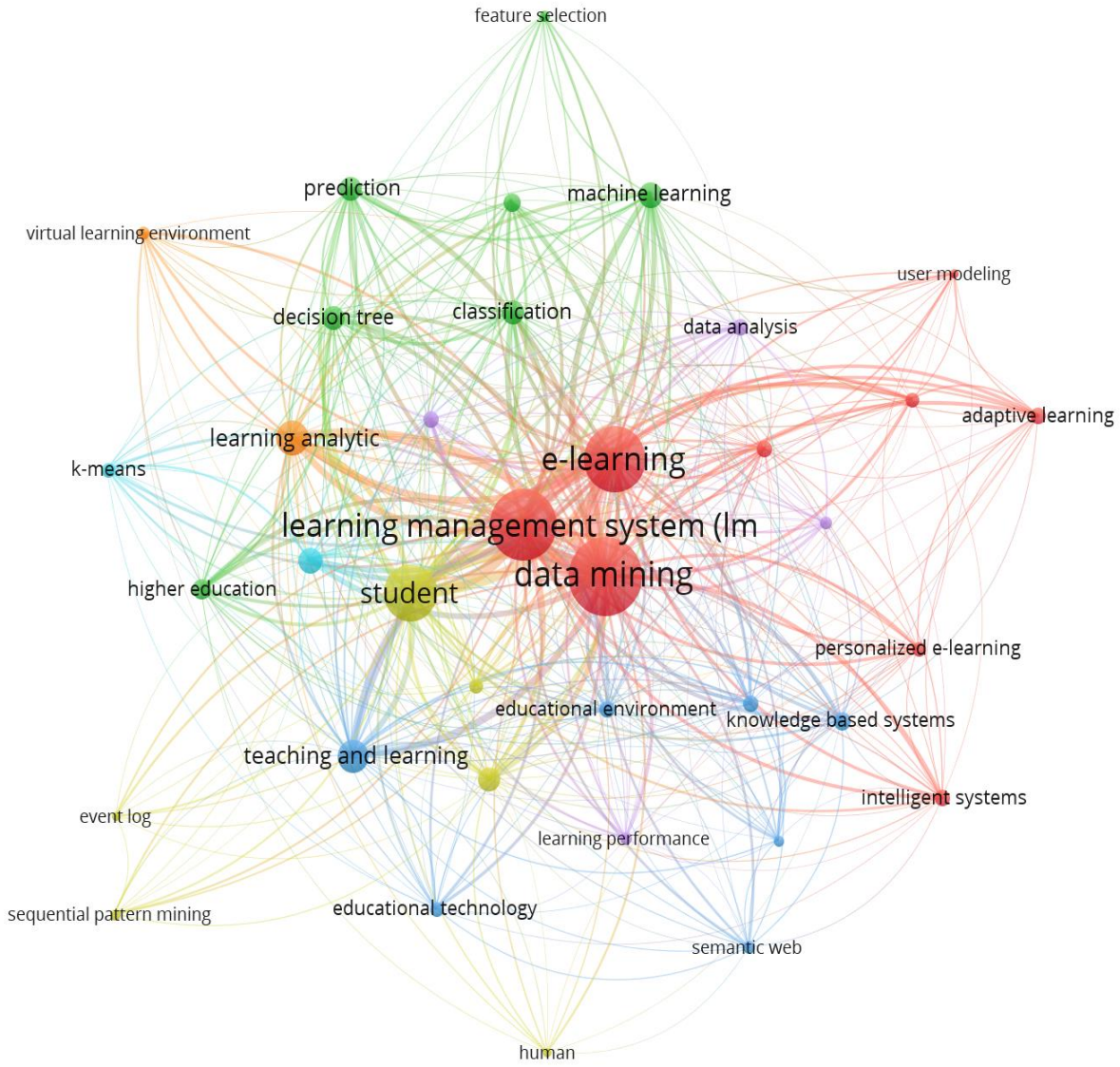


Figure 2. Network Visualization Map of Author Keywords Co-occurrence

The first cluster comprises keywords related to data mining (occurrence=533, links=36, total link strength=520), LMS (occurrence=438, links=36, total link strength=435), and e-learning (occurrence=360, links=36, total link strength=355). Additionally, the second cluster consists of keywords related to machine learning (occurrence=60, links=29, total link strength=60), classification (occurrence=53, links=30, total link strength=53), and prediction

(occurrence=48, links=28, total link strength=48). Subsequently, the third cluster focuses on teaching and learning (occurrence=94, links=35, total link strength=93), knowledge-based systems (occurrence=30, links=26, total link strength=30), and text mining (occurrence=27, links=25, total link strength=27). Moreover, the fourth cluster is centered around the student (occurrence=277, links=36, total link strength=277), algorithms (occurrence=49, links=36, total link strength=49), and learning materials (occurrence=22, links=30, total link strength=22). Additionally, the fifth cluster focuses on decision making (occurrence=27, links=27, total link strength=27), data analysis (occurrence=24, links=24, total link strength=24), and learning performance (occurrence=17, links=23, total link strength=17). The sixth cluster comprises keywords related to clustering (occurrence=56, links=31, total link strength=56) and k-means (occurrence=18, links=20, total link strength=18). Lastly, the seventh cluster encompasses learning analytics (occurrence=108, links=31, total link strength=106) and virtual learning environments (occurrence=15, links=17, total link strength=15).

