

# A critical review of data mining in education on the levels and aspects of education

Data mining in  
education

41

Inusah Fuseini

*Department of Mathematics and ICT Education, University for Development Studies, Tamale, Ghana and Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana, and*

Yaw Marfo Missah

*Department of Computer Science, Kwame Nkrumah University of Science and Technology, Kumasi, Ghana*

Received 13 January 2024  
Revised 2 May 2024  
Accepted 17 May 2024

## Abstract

**Purpose** – This systematic literature review aims to identify the pattern of data mining (DM) research by looking at the levels and aspects of education.

**Design/methodology/approach** – This paper reviews 113 conference and research papers from well-known publishers of educational data mining (EDM) and learning analytics-related research using a recognized literature review in computer science by Carrera-Rivera *et al.* (2022a). Two major stages, planning and conducting the review, were used. The databases of Elsevier, Springer, IEEE, SAI, Hindawi, MDPI, Wiley, Emerald and Sage were searched to retrieve EDM papers from the period 2017 to 2023. The papers retrieved were then filtered based on the application of DM to the three educational levels – basic, pre-tertiary and tertiary education.

**Findings** – EDM is concentrated on higher education. Basic education is not given the needed attention in EDM. This does not enhance inclusivity and equity. Learner performance is given much attention. Resource availability and teaching and learning are not given the needed attention.

**Research limitations/implications** – This review is limited to only EDM. Literature from the year 2017 to 2023 is covered. Other aspects of DM and other relevant literature published in EDM outside the research period are not considered.

**Practical implications** – As the current trend of EDM shows an increase in zeal, future research in EDM should concentrate on the lower levels of education to identify the challenges of basic education which serves as the core of education. This will enable addressing the challenges of education at an early stage and facilitate getting a quality education at all levels of education. Appropriate EDM techniques for mining the data at this level should be the focus of the research. Specifically, techniques that can cater for the variation in learner abilities and the appropriate identification of learner needs should be considered.

**Social implications** – Content sequencing is necessary in facilitating an easy understanding of concepts. Curriculum design from basic to higher education dwells much on this. Identifying the challenge of learning at

© Inusah Fuseini and Yaw Marfo Missah. Published by Emerald Publishing Limited. This article is published under the Creative Commons Attribution (CC BY 4.0) licence. Anyone may reproduce, distribute, translate and create derivative works of this article (for both commercial and non-commercial purposes), subject to full attribution to the original publication and authors. The full terms of this licence may be seen at <http://creativecommons.org/licenses/by/4.0/legalcode>

*Declaration of conflict of interest:* The authors declare that there is no known conflict of interest.

*Data availability:* Data sharing does not apply to this article as no datasets were generated or analysed during the current study.



---

the early stages will facilitate efficient learning. At the basic level of learning, data on learning should be collected by educational institutions just as it is done at the tertiary level. This will enable EDM to accurately identify the challenges and appropriate solutions to educational problems. Resource availability is a catalyst for effective teaching and learning. The attributes of a learner will enable knowing the true nature of the learner to determine the prospects of the learner.

**Originality/value** – This research has not been published in any journal. The information presented is the original knowledge of the authors. However, a pre-print of the work is in Research Square.

**Keywords** Basic education, Data mining, Tertiary, Pre-tertiary, Teaching and learning

**Paper type** Literature review

---

## 1. Introduction

Interestingly, data mining (DM) in education has played a tremendous role in the application of technology in education helping educational leaders to make accurate decisions (Inusah *et al.*, 2023a). To enhance more accuracy and consistency in educational data for efficient management, DM is needed (Inusah *et al.*, 2023b). Specifically, educational data mining (EDM) is the process of revealing hidden information from larger educational data for effective decision-making, while learning analytics looks at pedagogical content and the processes that yield to an outcome in learning. Learning analytics emphasises learning and the outcome (Lemay *et al.*, 2021). The significance of EDM in education is well noted in academic performance of learners through resources management, learner attributes. The application of EDM is therefore a necessity for life-long learning. As a novel methodology for solving problems in education, much research conducted in this area is basically on pedagogy which enriches the quality of tuition (Sokkhey *et al.*, 2020). Deep learning as a trending EDM technique has gained much attention in tracing patterns and detailed solutions to educational problems (Hernández-Blanco *et al.*, 2019). In a three-decade analysis of EDM, the application of clustering algorithms and their applications, as well as the usability of EDM, has been reviewed (Ashraf *et al.*, 2020). Enhancing more accuracy of using this methodology is necessary. Research by Tsiakmaki *et al.* (2020) considers determining accuracy using the rightful DM techniques. Gamification as a trend for arousing the interest of learners is a novel technique that is combined with EDM to spice up effective teaching and learning (Daghestani *et al.*, 2020; Arend *et al.*, 2022). In Issah *et al.* (2023), the attributes determining the performance of learners using DM techniques are reviewed.

The purpose of this review is to find out the levels of education given attention in EDM and the aspect of education which is usually considered. The research looks at three educational levels: basic education, secondary education and tertiary education. Basic education which is the lower level of education looks at the first 14 years of education. It is seen as the foundation (core) of education. Secondary education looks at the institutions after basic education which is non-tertiary. Tertiary education denotes the higher levels. The various aspects of education that are usually considered at the various levels consist of teaching and learning activities as well as the resources for learning which yields the performance in education. Attributes of learners and many other aspects of learning are considered. Overwhelming research on EDM stresses the attributes of the learner in determining academic performance. The following research questions guide the review:

- RQ1. What is the current trend in EDM that affects effective teaching and learning?
- RQ2. Which level of education is given much attention in EDM research?
- RQ3. Which aspect of education is given much attention in EDM to enhance effective learning?

These questions are necessary because the focus of recent trends in EDM regarding levels and aspects of education is the foundation for higher education. Also, resource availability and learner attributes contribute significantly to a learner's academic growth and development. The remaining part of the review is organized into six areas. The first part is the related reviews on EDM which comprises the EDM in education, the levels of education (basic, secondary and tertiary) and the major aspects of education (academic performance, teaching and learning resources and learner attributes). The second aspect is the methodology used in the literature search criteria which comprises how the search was conducted and the search strategy. The publication selection looks at inclusion and exclusion criteria. The third aspect is the presentation of the results for the literature used. The fourth aspect is the discussion of the research gap identified which looks at both the population gap and the knowledge gap. The fifth aspect is the conclusions drawn and the sixth is the recommendations to stakeholders of education. [Table 1](#) is related to reviews in EDM.

## 2. Related reviews on educational data mining

In [Ashraf \*et al.\* \(2020\)](#) and [Dutt \*et al.\* \(2017\)](#), the use of clustering algorithms and their application in EDM is specified using a trend of three decades. The trends of DM shortly are also included in this work. Deep learning approaches to EDM are reviewed to know the task that has already been taken and the future directions ([Hernández-Blanco \*et al.\*, 2019](#)). At the higher education level, a comprehensive review of EDM is conducted to enable potential authors to know the effects on students' performance and the possible direction to improve effective teaching and learning ([Aldowah \*et al.\*, 2019](#)). Focusing on only classroom learning, ([Khan and Ghosh, 2021](#)) used necessary predictors of time, methods and aims to identify significant performance predictions during a course. In a related paper, [Alshareef \*et al.\* \(2020\)](#) recommend using extended tools to measure and predict students' performance in teaching and learning. Using classification models, two attributes – grades and absence – are seen as the most significant in determining the performance of students in a year ([Fernandes \*et al.\*, 2019](#)). Video analytics and DM techniques are used in predicting students' performance at the higher education level to know the level of accuracy of the techniques applied ([Hasan \*et al.\*, 2020](#)). In [Du \*et al.\* \(2020\)](#), extensive review of EDM at the various levels of education reveals the rapid application of DM techniques across the levels of education but rare information available on K–12 which is the basic or early level of education. The potential influence of DM in identifying the learning process and progress is identified in higher education ([Aldowah \*et al.\*, 2019](#)).