Overall, the keyword co-occurrence analysis reveals seven clusters that depict common research areas on EDM in online learning. These include data mining in e-learning systems, data mining techniques, data mining for teaching and learning, students and learning materials, data-driven decisions and insights, and learning analytics. The findings provide valuable insights into the current research trends in this field, which can inform the research and practice of educators and researchers alike. Figure 3 illustrates the overlay visualization map.

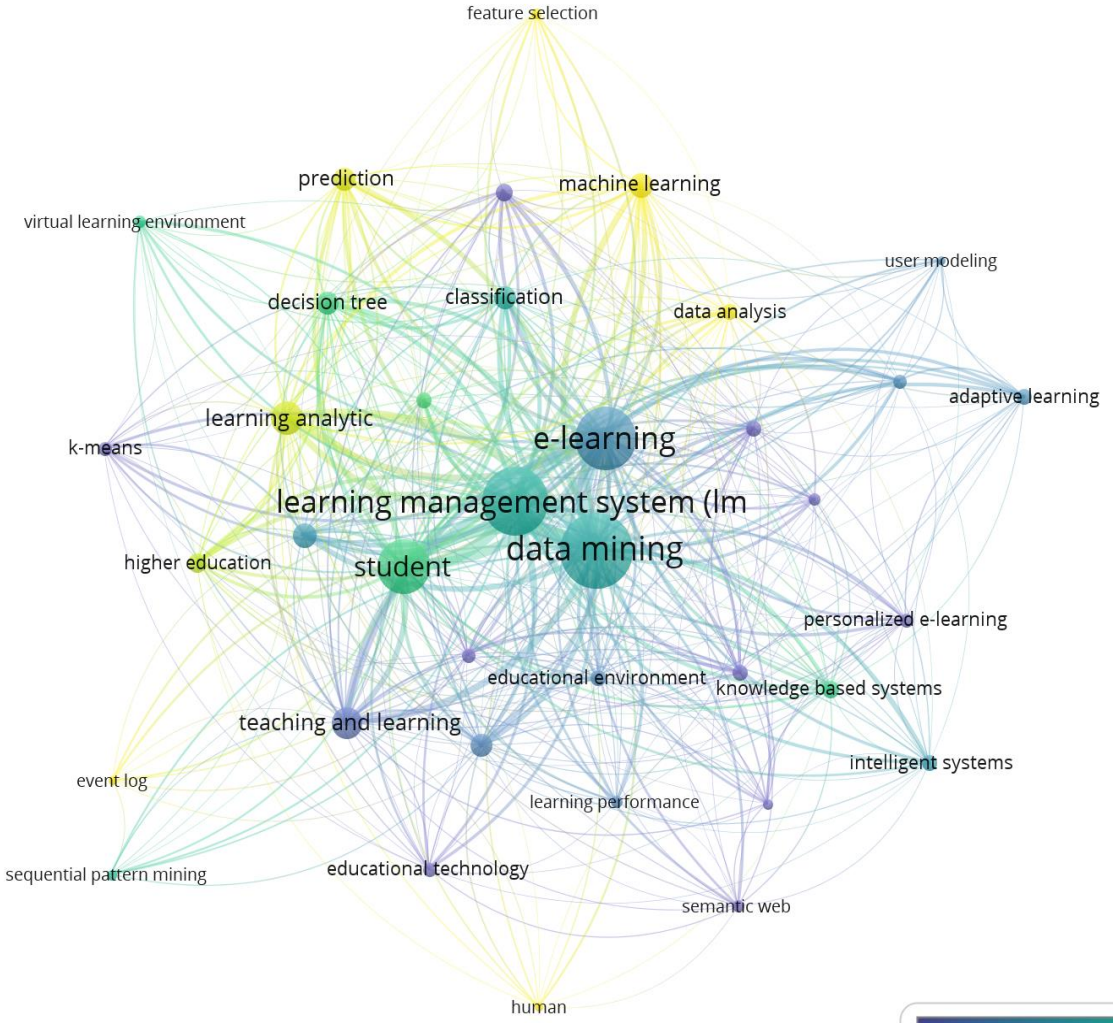


Figure 3. Overlay Visualization Map of Author Keywords

The overlay visualization map represents the currentness of the items for EDM in online learning publications. By default, colors range from blue to green to yellow, showing the least current to the most current (Ahmi, 2021). Therefore, the items marked as yellow, such as data analysis, machine learning, feature selection, prediction, learning analytic, and event log indicate EDM in online learning current publications focuses.

**RQ3: What is the impact and performance of EDM in online learning publications?**

**Impact and Performance of EDM in Online Learning Publications**

The study derived the citation metrics to gain insight into the impact and performance of examined publications.

***Citation Analysis***

Table 7 reports the citation metric of the publications evaluated through Perish or Publish software. There are 6273 citations reported in 10 years (2012–2022) for 615 documents, with an average of 627.3 citations per year. Figure 4 shows that the total citations gradually increased to 1537 in 2022.

Table 7. Citation Metric of the Publications

| Description | Results |
|---|---|
| Publication years | 2012-2022 |
| Citation years | 10 (2012-2022) |
| Total documents | 615 |
| Total citations | 6273 |
| Citations per year | 627.3 |
| Citations per paper | 10.23 |
| Authors per paper | 3.21 |
| h-index | 38 |
| g-index | 63 |

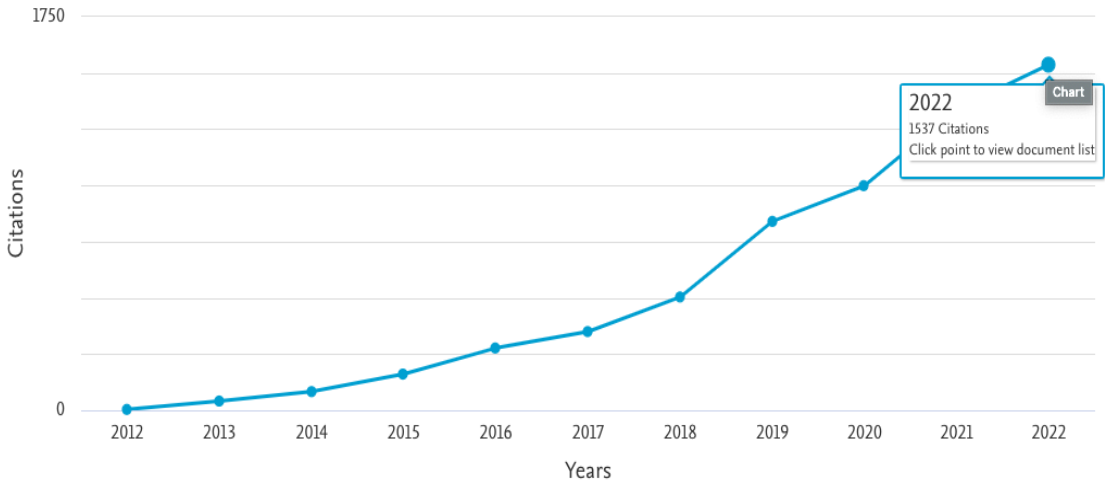


Figure 4. Total Citations from 2012 to 2022

Table 8 summarizes the ten most cited publications. The topmost cited document is "Sentiment Analysis in Facebook and Its Application to E-learning" by Ortigosa et al. (2014) (TC=382). It introduces SentBuk, a Facebook application that performs sentiment analysis on users' messages to extract information about their sentiment polarity and detect significant emotional changes. The hybrid approach used in SentBuk achieves high accuracy in sentiment classification and has potential applications in online learning for personalized learning and feedback for teachers. The proliferation of studies related to personalized online learning, such as Chang et al. (2022), Abhirami and Devi (2022), and Raj and Renumol (2022), justifies the high citations of this publication. By adapting the learning experience to each student's needs, preferences, and learning style, personalized e-learning can enhance student engagement, motivation, and achievement. Moreover, personalized learning can also help address the challenges of scalability and inclusivity in education by providing flexible and accessible learning opportunities to diverse student populations.

The second most cited document is "Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research", produced by Schwendimann et al. (2017) (TC=239), reviews the state-of-the-art of research on learning dashboards in the fields of Learning Analytics and EDM. It highlights the need for further longitudinal research and systematic comparisons of dashboard designs. Finally, "Web Usage Mining for Predicting Final Marks of Students that Use Moodle Courses" by Romero et al. (2013) is ranked the third most cited document with 198 citations. It reports how web usage mining can be applied in e-learning systems to predict university students' final examination marks. This study aims to develop a specific Moodle mining tool for data mining experts and novice users such as instructors and courseware authors (Romero et al., 2013). Additionally, it compared the performance of different data mining techniques for classifying students using students' web usage data in several Cordoba University Moodle engineering courses (Romero et al., 2013). The high citations explain the current research focus on implementing and comparing different EDM techniques in e-learning applications.

Table 8. Top 20 Most-Cited Documents

| Publication Year | Author | Title | Source | Total Citations (TC) |
|---|---|---|---|---|
| 2014 | Ortigosa, A., Martin. J. M., & Carro. R. M | Sentiment Analysis in Facebook and Its Application to E-learning | Computers in Human Behavior | 382 |
| 2017 | Schwendimann, B. A., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L.P., Boroujeni, M.S., Holzer, A., Gillet, D., & Dillenbourg, P. | Perceiving Learning at a Glance: A Systematic Literature Review of Learning Dashboard Research | IEEE Transactions on Learning Technologies | 239 |

| Publication Year | Author | Title | Source | Total Citations (TC) |
|---|---|---|---|---|
| 2013 | Romero, C., Espejo, P.G., Zafra, A., Romero, J.R., & Ventura, S. | Web Usage Mining for Predicting Final Marks of Students that use Moodle Courses | Computer Applications in Engineering Education | 198 |
| 2014 | Hu, Y. H., Lo, C. L., & Shih, S. P | Developing Early Warning Systems to Predict Students Online Learning Performance | Computers in Human Behavior | 182 |
| 2016 | Cerezo, R., Sanchez-Santillan, M., Paule-Ruiz, M.P., & Nunez, J.C. | Students' LMS Interaction Patterns and Their Relationship with Achievement: A Case Study in Higher Education | Computers and Education | 181 |
| 2013 | Aher, S. B., & Lobo, L. M. R. J. | Combination of Machine Learning Algorithms for Recommendation of Courses in E-Learning System based on Historical Data | Knowledge-Based Systems | 137 |
| 2018 | Burgos, C., Campanario, M.L., Pena, D.D.L., Lara, J.A., Lizcano, D., & Martinez, M.A. | Data Mining for Modeling Students' Performance: A Tutoring Action Plan to Prevent Academic Dropout | Computers and Electrical Engineering | 105 |
| 2016 | Elbadrawy, A., Polyzou, A., Ren, Z., Sweeney, M., Karypis, G., & Rangwala, H. | Predicting Student Performance Using Personalized Analytics | Computer | 111 |
| 2012 | Abdous, M., He, W., & Yen, C.-J. | Using Data Mining for Predicting Relationships between Online Question Theme and Final Grade | Educational Technology and Society | 102 |
| 2018 | Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson. S., Murray, D.J., & Long, Q. | Predicting Academic Performance by Considering Student Heterogeneity | Knowledge-Based Systems | 100 |
| 2014 | Lara, J.A., Lizcano, D., Martinez, M.A., | A system for knowledge discovery in e-learning | Computers and Education | 92 |