## 3. Methodology

This paper reviews 113 conference and research papers from well-known publishers of EDM and learning analytics-related research using a recognized literature review in computer science by [Carrera-Rivera \*et al.\* \(2022\)](#). Two major stages, planning and conducting the review, were used. The databases of Elsevier, Springer, IEEE, SAI, Hindawi, MDPI, Wiley, Emerald and Sage were searched to retrieve EDM papers from the period 2017 to 2023. The papers retrieved were then filtered based on the application of DM to the three educational levels; Basic, pre-tertiary and tertiary education. Furthermore, the papers were filtered based on three aspects of education; student performance, teaching and learning resources and attributes of learners. In all, 20 conference papers and 93 research articles met the criteria.

**Table 1.**  
Related reviews in  
educational data  
mining

Author(s)	Title	Years covered	Databases explored	Papers used
Mohamad and Tasir (2013) Du <i>et al.</i> (2020)	Educational data mining: a review Educational data mining: a systematic review of research and emerging trends	2004–2012 2007–2019	<a href="http://www.educationaldatamining.org">www.educationaldatamining.org</a> Web of Science, Proceedings of the Educational Data Mining conference, <i>Journal of Educational Data Mining</i> , ACM Digital Library and IEEE Xplore	15 217
Rodrigues <i>et al.</i> (2018)	Educational data mining: a review of the evaluation process in the e-learning	1994–2016	IEEE, Elsevier, International Educational Data- Mining Society, ACM Digital Library, EBSCO	72
Dutt <i>et al.</i> (2017)	A systematic review on educational data mining	1983–2016	IEEE Xplore, ACM Digital, JEDM ( <i>Journal of Educational Data Mining</i> ), ProQuest and ScienceDirect	166
Chango <i>et al.</i> (2022)	A review of data fusion in multimodal learning analytics and educational data mining	Earlier–2021	Google Scholar, Web of Science and Scopus	42
Hernández-Blanco <i>et al.</i> (2019)	A systematic review of deep learning approaches to educational data mining	2015–2018	ACM Digital Library ( <a href="https://dl.acm.org/">https://dl.acm.org/</a> ), Google Scholar ( <a href="https://scholar.google.es/">https://scholar.google.es/</a> ) and IEEE Xplore	41
Shafiq <i>et al.</i> (2022)	Student retention using educational data mining and predictive analytics: a systematic literature review	2017–2021	IEEE, Springer, Elsevier, Google Scholar, Research Gate, MDPI, Taylor and Francis, Semantic Scholar, ACM Library, Emerald, Wiley, IOP science	100
Okewu <i>et al.</i> (2021)	Artificial neural networks for educational data mining in higher education: a systematic literature review	2010–2018	Web of Science, Scopus, IEEE Xplore, ACM Digital Library and Springer link	190
Khan and Ghosh (2021)	Student performance analysis and prediction in classroom learning: a review of educational data mining studies	2000–2018	Google Scholar	140
Aldowah <i>et al.</i> (2019)	Educational data mining and learning analytics for 21st century higher education: a review and synthesis	2000–2017	Scopus, Web of Science, Google Scholar, ERIC, Science Direct, DBLP and ACM Digital Library, IEEE Xplore, Springer Link	402

**Source:** Authors' construct

---

### 3.1 Planning

After the initial idea formulation, six major steps were considered in the planning of the literature search. The first step was to define the PICOC and synonyms for clarity and accuracy in identifying the literature. The second step was to formulate the research questions that guide the literature review. This was followed by the selection of digital library sources. The fourth step was to define the inclusion and exclusion criteria. The fifth stage was to design the quality assessment checklist and the last step was designing the data extraction form.

### 3.2 Conducting the search

After the planning of the literature stage, the next stage is the conducting of a literature search. This comprises five major steps. The first step is to build digital library search strings. This was to get the specific publications from the databases selected. The second step was to gather the studies from the digital libraries using the search strings specified in the first step. All the nine databases selected were thoroughly searched to retrieve the information. The third step was to do a study selection and refinement. Three conditions were used in this stage for selection. The first condition was to ensure there was no duplicate. The next was to evaluate the inclusion and exclusion criteria and the last was to assign quality scores. After these conditions were met in the third step, the fourth step was to do data extraction from the selected papers and a final analysis and report presentation was done in the last stage.

To identify relevant papers for the criteria, a search phrase was necessary. The phrase “educational data mining” or “data mining in education” was used to locate the related literature. Further filtering by specifying the years of publication was done to limit the retrieved papers to the most relevant works from the year 2017 to 2023. Papers that were applied to the various levels of education were further filtered by manually reading the abstract and the methodology to know what the paper was about. Concentration was given to the aspect of education applied in the research; student performance, teaching and learning resources and learner attributes.

*3.2.1 Search strategies.* A Boolean search was used to locate all necessary publications before filtering. The search string included OR, NOT and AND to enable specifically getting the relevant literature. In all nine databases of well-known publishers, a thorough search was conducted. The nine publishers and the literature rating are in [Table 1](#). The search strings used in the search can be seen below (“educational data mining” or “data mining in education” or “education and data mining”) AND (“learner attributes” or “performance” or “level of learning”) NOT (“learning systems”).

### 3.3 Publication selection

Duplicates were checked to ensure the same information from different databases was not included more than once. This was done by the reading of titles, abstracts and the authors as well as the year of publication.

*3.3.1 The inclusion criteria.* The most relevant papers to the research were from only conferences and articles from journal publications in the specified publishers. Specifically, research on how EDM is applied to a level of education and an aspect of education was the inclusion criteria.

*3.3.2 Exclusion criteria.* Three significant steps were used to exclude research works that were not most recently published and relevant to the research.

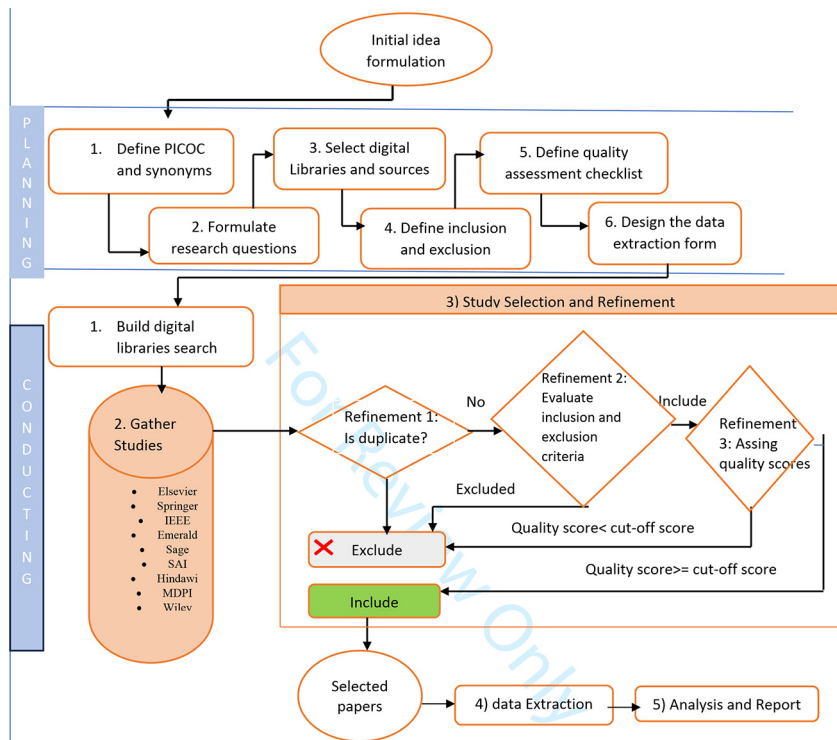
- Papers that are not relevant to the objectives of the research.

- Papers that are not published from 2017 to 2023.
- Papers that do not point to a particular level of education and the aspects of education.

3.3.3 *Selecting primary sources.* The selection process included the string “educational data mining” and the use of keywords; education, data and mining. Papers that did not have the string phrase but included the keywords specified were thoroughly read to know the full-text content and include or exclude. To avoid errors and duplications, the primary reviewer’s work was reviewed by a second reviewer to ascertain its accuracy and precision. Random selection of the papers under the various categorizations was used as a criterion by the second reviewer. This methodology is the adoption of [Carrera-Rivera et al. \(2022\)](#) guide for systematic literature review in computer science. [Fig 1](#) is the flowchart.

#### 4. Results

Revelations from the investigation into the already existing literature produced an interesting trend for publications in EDM. For clarity and consistency in the result presentation, the results and findings of the review are presented in line with the research questions. [Table 2](#) is the source and results of the literature search.



**Figure 1.** Flowchart for literature guide in a systematic review

Source: (Carrera-Rivera et al. 2022)

**Table 2.**

Data source and results for literature search

No.	Publisher	Source	Total hits	First filtering	Final filtering
1	Elsevier	<a href="http://www.sciencedirect.com/">www.sciencedirect.com/</a>	105	53	16
2	Springer	<a href="http://www.springer.com/gp">www.springer.com/gp</a>	98	65	13
3	IEEE	<a href="http://www.ieee.org/">www.ieee.org/</a>	120	84	29
4	SAI	<a href="https://thesai.org">https://thesai.org</a>	51	23	7
5	Hindawi	<a href="http://www.hindawi.com/">www.hindawi.com/</a>	72	31	10
6	MDPI	<a href="http://www.mdpi.com/">www.mdpi.com/</a>	20	10	12
7	Wiley	<a href="http://www.wiley.com/en-gb">www.wiley.com/en-gb</a>	15	8	4
8	Emerald	<a href="http://www.emeraldgroupublishing.com">www.emeraldgroupublishing.com</a>	21	5	7
9	Sage	<a href="https://sagepub.com">https://sagepub.com</a>	34	23	15
Total			536	317	113

Source: Authors' construct

From [Table 2](#), the number of databases searched for the literature totalled nine. These databases are for identified academic publishers of EDM for this work. The total number of responses gotten from the search is 536. After the first filtering, 317 papers were qualified. When the final filtering was done, only 113 papers were qualified to be used in the review. These papers were those that met the inclusion criteria specified in the methodology. [Table 2](#) is the categorization of the qualified papers into the levels and aspects of education. Both the basic level which is seen as the core has insignificant publications of nine which is about 7.96%, the pre-tertiary (secondary) also has nine which is about 7.96% and the tertiary has 95 which is about 84%. Specifically, for the aspects of education, academic performance at the basic level recorded only one publication, teaching and learning resource management recorded seven and learner attributes recorded only one. For the pre-tertiary level, academic performance was recorded six, teaching and learning resource management recorded one and learner attributes recorded two. The tertiary level which received overwhelming recognition had 47 for academic performance, 32 for teaching and learning resource management and 16 for learner attributes.

## 5. Discussion

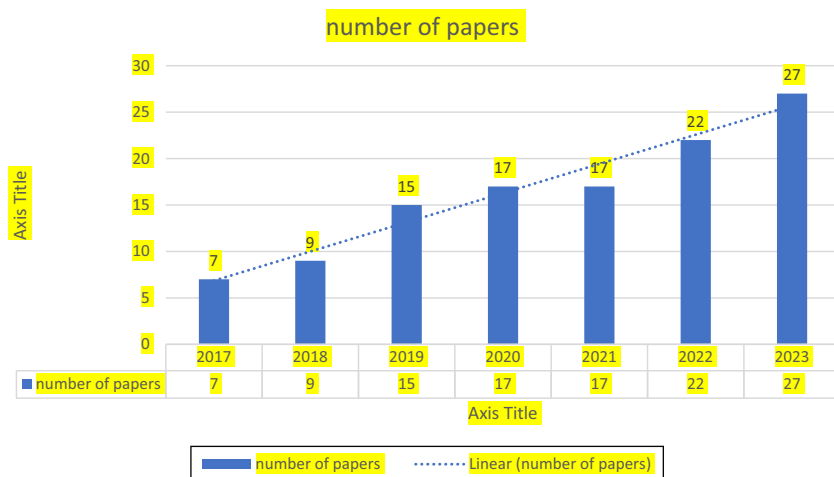
Conscious efforts are made by researchers such as [Inusah \*et al.\* \(2021\)](#) in tracing problems of education using EDM as indicated in [Table 3](#). However, the current trend in the rise of EDM with the relegation of basic education raises a concern (see [Figures 2](#) and [3](#)). This supports the [Jiang \(2022\)](#) assertion that researchers do not have enough information to predict the challenges of basic education. With [Yang \*et al.\* \(2022\)](#) alone spotted as the literature for performance and [Francis and Babu \(2019\)](#) as the only literature on learner attributes at the basic level, the results presented in [Table 3](#) confirm the insufficient publications though the publications are limited to full manuscript journal research and conference papers. There is enough evidence to believe that the basic level is relegated. Resource management and utilization are essential for effective teaching and learning ([Inusah \*et al.\*, 2022](#)). The basic (foundation) of the success of education in every nation has an insignificant number of research papers in EDM. Challenges of education can be identified at this level of education for lifelong learning of the citizenry. Accurate planning and prediction of the needs of learners in the future can be done to improve teaching and learning. If the foundation is weak, then the entire structure of education will be weaker. Learner attributes must be carefully identified at this stage and used to improve learning for good education. Proper storage of information in schools at the higher levels of education in databases makes it