| Publication Year | Author | Title | Source | Total Citations (TC) |
|---|---|---|---|---|
| | Pazos, J., & Riera, T. | environments within the European Higher Education Area - Application to student data from Open University of Madrid, UDIMA | | |
| 2016 | Krouska, A., Troussas, C., & Virvou, M. | The Effect of Preprocessing Technique in Twitter Sentiment Analysis | IISA 2016 - 7th International Conference on Information, Intelligence, Systems and Applications | 89 |
| 2016 | Park Y., Yu J.H., & Jo I.-H. | Clustering Blended Learning Courses by Online Behavior Data Case Study in a Korean Higher Education Institute | Internet and Higher Education | 86 |
| 2015 | Amrieh, E.A., Hamtini, T., & Aljarah, I. | Preprocessing and Analyzing Educational Data Set using X-API for Improving Student's Performance | 2015 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies, AEECT 2015 | 74 |
| 2013 | Del-Blanco, A., Serrano, A., Freire, M., Martinez-Ortiz, I., & Fernandez-Manjon, B. | E-learning Standards and Learning Analytics. Can Data Collection Be Improved by Using Standard Data Models? | IEEE Global Engineering Education Conference, EDUCON | 69 |
| 2019 | Juhanak, L., Zounek, J., & Rohlikova, L. | Using Process Mining to Analyse Students' Quiz-taking Behaviour Patterns in a Learning Management System | Computers in Human Behavior | 63 |
| 2018 | Bharara, S., Sabitha, S., & Bansal, A. | Application of Learning Analytics Using Clustering Data Mining for | Education and Information | 58 |

| Publication Year | Author | Title | Source | Total Citations (TC) |
|---|---|---|---|---|
| | | Students' Disposition Analysis | Technologies | |
| 2020 | Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K., & Sattar, M. | Predicting Student Performance in Higher Educational Institutions Using Video Learning Analytics and Data Mining Techniques | Applied Sciences (Switzerland) | 56 |
| 2018 | Adejo, O., & Connolly, T. | Predicting Student Academic Performance Using Multi-model Heterogeneous Ensemble Approach | Journal of Applied Research in Higher Education | 54 |
| 2012 | Fernández, A., Peralta, D., Herrera, F., & Benitez, J. M. | An Overview of E-learning in Cloud Computing | Advances in Intelligent Systems and Computing | 55 |

**RQ4: Who are the prominent authors of EDM in online learning publications, and how have they collaborated?**

**Prominent Authors and Collaborations among EDM in Online Learning Publications**

There are several aspects used to explain the prominent authors and collaborators of EDM in online learning publications, namely (i) countries with the most publications, (ii) institutions with the most publications, and (iii) the co-authorship analysis.

***Countries with the Most Publications***

Table 9 tabulates the data relevant to the top 20 countries contributing the most to EDM in online learning publications.

Table 9. Top 20 Most-Contributed Countries

| Countries | Number of Documents | Percentage (%) |
|---|---|---|
| India | 72 | 11.71 |
| Spain | 56 | 9.11 |
| China | 53 | 8.62 |
| United States | 44 | 7.15 |
| Japan | 33 | 5.37 |
| Australia | 22 | 3.58 |
| Greece | 22 | 3.58 |

| Countries | Number of Documents | Percentage (%) |
|---|---|---|
| Indonesia | 21 | 3.41 |
| Malaysia | 20 | 3.25 |
| Canada | 17 | 2.76 |
| Brazil | 16 | 2.60 |
| Pakistan | 15 | 2.44 |
| Saudi Arabia | 15 | 2.44 |
| Taiwan | 15 | 2.44 |
| Turkey | 15 | 2.44 |
| Croatia | 14 | 2.28 |
| Ecuador | 14 | 2.28 |
| Oman | 14 | 2.28 |
| Germany | 13 | 2.11 |
| Russian Federation | 12 | 1.95 |

India (n=72, 11.71%) is the leading country with the most publications, followed by Spain (n=56, 9.11%), China (n=53, 8.62%), and the United States (n=44, 7.15%). Subsequently, Japan, Australia, Greece, Indonesia, and Malaysia contributed 5.37%, 3.58%, 3.58%, 3.41%, and 3.25% of the total publications. The remaining countries spread across the globe, such as Canada, Brazil, Pakistan, Saudi Arabia, Taiwan, Turkey, Croatia, Ecuador, Oman, Germany, and the Russian Federation, with each country contributes less than 3% of the publications.

***Institutions with Most Publications***

Hacettepe Üniversitesi (n=8, 1.30%), Middle East College (n=8, 1.30%), King Abdulaziz University (n=7, 1.14%), and Amity University (n=7, 1.14%) are the universities that have produced more than 1% of the total publications in the past decade. Table 10 summarizes the top 20 institutions that contributed the most to the publications on EDM in online learning.