QEA 1,2	Level	Aspect	Published papers	No. of papers
48	Basic	Academic performance	Yang <i>et al.</i> (2022)	1
		Teaching and learning resource management	Mushtaq <i>et al.</i> (2019); Inusah <i>et al.</i> (2022); Inusah <i>et al.</i> (2021); Jiang (2022); Tsiakmaki <i>et al.</i> (2020); Torcate and De Oliveira Rodrigues (2021); Marfo Missah <i>et al.</i> (2023)	7
	Pre-tertiary	Learner attributes	Francis and Babu (2019)	1
		Academic Performance	Amjad <i>et al.</i> (2022); Abu Zohair (2019); Rebai <i>et al.</i> (2020); Livieris <i>et al.</i> (2018); Abideen <i>et al.</i> (2023); Alghamdi and Rahman (2023)	6
		Teaching and learning resource management	Arfaee <i>et al.</i> (2022)	1
	Tertiary	Learner attributes	Shao <i>et al.</i> (2020), Al-Barrak (2016)	2
		Academic performance	Francis and Babu (2019); Asif <i>et al.</i> (2017); Haryani <i>et al.</i> (2022); Yağcı (2022); Dabhade <i>et al.</i> (2021); Hung <i>et al.</i> (2020); Aldowah <i>et al.</i> (2019); Kurdi <i>et al.</i> (2018); Trakunphutthirak and Lee (2021); Hassan <i>et al.</i> (2020); Wu (2022a, 2022b, 2022c); Hassan <i>et al.</i> (2024); Feng <i>et al.</i> (2022); Abu Zohair (2019); Sokkhey <i>et al.</i> (2020); Tsiakmaki <i>et al.</i> (2020); Fernandes <i>et al.</i> (2019); Tasnim <i>et al.</i> (2019); Ali and Hanna (2022); Dutt <i>et al.</i> (2017); Karthikeyan <i>et al.</i> (2020); Sultana <i>et al.</i> (2017); Adejo and Connolly (2018), (Adekitan <i>et al.</i> (2019); Cardona <i>et al.</i> (2023); Ashraf <i>et al.</i> (2020); Chaturvedi (2017); Roslan <i>et al.</i> (2023); Ordoñez-Avila <i>et al.</i> (2023); Sánchez <i>et al.</i> (2023); Chytas <i>et al.</i> (2023); Yürüm <i>et al.</i> (2023); Brandon (2023); Tan (2023); Sun (2023); Sghir <i>et al.</i> (2023); Guleria and Sood (2023); Jha (2023); Okewu <i>et al.</i> (2021); Aman <i>et al.</i> (2019); Al-Obeidat (2017); Mengash (2020); Yakubu and Abubakar (2021); Widyaningsih <i>et al.</i> (2019); Adekitan and Noma-Osaghae (2019); Lin <i>et al.</i> (2023); Rodrigues <i>et al.</i> (2018); Khan and Ghosh (2021)	47
	Teaching and learning resource management	Hung <i>et al.</i> (2020); Arfaee <i>et al.</i> (2022) (Chu <i>et al.</i> (2022); Wu (2022a, 2022b, 2022c); Li <i>et al.</i> (2022); Zhao (2022); Wu (2022a, 2022b, 2022c); Peng and Tang (2022); Shafiq <i>et al.</i> (2022); Zhou and Wu (2022); Prada <i>et al.</i> (2020); Daoudi <i>et al.</i> (2021); Mazidi and Abusham (2018); Tang <i>et al.</i> (2018); Chen (2019); Du <i>et al.</i> (2023); H. Wang (2023); Du <i>et al.</i> (2022); Tang <i>et al.</i> (2018); Biao (2023); Contreras-Luján <i>et al.</i> (2022); Zhang and Qin (2018); Jiang and Zeng (2019); Lv (2021) (Hu and Guo (2018); Shayakhmetova <i>et al.</i> (2021); Yang (2021); Bowen (2021); Paredes <i>et al.</i> (2021); Joshy <i>et al.</i> (2023); Shaik <i>et al.</i> (2023); Alam (2023)	32	
	Learner attributes	Nascimento <i>et al.</i> (2018); Leitner <i>et al.</i> (2019); Martinez-Abad <i>et al.</i> (2020); Xing and Du (2019); Cardona <i>et al.</i> (2020); Hung <i>et al.</i> (2020); Wang <i>et al.</i> (2020); Davies <i>et al.</i> (2021); Xilin and Honglian, 2021); Wang, 2018); Aldowah <i>et al.</i> (2019); Mohd Talib <i>et al.</i> (2023); Rafique <i>et al.</i> (2023); Zárate <i>et al.</i> (2023); Gera <i>et al.</i> (2024); Wongvorachan <i>et al.</i> (2023)	16	

**Table 3.**  
EDM as applied in  
levels and aspects

**Source:** Authors' construct

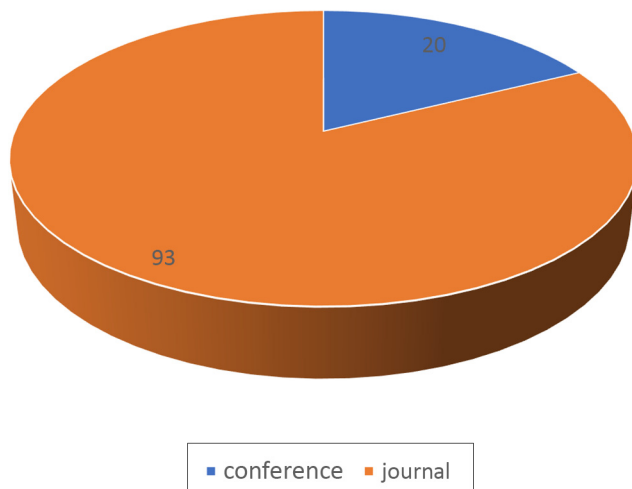




Source: Authors' construct

Figure 2.  
Number of publications by year

Number of papers



Source: Authors' construct

Figure 3.  
Papers by academic sources

relatively easier to analyse data. Pre-tertiary level of education has meaningful data available from learners and this information can be used to predict the likely occurrence of events in the schools using the attributes of the learner (Shao *et al.*, 2020) As spotted by Amjad *et al.* (2022), Yağcı (2022) and Khan and Ghosh (2021) among others, the academic

performance of learners at this level of education can easily be predicated using other factors. With only two spotted literature on resource management (Arfaee *et al.*, 2022; Rahman *et al.*, 2022a, 2022b), EDM for effective management of resources is insufficiently utilized. Unfortunately, this level of education is not given much attention in EDM. As the progress of learners to tertiary institutions is best determined at this level, the use of EDM to identify the appropriate programme a learner can pursue at the tertiary level should be done with ease. Challenges of education at this level can help both the basic and tertiary levels. Learners' performance at this level can help stakeholders in basic education to know how to prepare learners while the universities will also know how to deal with these learners.

Tertiary institutions are well organized and properly structured to make predictions easier (Francis and Babu, 2019; Yağci, 2022) which makes them suitable for EDM. This could be the reason for the overwhelming research in EDM at this level. A larger chunk of research on EDM is on tertiary institutions where independent learners and their attributes are used in predicting academic performance. In all aspects of EDM, the tertiary level has a majority of publications. This could also be attributed to the universal nature of the tertiary system. However, the research works at this level failed to look at the previous performance and attributes of learners at the secondary level and its effects on the learner at the tertiary level. Also, revisiting learners who complete institutions and return for further studies should be identified. This will help universities to know the true needs of the learner and the demands of society. As resources are very necessary for effective delivery, conscious efforts by researchers to predict the effects of resource availability on the performance of the learner are necessary (Inusah *et al.*, 2022). This will help institutions to carefully plan and provide educational resources that can boost the performance of learners. The attributes of learners which also affect the performance of learners should be looked at in the direction of programme choice which affects the occupations for employment. This will enable us to know the attributes associated with a particular occupation. In totality, the overwhelming research on EDM centred on the academic performance of learners which in some cases are replication and repetition of research works should be given a different direction. The factors determining the performance of a learner should be given a chance in EDM. The population gap identified in this research is the learners at the basic level. These groups of learners are neglected in EDM both in attributes and academic performance. They are not also considered in the provision and utilization of teaching and learning resources for effective learning. This creates a bigger challenge in education as they are not adequately prepared for the pre-tertiary and tertiary levels. The knowledge gap in this research is the limitation in the application of DM in education. Concentration is on the academic performance of learners with little or no attention to the learner attributes and teaching and learning resource management. Also the techniques used are limited to academic achievements with little or no information on the possibility of dropout or the cost of educating the learner.

### 5.1 Theoretical contribution

Content sequencing is necessary to facilitate an easy understanding of concepts. Curriculum design from basic to higher education dwells much on this. Identifying the challenge of learning at the early stages will facilitate efficient learning. At the basic level of learning, data on learning should be collected by educational institutions just as it is done at the tertiary level. This will enable EDM to accurately identify the challenges and appropriate solutions to educational problems. Resource availability is a catalyst for effective teaching and learning. The attributes of a learner will enable knowing the true nature of the learner to determine the prospects of the learner.

## 5.2 Practical implication

As the current trend of EDM shows an increase in zeal, future research in EDM should concentrate on the lower levels of education to identify the challenges of basic education which serves as the core of education. This will enable addressing the challenges of education at an early stage and facilitate getting a quality education at all levels of education. Appropriate EDM techniques for mining the data at this level should be the focus of the research. Specifically, techniques that can cater for the variation in learner abilities and the appropriate identification of learner needs should be considered.

## 6. Conclusion

This paper presents a thorough review of EDM by considering conference papers and articles from nine well-known publishers. All levels of education and research work are considered to know the application DM and the aspect of the application. The pattern of EDM is identified and the research gaps that need more attention in the area of EDM are identified. In summary, the tertiary level of education is given much attention in EDM at the expense of basic education for growth and development in education. The academic performance of learners is the most recognized aspect of education by researchers with little recognition of resources for a learner. Careful application of EDM at the lower levels of education can help in identifying the challenges of education to gain sustainable solutions for the improvement of the education sector and the development of nations as a whole.

## 7. Recommendation

Stakeholders in education must recognize basic education as the pivot for education and assist in the utilization of EDM methodologies to accurately identify the challenges and the appropriate interventions at this level. This is necessary to help enhance efficiency in educational management for effective teaching and learning using the available limited resources. It will also help in leveraging the success of education to other levels of education. Adequate provision of resources at this level will help in laying a foundation for the future of education. It will enhance inclusivity and equity to enable every child of school-going age to gain quality education. The progression of children of school-going ages will also be smooth as they have a good foundation of learning.

## References

- Abideen, Z. U., Mazhar, T., Razzaq, A., Haq, I., Ullah, I., Alasmay, H. and Mohamed, H.G. (2023), "Analysis of enrollment criteria in secondary schools using machine learning and data mining approach", *Electronics (Switzerland)*, Vol. 12 No. 3, doi: [10.3390/electronics12030694](https://doi.org/10.3390/electronics12030694).
- Abu Zohair, L.M. (2019), "Prediction of student's performance by modelling small dataset size", *International Journal of Educational Technology in Higher Education*, Vol. 16 No. 1, doi: [10.1186/s41239-019-0160-3](https://doi.org/10.1186/s41239-019-0160-3).
- Adejo, O.W. and Connolly, T. (2018), "Predicting student academic performance using multi-model heterogeneous ensemble approach", *Journal of Applied Research in Higher Education*, Vol. 10 No. 1, pp. 61-75, doi: [10.1108/JARHE-09-2017-0113](https://doi.org/10.1108/JARHE-09-2017-0113).
- Adekitan, A.I. and Noma-Osaghae, E. (2019), "Data mining approach to predicting the performance of first year student in a university using the admission requirements", *Education and Information Technologies*, Vol. 24 No. 2, pp. 1527-1543, doi: [10.1007/s10639-018-9839-7](https://doi.org/10.1007/s10639-018-9839-7).
- Adekitan, A.I., Salau, O. and Adekitan, A.I. (2019), "The impact of engineering students' performance in the first three years on their graduation result using educational data mining", *Heliyon*, Vol. 5 No. 2, p. e01250, doi: [10.1016/j.heliyon.2019](https://doi.org/10.1016/j.heliyon.2019).

- 
- Alam, A. (2023), "The secret sauce of student success: cracking the code by navigating the path to personalized learning with educational data mining", 2nd International Conference on Smart Technologies and Systems for Next Generation Computing, ICSTSN 2023, doi: [10.1109/ICSTSN57873.2023.10151558](https://doi.org/10.1109/ICSTSN57873.2023.10151558).
- Al-Barrak, M.A.-R. (2016), "Predicting students final GPA using decision trees: a case study", *International Journal of Information and Education Technology*, Vol. 6 No. 7, pp. 528-533.
- Aldowah, H., Al-Samarraie, H. and Fauzy, W.M. (2019), "Educational data mining and learning analytics for 21st century higher education: a review and synthesis", *In Telematics and Informatics*, Vol. 37, pp. 13-49, doi: [10.1016/j.tele.2019.01.007](https://doi.org/10.1016/j.tele.2019.01.007).
- Alghamdi, A.S. and Rahman, A. (2023), "Data mining approach to predict success of secondary school students: a Saudi Arabian case study", *Education Sciences*, Vol. 13 No. 3, doi: [10.3390/educsci13030293](https://doi.org/10.3390/educsci13030293).
- Ali, A.D. and Hanna, W.K. (2022), "Predicting students' achievement in a hybrid environment through self-regulated learning, log data, and course engagement: a data mining approach", *Journal of Educational Computing Research*, Vol. 60 No. 4, pp. 960-985, doi: [10.1177/07356331211056178](https://doi.org/10.1177/07356331211056178).
- Al-Obeidat, A.T.A.D.B.S. (2017), "Analyzing students' performance using multi-criteria classification", *Cluster Computing*, Vol. 21 No. 1, pp. 623-632.
- Alshareef, F., Alhakami, H., Alsubait, T. and Baz, A. (2020), "Educational data mining applications and techniques", *International Journal of Advanced Computer Science and Applications*, Vol. 11 No. 4, pp. 729-734, doi: [10.14569/IJACSA.2020.0110494](https://doi.org/10.14569/IJACSA.2020.0110494).
- Aman, F., Rauf, A., Ali, R., Iqbal, F. and Khattak, A.M. (2019), "A predictive model for predicting students academic performance", 10th International Conference on Information, Intelligence, Systems and Applications, IISA 2019, January 2020, doi: [10.1109/IISA.2019.8900760](https://doi.org/10.1109/IISA.2019.8900760)
- Amjad, S., Younas, M., Anwar, M., Shaheen, Q., Shiraz, M. and Gani, A. (2022), "Data mining techniques to analyze the impact of social media on academic performance of high school students", *Wireless Communications and Mobile Computing*, Vol. 2022, doi: [10.1155/2022/9299115](https://doi.org/10.1155/2022/9299115).
- Arend, D., Yuwono, S., Diprasetya, M.R. and Schwung, A. (2022), "MLPro 1.0 – standardized reinforcement learning and game theory in python", *Machine Learning with Applications*, Vol. 9, p. 100341, doi: [10.1016/j.mlwa.2022.100341](https://doi.org/10.1016/j.mlwa.2022.100341).
- Arfaee, M., Bahari, A. and Khalilzadeh, M. (2022), "A novel prediction model for educational planning of human resources with data mining approach: a national tax administration case study", *Education and Information Technologies*, Vol. 27 No. 2, pp. 2209-2239, doi: [10.1007/S10639-021-10699-6/TABLES/14](https://doi.org/10.1007/S10639-021-10699-6/TABLES/14).
- Ashraf, M., Zaman, M. and Ahmed, M. (2020), "An intelligent prediction system for educational data mining based on ensemble and filtering approaches", *Procedia Computer Science*, Vol. 167, pp. 1471-1483, doi: [10.1016/j.procs.2020.03.358](https://doi.org/10.1016/j.procs.2020.03.358).
- Asif, R., Merceron, A., Abbas Ali, S. and Ghani Haider, N. (2017), "Accepted manuscript analyzing undergraduate students' performance using educational data mining", *Computers and Education*, doi: [10.1016/j.compedu.2017.05.007](https://doi.org/10.1016/j.compedu.2017.05.007).
- Biao, Z.Z. (2023), "Design and realization of data mining simulation and methodological models", *Journal of King Saud University - Science*, Vol. 35 No. 10, doi: [10.1016/j.jksus.2023.102964](https://doi.org/10.1016/j.jksus.2023.102964).
- Bowen, D. (2021), "Construction of business English subject system based on data mining algorithm", IEEE 4th International Conference on Information Systems and Computer Aided Education, ICISCAE 2021, pp. 441-445, doi: [10.1109/ICISCAE52414.2021.9590650](https://doi.org/10.1109/ICISCAE52414.2021.9590650).
- Brandon, D. (2023), "Data mining twitter for COVID-19 sentiments concerning college online education", *Future Business Journal*, Vol. 9 No. 1, doi: [10.1186/s43093-023-00284-3](https://doi.org/10.1186/s43093-023-00284-3).
- Cardona, T., Cudney, E.A., Hoerl, R. and Snyder, J. (2020), "Data mining and machine learning retention models in higher education", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 25 No. 1, doi: [10.1177/1521025120964920](https://doi.org/10.1177/1521025120964920).