Table 10. Top 20 Most-Contributed Institutions

| Institutions | Number of Documents | Percentage (%) |
|---|---|---|
| Hacettepe Üniversitesi | 8 | 1.30 |
| Middle East College | 8 | 1.30 |
| King Abdulaziz University | 7 | 1.14 |
| Amity University | 7 | 1.14 |
| Panepistimion Makedonias | 6 | 0.98 |
| Universidad de Córdoba | 6 | 0.98 |
| Universidad de Oviedo | 6 | 0.98 |
| Universitat d'Alacant | 6 | 0.98 |
| Kyushu University | 6 | 0.98 |

| Institutions | Number of Documents | Percentage (%) |
|---|---|---|
| Universidad de las Americas - Ecuador | 6 | 0.98 |
| Eastern Macedonia and Thrace Institute of Technology | 5 | 0.81 |
| Université Sidi Mohamed Ben Abdellah | 5 | 0.81 |
| Athabasca University | 5 | 0.81 |
| University of Science and Technology of China | 5 | 0.81 |
| University of Rijeka | 5 | 0.81 |
| Democritus University of Thrace | 5 | 0.81 |
| University of Patras | 5 | 0.81 |
| Aichi University | 5 | 0.81 |
| Telkom University | 5 | 0.81 |
| Universidad a Distancia de Madrid | 5 | 0.81 |

***Authorship Analysis***

Ioannis Kazanidis, affiliated with International Hellenic University, Greece, and Stavros Valsamidis of Eastern Macedonia and Thrace Institute of Technology, Greece, produced the most publications on EDM in online learning with seven publications or 1.14% each. This is followed by Sotirios Kontogiannis from the University of Ioannina, Greece; Cristobal Romero from Universidad de Córdoba, Spain; and William Villegas-Ch from Universidad Latina de Costa Rica, Costa Rica, with 6 (0.98%) each. Table 11 shows the top 20 most productive authors.

Table 11. Top 20 Most-Contributed Authors

| Authors | Number of Documents | Percentage (%) |
|---|---|---|
| Kazanidis, I. | 7 | 1.14 |
| Valsamidis, S. | 7 | 1.14 |
| Kontogiannis, S. | 6 | 0.98 |
| Romero, C. | 6 | 0.98 |
| Villegas-Ch, W. | 6 | 0.98 |
| Bansal, A. | 5 | 0.81 |
| Cerezo, R. | 5 | 0.81 |
| Dobashi, K. | 5 | 0.81 |
| Hasan, R. | 5 | 0.81 |
| Hirokawa, S. | 5 | 0.81 |
| Huang, Z. | 5 | 0.81 |
| Luján-Mora, S. | 5 | 0.81 |
| Su, Y. | 5 | 0.81 |
| Sweta, S. | 5 | 0.81 |
| Abbas, A. | 4 | 0.65 |

| Authors | Number of Documents | Percentage (%) |
|---|---|---|
| Araka, E. | 4 | 0.65 |
| Chen, E. | 4 | 0.65 |
| Distante, D. | 4 | 0.65 |
| Gitonga, R. | 4 | 0.65 |
| Karagiannis, I. | 4 | 0.65 |

This study used VOSviewer software to produce network visualization showing co-authorship mapping among authors. Due to the very particular search query performed for the bibliometric analysis in this study that led to minimal yet specific results, this mapping used the fractional counting method and data of those authors who have at least one document on EDM in online learning with at least one citation.

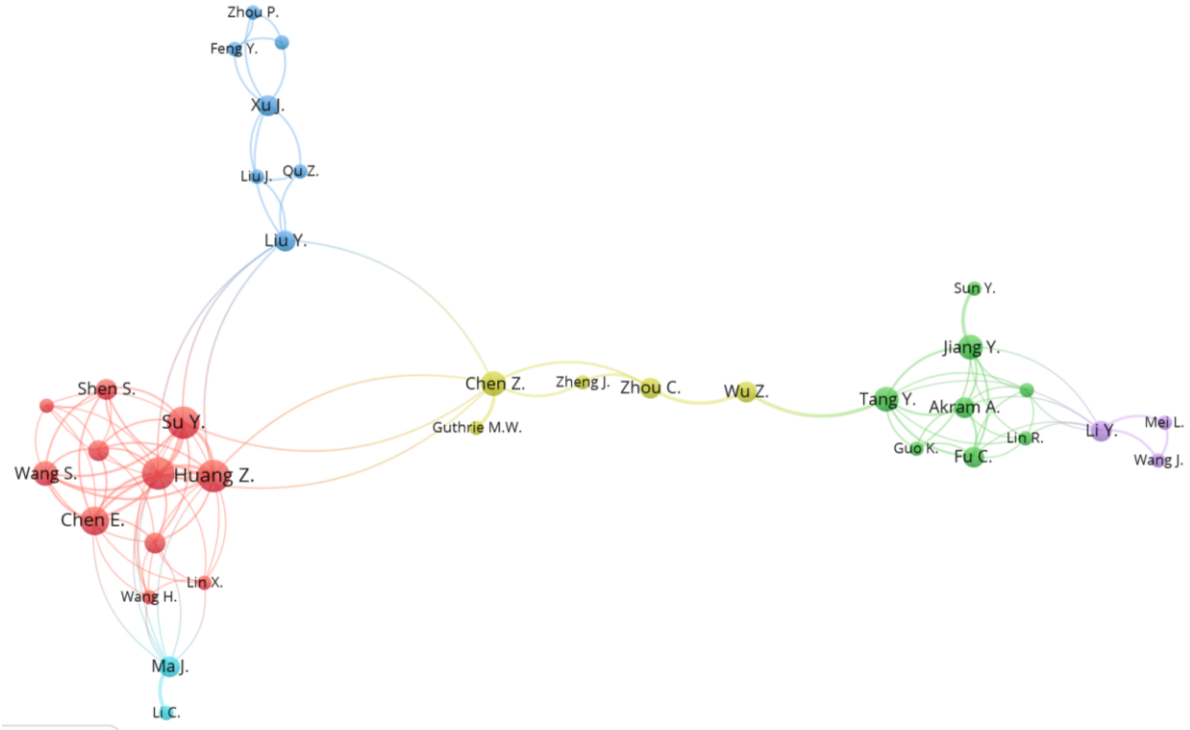


Figure 5. The Network Visualization Map of Co-authorship among Authors

According to the network visualization, the color, circle size, font size, and thickness of connecting lines represent the strength of collaborations among authors. Associated authors are typically clustered together as indicated by the same color. For example, Figure 5 depicts Chen Z., Zhou C., Wu Z., and Guthrie M. W. collaborating closely for EDM in online learning publications. The overall network visualization of the co-authorship mapping is viewable at https://tinyurl.com/3fm4syxn.