- Cardona, T., Cudney, E.A., Hoerl, R. and Snyder, J. (2023), "Data mining and machine learning retention models in higher education", *Journal of College Student Retention: Research, Theory and Practice*, Vol. 25 No. 1, pp. 51-75, doi: [10.1177/1521025120964920](https://doi.org/10.1177/1521025120964920).
- Carrera-Rivera, A., Larrinaga, F. and Lasar, G. (2022), "Context-awareness for the design of smart-product service systems: literature review", *In Computers in Industry*, Vol. 142, p. 103730, doi: [10.1016/j.compind.2022.103730](https://doi.org/10.1016/j.compind.2022.103730).
- Chango, W., Lara, J.A., Cerezo, R. and Romero, C. (2022), "A review on data fusion in multimodal learning analytics and educational data mining", *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, Vol. 12 No. 4, doi: [10.1002/widm.1458](https://doi.org/10.1002/widm.1458).
- Chaturvedi, M. (2017), "Data mining and its application in EDM domain", *Proceedings of the 2017 International Conference on Intelligent Computing and Control Systems, ICICCS 2017, 2018-January*, pp. 829-834, doi: [10.1109/ICCONS.2017.8250579](https://doi.org/10.1109/ICCONS.2017.8250579).
- Chen, L.-L. (2019), "Enhancing teaching with effective data mining protocols", *Journal of Educational Technology Systems*, Vol. 47 No. 4, pp. 500-512, doi: [10.1177/0047239519835846](https://doi.org/10.1177/0047239519835846).
- Chu, X., Cao, F., Jiao, L., Wang, J. and Jiao, Y. (2022), "Optimal allocation of higher education resources based on data mining and cloud computing", *Wireless Communications and Mobile Computing*, Vol. 2022, doi: [10.1155/2022/7067676](https://doi.org/10.1155/2022/7067676).
- Chytas, K., Tsolakidis, A., Triperina, E. and Skourlas, C. (2023), "Educational data mining in the academic setting: employing the data produced by blended learning to ameliorate the learning process", *Data Technologies and Applications*, Vol. 57 No. 3, pp. 366-384, doi: [10.1108/DTA-06-2022-0252](https://doi.org/10.1108/DTA-06-2022-0252).
- Contreras-Luján, E.E., García-Guerrero, E.E., López-Bonilla, O.R., Tlelo-Cuautle, E., López-Mancilla, D. and Inzunza-González, E. (2022), "Evaluation of machine learning algorithms for early diagnosis of deep venous thrombosis", *Mathematical and Computational Applications*, Vol. 27 No. 2, p. 24, doi: [10.3390/mca27020024](https://doi.org/10.3390/mca27020024).
- Dabhade, P., Agarwal, R., Alameen, K.P., Fathima, A.T., Sridharan, R. and Gopakumar, G. (2021), "Educational data mining for predicting students' academic performance using machine learning algorithms", *Materials Today: Proceedings*, Vol. 47, pp. 5260-5267, doi: [10.1016/j.matpr.2021.05.646](https://doi.org/10.1016/j.matpr.2021.05.646).
- Daghestani, L.F., Ibrahim, L.F., Al-Towirgi, R.S. and Salman, H.A. (2020), "Adapting gamified learning systems using educational data mining techniques", *Computer Applications in Engineering Education*, Vol. 28 No. 3, pp. 568-589, doi: [10.1002/cae.22227](https://doi.org/10.1002/cae.22227).
- Daoudi, I., Chebil, R., Tranvouez, E., Lejouad Chaari, W. and Espinasse, B. (2021), "Improving learners' assessment and evaluation in crisis management serious games: an emotion-based educational data mining approach", *Entertainment Computing*, Vol. 38, p. 100428, doi: [10.1016/j.entcom.2021.100428](https://doi.org/10.1016/j.entcom.2021.100428).
- Davies, R., Allen, G., Albrecht, C., Bakir, N. and Ball, N. (2021), "Using educational data mining to identify and analyze student learning strategies in an online flipped classroom", *Education Sciences*, Vol. 11 No. 11, doi: [10.3390/educsci11110668](https://doi.org/10.3390/educsci11110668).
- Dutt, A., Ismail, M.A. and Herawan, T. (2017), "A systematic review on educational data mining", *In IEEE Access*, Vol. 5, pp. 15991-16005, doi: [10.1109/ACCESS.2017.2654247](https://doi.org/10.1109/ACCESS.2017.2654247).
- Du, H., Xing, W. and Zhu, G. (2022), "Mining teacher informal online learning networks: insights from massive educational chat tweets", *Journal of Educational Computing Research*, Vol. 61 No. 1, doi: [10.1177/07356331221103764](https://doi.org/10.1177/07356331221103764).
- Du, H., Xing, W. and Zhu, G. (2023), "Mining teacher informal online learning networks: Insights from massive educational chat tweets", *Journal of Educational Computing Research*, Vol. 61 No. 1, pp. 127-150, doi: [10.1177/07356331221103764](https://doi.org/10.1177/07356331221103764).
- Du, X., Yang, J., Hung, J.L. and Shelton, B. (2020), "Educational data mining: a systematic review of research and emerging trends", *In Information Discovery and Delivery*, Vol. 48 No. 4, pp. 225-236, doi: [10.1108/IDD-09-2019-0070](https://doi.org/10.1108/IDD-09-2019-0070).

- Feng, G., Fan, M. and Chen, Y. (2022), "Analysis and prediction of students' academic performance based on educational data mining", *IEEE Access*, Vol. 10, pp. 19558-19571, doi: [10.1109/ACCESS.2022.3151652](https://doi.org/10.1109/ACCESS.2022.3151652).
- Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R. and Erven, G.V. (2019), "Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil", *Journal of Business Research*, Vol. 94, pp. 335-343, doi: [10.1016/j.jbusres.2018.02.012](https://doi.org/10.1016/j.jbusres.2018.02.012).
- Francis, B.K. and Babu, S.S. (2019), "Predicting academic performance of students using a hybrid data mining approach", *Journal of Medical Systems*, Vol. 43 No. 6, doi: [10.1007/s10916-019-1295-4](https://doi.org/10.1007/s10916-019-1295-4).
- Gera, J., Bhambri Marwaha, E., Thareja, R. and Jain, A. (2024), "Predicting and improving behavioural factors that boosts learning abilities in post-pandemic times using AI techniques", *In IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 14 No. 11, available at: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- Guleria, P. and Sood, M. (2023), "Explainable AI and machine learning: performance evaluation and explainability of classifiers on educational data mining inspired career counseling", *Education and Information Technologies*, Vol. 28 No. 1, pp. 1081-1116, doi: [10.1007/S10639-022-11221-2/TABLES/9](https://doi.org/10.1007/S10639-022-11221-2/TABLES/9).
- Haryani, C.A., Daicy, W., Widjaja, A.E., Aribowo, A. and Hery, K. (2022), "Educational data mining: the application in the university's feedback survey analysis using classification and clustering techniques", *2022 International Conference on Science and Technology, ICOSTECH 2022*, doi: [10.1109/ICOSTECH54296.2022.9829148](https://doi.org/10.1109/ICOSTECH54296.2022.9829148)
- Hasan, R., Palaniappan, S., Mahmood, S., Abbas, A., Sarker, K.U. and Sattar, M.U. (2020), "Predicting student performance in higher educational institutions using video learning analytics and data mining techniques", *Applied Sciences*, Vol. 10 No. 11, doi: [10.3390/app10113894](https://doi.org/10.3390/app10113894).
- Hassan, H., Ahmad, N.B. and Anuar, S. (2020), "Improved students' performance prediction for multi-class imbalanced problems using hybrid and ensemble approach in educational data mining", *Journal of Physics: Conference Series*, Vol. 1529 No. 5, doi: [10.1088/1742-6596/1529/5/052041](https://doi.org/10.1088/1742-6596/1529/5/052041).
- Hassan, K.M., Khafagy, M.H. and Thabet, F.M. (2024), "Mining educational data to analyze the student's performance in TOEFL iBT reading, listening and writing scores", *In IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 13 No. 7, available at: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- Hernández-Blanco, A., Herrera-Flores, B., Tomás, D. and Navarro-Colorado, B. (2019), "A systematic review of deep learning approaches to educational data mining", *In Complexity*, Vol. 2019, doi: [10.1155/2019/1306039](https://doi.org/10.1155/2019/1306039).
- Hu, D. and Guo, Z. (2018), "Exploring deep integration of information technology and china's higher education in the era of big data", *Proceedings – 6th International Conference of Educational Innovation Through Technology, EITT 2017*, 2018-March, pp. 262-267, doi: [10.1109/EITT.2017.71](https://doi.org/10.1109/EITT.2017.71)
- Hung, H.C., Liu, I.F., Liang, C.T. and Su, Y.S. (2020), "Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education", *Symmetry*, Vol. 12 No. 2, doi: [10.3390/sym12020213](https://doi.org/10.3390/sym12020213).
- Inusah, F., Missah, Y.M., Ussiph, N. and Twum, F. (2021), "Expert system in enhancing efficiency in basic educational management using data mining techniques", *In IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 12 No. 11, available at: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org).
- Inusah, F., Missah, Y.M., Najim, U. and Twum, F. (2022), "Data mining and visualisation of basic educational resources for quality education", *International Journal of Engineering Trends and Technology*, Vol. 70 No. 12, pp. 296-307, doi: [10.14445/22315381/IJETT-V70I12P228](https://doi.org/10.14445/22315381/IJETT-V70I12P228).
- Inusah, F., Missah, Y.M., Najim, U. and Twum, F. (2023a), "Agile neural expert system for managing basic education", *Intelligent Systems with Applications*, Vol. 17, doi: [10.1016/j.iswa.2023.200178](https://doi.org/10.1016/j.iswa.2023.200178).

- Inusah, F., Missah, Y.M., Najim, U. and Twum, F. (2023b), "Integrating expert system in managing basic education: a survey in Ghana", *International Journal of Information Management Data Insights*, Vol. 3 No. 1, p. 100166, doi: [10.1016/j.jjime.2023.100166](https://doi.org/10.1016/j.jjime.2023.100166).
- Issah, I., Appiah, O., Appiahene, P. and Inusah, F. (2023), "A systematic review of the literature on machine learning application of determining the attributes influencing academic performance", *Decision Analytics Journal*, Vol. 7, p. 100204, doi: [10.1016/j.dajour.2023.100204](https://doi.org/10.1016/j.dajour.2023.100204).
- Jha, R. (2023), "A systematic study on student performance prediction from the perspective of machine learning and data mining approaches", *Proceedings of the 8th International Conference on Communication and Electronics Systems, ICCES 2023*, pp. 1336-1342, doi: [10.1109/ICCES57224.2023.10192866](https://doi.org/10.1109/ICCES57224.2023.10192866)
- Jiang, L. (2022), "Development and implementation path of kindergarten stem educational activities based on data mining", *Computational Intelligence and Neuroscience*, Vol. 2022, doi: [10.1155/2022/2700674](https://doi.org/10.1155/2022/2700674).
- Jiang, J. and Zeng, L. (2019), "Research on individualized teaching based on big data mining", 14th International Conference on Computer Science and Education, ICCSE 2019, pp. 56-59, doi: [10.1109/ICCSE.2019.8845054](https://doi.org/10.1109/ICCSE.2019.8845054)
- Joshya, K.X., Thakurta, R. and Sekh, A.A. (2023), "Future educational environment – identification of smart campus topic trends using text mining", In *International Journal of Educational Management*, Vol. 37 No. 4, pp. 884-906, doi: [10.1108/IJEM-01-2023-0005](https://doi.org/10.1108/IJEM-01-2023-0005).
- Karthikeyan, V.G., Thangaraj, P. and Karthik, S. (2020), "Towards developing hybrid educational data mining model (HEDM) for efficient and accurate student performance evaluation", *Soft Computing*, Vol. 24 No. 24, pp. 18477-18487, doi: [10.1007/s00500-020-05075-4](https://doi.org/10.1007/s00500-020-05075-4).
- Khan, A. and Ghosh, S.K. (2021), "Student performance analysis and prediction in classroom learning: a review of educational data mining studies", *Education and Information Technologies*, Vol. 26 No. 1, pp. 205-240, doi: [10.1007/s10639-020-10230-3](https://doi.org/10.1007/s10639-020-10230-3).
- Kurdi, M.M., Al-Khafagi, H. and Elzein, I. (2018), "Mining educational data to analyze students' behavior and performance", Proceedings of 2018 JCCO Joint International Conference on ICT in Education and Training, International Conference on Computing in Arabic, and International Conference on Geocomputing, *JCCO: TICET-ICCA-GECO 2018*, pp. 171-175, doi: [10.1109/ICCA-TICET.2018.8726203](https://doi.org/10.1109/ICCA-TICET.2018.8726203)
- Leitner, P., Ebner, M. and Ebner, M. (2019), "Learning analytics challenges to overcome in higher education institutions", *Utilizing Learning Analytics to Support Study Success*, pp. 91-104, doi: [10.1007/978-3-319-64792-0\\_6](https://doi.org/10.1007/978-3-319-64792-0_6).
- Lemay, D.J., Baek, C. and Doleck, T. (2021), "Comparison of learning analytics and educational data mining: a topic modeling approach", *Computers and Education: Artificial Intelligence*, Vol. 2, p. 100016, doi: [10.1016/j.CAEAI.2021.100016](https://doi.org/10.1016/j.CAEAI.2021.100016).
- Li, K., Huang, J. and Lin, J. (2022), "The architecture of college psychological teaching management system based on data mining technology", *Security and Communication Networks*, Vol. 2022, pp. 1-12, doi: [10.1155/2022/7968756](https://doi.org/10.1155/2022/7968756).
- Lin, X., Guan, W. and Zhang, Y. (2023), "Application of data mining technology with improved clustering algorithm in library personalized book recommendation system", in *International Journal of Advanced Computer Science and Applications (IJACSA)*, Vol. 14 No. 11, available at: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- Livieris, I.E., Drakopoulou, K., Tampakas, V.T., Mikropoulos, T.A. and Pintelas, P. (2018), "Predicting secondary school students' performance utilizing a semi-supervised learning approach", *Journal of Educational Computing Research*, Vol. 57 No. 2, pp. 448-470, doi: [10.1177/0735633117752614](https://doi.org/10.1177/0735633117752614).
- Lv, X. (2021), "A quality evaluation scheme for curriculum in ideological and political education based on data mining", *Proceedings – 2021 13th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2021*, pp. 649-652, doi: [10.1109/ICMTMA52658.2021.00149](https://doi.org/10.1109/ICMTMA52658.2021.00149).