Figure 6 shows the network visualization map of co-authorship among authors by country. The network visualization map shows active cross-countries co-authorship. India, the United States, and China are collaborating. However, Spain and China have yet to establish any collaboration. The detailed findings are viewable at https://tinyurl.com/5yfe64z9.

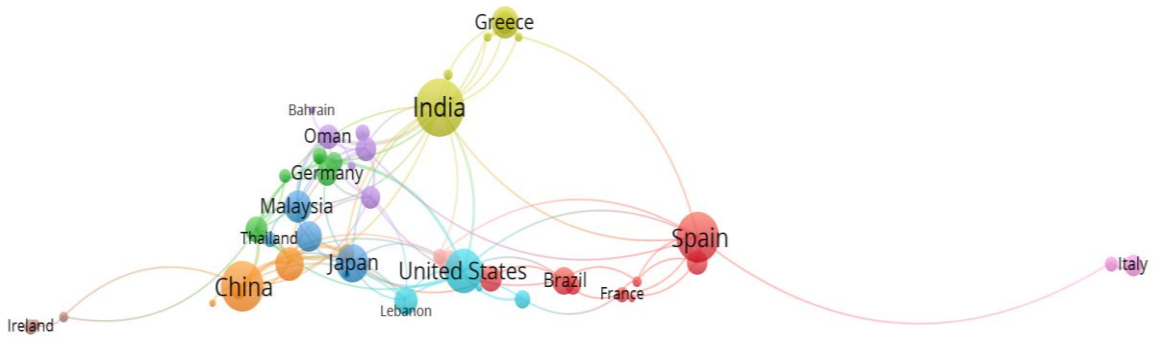


Figure 6. The Network Visualization Map of Co-authorship by Country

Findings revealed that India, Spain, China, and the United States are the top four most contributed countries. Moreover, the top four institutions contributing to the publications are Hacettepe University in Turkey, Middle East College in Oman, King Abdulaziz University in Saudi Arabia, and Amity University in India. Collaboration among researchers, regardless of local or international, is a crucial relationship for providing strategic insights into how best to design, manage, and implement EDM in e-learning and eventually facilitate electronic educational systems (Slater & Ravid, 2010; Ichou, 2018; Wahid et al., 2020). Also, collaborations among countries, institutions, and scholars are often associated with transferring knowledge and technology, which is extremely important for the world's economies (Wahid et al., 2020). For example, the visualization map of the collaboration among authors by country shows that in current affairs, authors who focus on EDM in online learning from Spain work closely with authors from the United States, India, Saudi Arabia, and Italy. Nevertheless, countries such as China, Japan, Brazil, Greece, and Turkey also play a crucial role in the publications of EDM in e-learning. India, the United States, and China are collaborating. However, Spain and China have yet to establish any collaboration. Further collaborations between countries enable the discovered knowledge from different researchers and studies in different contexts to be transferred, combined, and reinforced. Furthermore, countries can better understand the overview of EDM in online learning across geographical boundaries through international collaboration for global trend analysis and cross-cultural studies.

## Conclusion

This study indicates an increase in publications on EDM in online learning over the past decade, with conference papers being the most prevalent format. Journal article publications are expected to increase as the field matures. The increased occurrences of EDM in online learning publications in educational technology-related journals infers the robustness of the data mining techniques derived from the computer science field and, hence, the feasibility of adopting these techniques in online learning. As this bibliometric study reveals existing studies focusing on the learning analytics aspect of LMS, future research can expand the current focus to incorporate data mining techniques into LMS. Near future studies can also harness the benefits of less conventional data mining techniques, such as time series analysis, natural language processing, network analysis, and outlier detection.

The keyword co-occurrence analysis highlights data mining in e-learning systems, data mining techniques, data

mining for teaching and learning, students and learning materials, data-driven decisions and insights, and learning analytics as current research trends. In addition to these trends, future researchers may also look into other research focuses, such as the ethical issues surrounding EDM in online learning for responsible data handling, techniques to create universally accessible online content by considering learners of diverse learning needs, integrating EDM with other emerging technologies such as augmented reality and virtual reality to gain insights into leveraging these technologies for engaging online learning, and develop real-time EDM support systems for different educational stakeholders to provide timely insights and interventions for enhancing online learning. This study also reveals that highly cited publications are relevant to the current research on personalized learning and providing learning analytics and EDM access to relevant educational stakeholders. The co-authorship by country analysis also points to limited collaborations among authors of different countries, and hence, intensifying international collaborations is necessary for scalability and inclusivity in the field.

This study focuses on bibliometric analysis. Future studies can consider embedding other methods to understand the field of EDM in e-learning from different perspectives. For example, text mining can transform unstructured text into a structured format to identify meaningful patterns and new insights from the retrieved publications. Future studies can also employ a systematic literature review to identify all empirical evidence that fits the pre-specified inclusion criteria to answer a particular research question of EDM in online learning. The current study retrieved publications from the Scopus database. Hence, future work can include the Web of Sciences (WOS), PubMed, and Google Scholar databases and conduct comparative or combined databases for bibliometrics analysis study of EDM in online learning.

## Acknowledgements

The author acknowledges the support provided by Universiti Malaysia Sarawak through the Malaysia Comprehensive University Network Grant Scheme [GL/F04/MCUN/20/2020].

## References

Abana, E. (2019). A decision tree approach for predicting student grades in research project using Weka. *International Journal of Computer Science and Applications, 10*(7), 285–289.

Abdous, M., He, W., & Yen, C. (2012). Using data mining for predicting relationships between online question theme and final grade. *Educational Technology and Society*, *15*(3), 61-70.

Abhirami, K., & Devi, M. K. (2022). Student behavior modeling for an e-learning system offering personalized learning experiences. *Comput. Syst. Sci. Eng.*, *40*(3), 1127-1144.

Adejo, O. W., & Connolly, T. (2018). Predicting student academic performance using multi-model heterogeneous ensemble approach. *Journal of Applied Research in Higher Education*, *10*(1), 61-75.

Adeniji, B. A. (2019). *A bibliometric study on learning analytics* (Doctoral dissertation, Long Island University, CW Post Center).

Aher, S., & Lobo, L. (2013). Combination of machine learning algorithms for recommendation of courses in e-learning system based on historical data. *Knowledge-Based Systems, 51*, 96-104.

Ahmi, A. (2021). *Bibliometric analysis for beginner*. Malaysia: UUM.

Amrieh, E. A., Hamtini, T., & Aljarah, I. (2015). Preprocessing and analyzing educational data set using X-API for improving student's performance. In *2015 IEEE Jordan conference on applied electrical engineering and computing technologies (AEECT)* (pp. 1-5). IEEE.

Bakhshinategh, B., Zaiane, O., ElAtia, S., & Ipperciel, D. (2018). Educational data mining applications and tasks: A survey of the last 10 years. *Education and Information Technologies, 23*(1), 537–553.

Bharara, S., Sabitha, S., & Bansal, A. (2018). Application of learning analytics using clustering data Mining for Students' disposition analysis. *Education and Information Technologies*, *23*, 957-984.

Bienkowski, M., Feng, M., & Means, B. (2012). Enhancing Teaching and Learning through Educational Data Mining and Learning Analytics: An Issue Brief. *Office of Educational Technology, US Department of Education*.

Burgos, C., Campanario, M. L., de la Peña, D., Lara, J. A., Lizcano, D., & Martínez, M. A. (2018). Data mining for modeling students' performance: A tutoring action plan to prevent academic dropout. *Computers & Electrical Engineering*, *66*, 541-556.

Cerezo, R., Sánchez-Santillán, M., Paule-Ruiz, M. P., & Núñez, J. C. (2016). Students' LMS interaction patterns and their relationship with achievement: A case study in higher education. *Computers & Education*, *96*, 42-54.

Chang, Y. C., Li, J. W., & Huang, D. Y. (2022). A personalized learning service compatible with Moodle e-learning management system. *Applied Sciences, 12*(7), 3562.

Del Blanco, Á., Serrano, Á., Freire, M., Martínez-Ortiz, I., & Fernández-Manjón, B. (2013). E-Learning standards and learning analytics. Can data collection be improved by using standard data models?. In *2013 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1255-1261). IEEE.

Dondorf, T., Pyka, C., Gramlich, R., Sewilam, H. & Nacken, H. (2019). Learning analytics software implementation for the moodle learning management system. In *12th Annual International Conference of Education, Research and Innovation* (pp. 6957-6964). IATED.

Dormezil, S., Khoshgoftaar, T. M., & Robinson-Bryant, F. (2019). Differentiating between Educational Data Mining and Learning Analytics: A Bibliometric Approach. In *EDM (Workshops)* (pp. 17-22).

Dutt, A., Ismail, M.A., & Herawan, T. (2017). A systematic review on educational data mining. *IEEE Access, 5*, 15991–16005.

Elbadrawy A., Polyzou A., Ren Z., Sweeney M., Karypis G., & Rangwala H. (2016). Predicting student performance using personalized analytics. *Computer, 49*(4), 61–69.

Estacio, R., & Raga Jr, R. (2017). Analyzing students online learning behavior in blended courses using Moodle. *Asian Association of Open Universities Journal, 12*(1), 52-68.

Fernández, A., Peralta, D., Herrera, F., & Benítez, J. M. (2012). An overview of e-learning in cloud computing. *Advances in Intelligent Systems and Computing*, 35–46.

Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K. U., & Sattar, M. U. (2020). Predicting student performance in higher educational institutions using video learning analytics and data mining techniques. *Applied Sciences, 10*(11), 3894.

Helal, S., Li, J., Liu, L., Ebrahimie, E., Dawson, S., Murray, D., & Long, Q. (2018). Predicting academic performance by considering student heterogeneity. *Knowledge-Based Systems*, *161*, 134-146.

Hu, Y., Lo, C., & Shih, P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior, 36*, 469-478.

Huebner, R. A. (2013). A survey of educational data mining research. *Research in Higher Education Journal, 19*, 1-13.

Ichou, R. P. (2018). Can MOOCs reduce global inequality in education? *Australasian Marketing Journal*, *26*(2), 116–120.

Juhanak, L., Zounek, J., & Rohlikova, L. (2019). Using process mining to analyse students quiz taking behaviour patterns in a learning management system. *Computers in Human Behaviour, 92*, 496-506.

Kabakchieva, D. (2013). Predicting student performance by using data mining methods for classification. *Cybernetics and Information Technologies, 13*(1), 61–72.

Krouska, A., Troussas, C., & Virvou, M. (2016). The effect of preprocessing techniques on Twitter sentiment analysis. In *2016 7th International Conference on Information, Intelligence, Systems & Applications (IISA)* (pp. 1-5). IEEE.

Lara, J. A., Lizcano, D., Martínez, M. A., Pazos, J., & Riera, T. (2014). A system for knowledge discovery in e-learning environments within the European Higher Education Area–Application to student data from Open University of Madrid, UDIMA. *Computers & Education*, *72*, 23-36.

Lee, J., & Recker, M. (2022). The effects of instructors' use of online discussions strategies on student participation and performance in university online introductory mathematics courses. *Computer and Education, 162*, 104084.

Ortigosa, A., Martín, J. M., & Carro, R. M. (2014). Sentiment analysis in Facebook and its application to e-learning. *Computers in Human Behavior*, 31, *527-541*.

Park, Y., Yu, J. H., & Jo, I. H. (2016). Clustering blended learning courses by online behavior data: A case study in a Korean higher education institute. *The Internet and Higher Education*, *29*, 1-11.

Papamitsiou, Z., & Economides, A. (2014). Learning analytics and educational data mining in practice: A systematic literature review of empirical evidence. *Educational Technology & Society*, *17*(4), 49-64.

Raj, N. S., & Renumol, V. G. (2022). A systematic literature review on adaptive content recommenders in personalized learning environments from 2015 to 2020. *Journal of Computers in Education*, *9*(1), 113-148.

Real, E. M., Pinheiro-Pimentel, L. V., de Oliveira, J., Cristina, B., & Stiubiener, I. (2020). Educational process mining for verifying student learning paths in an introductory programming course. In *2020 IEEE Frontiers in Education Conference (FIE)* (pp. 1-9). Sweden.

Realinho, V., Machado, J., Baptista, L., & Vieira Martins, M. (2022). Predicting student dropout and academic success. *Data*, *7*(11), 146.

Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web usage mining for predicting final marks of students that use Moodle courses. *Computer Applications in Engineering Education*, *21*(1), 135-146.

Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications, 33*, 135–146.

Romero, C., & Ventura, S. (2020). Educational data mining and learning analytics: An updated survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *10*(3).

Salloum, S. A., Alshurideh, M., Elnagar, A., & Shaalan, K. (2020). Mining in educational data: Review and future directions. In *Proceedings of the International Conference on Artificial Intelligence and Computer Vision (AICV2020)* (pp. 92–102). Springer International Publishing.

Schwendimann, B., Rodriguez-Triana, M. J., Vozniuk, A., Prieto, L. P., Boroujeni, M. S., Holzer, A., Gillet, D., & Dillenbourg, P. (2017). Perceiving learning at a glance: A systematic literature review of learning dashboard research. *IEEE Transactions on Learning Technologies, 10*(1), 30-41.

Slater, J., & Ravid, R. (2010). *Collaboration in education* (2nd ed.). Taylor & Francis.

Tawfeeq, N., Ghazal, O., & Abed, W. (2022). Using data mining techniques to extract students' attitudes toward e-learning. *Indonesian Journal of Electrical Engineering and Computer Science, 28*(2), 1037–1048.

Usov, S., Safonov, M., Akbilek, E., & Sorokina, L. (2020). Swot analysis of Moodle platform application in the assessment of foreign language knowledge. In *The 4th International Conference on Education and Multimedia Technology* (pp. 31-34).

Wahid, R., Ahmi, A., & Alam, F. (2020). Growth and collaboration in massive open online courses. *International Review of Research in Open and Distributed Learning, 21*(4), 293-322.

Yu, Y., Li, Y., Zhang, Z., Gu, Z., Zhong, H., Zha, Q., Yang, L., Zhu, C., & Chen, E. (2020). A bibliometric analysis using VOSviewer of publications on COVID-19. *Annals of Translational Medicine, 8*(13), 816-827.

Zhang, Y., Ghandour, A., & Shestak, V. (2020). Using learning analytics to predict students performance in Moodle LMS. *International Journal of Emerging Technologies in Learning (iJET), 15*(20), 102.

## Author Information

**Nie Hui Ling**
https://orcid.org/0009-0008-2685-4158
Universiti Malaysia Sarawak
94300, Kota Samarahan, Sarawak
Malaysia

**Chwen Jen Chen**
https://orcid.org/0000-0001-5175-7060
Universiti Malaysia Sarawak
94300, Kota Samarahan, Sarawak
Malaysia
Contact e-mail: *cjchen@unimas.my*

**Chee Siong Teh**
https://orcid.org/0000-0003-2474-5958
Universiti Malaysia Sarawak
94300, Kota Samarahan, Sarawak
Malaysia

**Dexter Sigan John**
https://orcid.org/0000-0003-4096-101X
Universiti Malaysia Sarawak
94300, Kota Samarahan, Sarawak
Malaysia

**Looi Chin Ch'ng**
https://orcid.org/0000-0001-5280-1393
Universiti Teknologi MARA Sarawak
Jalan Meranek, 94300, Kota Samarahan, Sarawak
Malaysia

**Yoon Fah Lay**
https://orcid.org/0000-0002-5219-6696
Universiti Malaysia Sabah
Jalan UMS, 88400, Kota Kinabalu, Sabah
Malaysia