- 
- Marfo Missah, Y., Inusah, F., Najim, U. and Twum, F. (2023), "Evaluating agile neural educational system for effective resource management", *SAGE Open*, Vol. 13 No. 4, doi: [10.1177/21582440231214843](https://doi.org/10.1177/21582440231214843).
- Martínez-Abad, F., Gamazo, A. and Rodríguez-Conde, M.J. (2020), "Educational data mining: Identification of factors associated with school effectiveness in PISA assessment", *Studies in Educational Evaluation*, Vol. 66, p. 100875, doi: [10.1016/J.STUEDUC.2020.100875](https://doi.org/10.1016/J.STUEDUC.2020.100875).
- Mazidi, A. and Abusham, E. (2018), "Study of general education diploma students' performance and prediction in sultanate of Oman, based on data mining approaches", *International Journal of Engineering Business Management*, Vol. 10, doi: [10.1177/1847979018807020/ASSET/IMAGES/LARGE/10.1177\\_1847979018807020-FIG1.JPEG](https://doi.org/10.1177/1847979018807020/ASSET/IMAGES/LARGE/10.1177_1847979018807020-FIG1.JPEG).
- Mengash, H.A. (2020), "Using data mining techniques to predict student performance to support decision making in university admission systems", *IEEE Access*, Vol. 8, pp. 55462-55470, doi: [10.1109/ACCESS.2020.2981905](https://doi.org/10.1109/ACCESS.2020.2981905).
- Mohamad, S.K. and Tasir, Z. (2013), "Educational data mining: a review", *Procedia – Social and Behavioral Sciences*, Vol. 97, pp. 320-324, doi: [10.1016/j.sbspro.2013.10.240](https://doi.org/10.1016/j.sbspro.2013.10.240).
- Mohd Talib, N.I., Abd Majid, N.A. and Sahran, S. (2023), "Identification of student behavioral patterns in higher education using K-means clustering and support vector machine", *Applied Sciences*, Vol. 13 No. 5, doi: [10.3390/app13053267](https://doi.org/10.3390/app13053267).
- Mushtaq, H., Siddique, I., Malik, B.H., Ahmed, M., Butt, U.M., Tahir Ghafoor, R.M., Zubair, H. and Farooq, U. (2019), "Educational data classification framework for community pedagogical content management using data mining", *International Journal of Advanced Computer Science and Applications*, Vol. 10 No. 1, pp. 329-338, doi: [10.14569/IJACSA.2019.0100144](https://doi.org/10.14569/IJACSA.2019.0100144).
- Nascimento, R.L.S., das Neves Junior, R.B., de Almeida Neto, M.A. and de Araújo Fagundes, R.A. (2018), "Educational data mining: an application of regressors in predicting school dropout", *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 10935 LNAI, pp. 246-257, doi: [10.1007/978-3-319-96133-0\\_19](https://doi.org/10.1007/978-3-319-96133-0_19).
- Okewu, E., Adewole, P., Misra, S., Maskeliunas, R. and Damasevicius, R. (2021), "Artificial neural networks for educational data mining in higher education: a systematic literature review", *Applied Artificial Intelligence*, Vol. 35 No. 13, pp. 983-1021, doi: [10.1080/08839514.2021.1922847](https://doi.org/10.1080/08839514.2021.1922847).
- Ordoñez-Avila, R., Salgado Reyes, N., Meza, J. and Ventura, S. (2023), "Data mining techniques for predicting teacher evaluation in higher education: a systematic literature review", *In Heliyon*, Vol. 9 No. 3, doi: [10.1016/j.heliyon.2023.e13939](https://doi.org/10.1016/j.heliyon.2023.e13939).
- Paredes, Y.V., Siegle, R.F., Hsiao, I.-H. and Craig, S.D. (2021), "Educational data mining and learning analytics for improving online learning environments", *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 64 No. 1, pp. 500-504, doi: [10.1177/1071181320641113](https://doi.org/10.1177/1071181320641113).
- Peng, W. and Tang, Z. (2022), "Construction and application of farmers' practical teaching system in vocational education based on big data mining technology", *Computational Intelligence and Neuroscience*, Vol. 2022, pp. 1-9, doi: [10.1155/2022/6075719](https://doi.org/10.1155/2022/6075719).
- Prada, M.A., Dominguez, M., Vicario, J.L., Alves, P.A.V., Barbu, M., Podpora, M., Spagnolini, U., Pereira, M.J.V. and Vilanova, R. (2020), "Educational data mining for tutoring support in higher education: a web-based tool case study in engineering degrees", *IEEE Access*, Vol. 8, pp. 212818-212836, doi: [10.1109/ACCESS.2020.3040858](https://doi.org/10.1109/ACCESS.2020.3040858).
- Rafique, A., Ameen, K. and Arshad, A. (2023), "E-book data mining: real information behavior of university academic community", *Library Hi Tech*, Vol. 41 No. 2, pp. 413-431, doi: [10.1108/LHT-07-2020-0176](https://doi.org/10.1108/LHT-07-2020-0176).
- Rahman, M., Ming, T.H., Baigh, T.A. and Sarker, M. (2022a), "Adoption of artificial intelligence in banking services: an empirical analysis", *International Journal of Emerging Markets*, doi: [10.1108/IJOEM-06-2020-0724](https://doi.org/10.1108/IJOEM-06-2020-0724).



- Rahman, M.M., Watanobe, Y., Matsumoto, T., Kiran, R.U. and Nakamura, K. (2022b), "Educational data mining to support programming learning using problem-solving data", *IEEE Access*, Vol. 10, pp. 26186-26202, doi: [10.1109/ACCESS.2022.3157288](https://doi.org/10.1109/ACCESS.2022.3157288).
- Rebai, S., Ben Yahia, F. and Essid, H. (2020), "A graphically based machine learning approach to predict secondary schools performance in Tunisia", *Socio-Economic Planning Sciences*, Vol. 70, doi: [10.1016/j.seps.2019.06.009](https://doi.org/10.1016/j.seps.2019.06.009).
- Rodrigues, M.W., Isotani, S. and Zárate, L.E. (2018), "Educational data mining: a review of evaluation process in the e-learning", *Telematics and Informatics*, Vol. 35 No. 6, pp. 1701-1717, doi: [10.1016/j.TELE.2018.04.015](https://doi.org/10.1016/j.TELE.2018.04.015).
- Roslan, M.H., Bin, C. and Chen, C.J. (2023), "Predicting students' performance in English and mathematics using data mining techniques", *Education and Information Technologies*, Vol. 28 No. 2, pp. 1427-1453, doi: [10.1007/s10639-022-11259-2](https://doi.org/10.1007/s10639-022-11259-2).
- Sánchez, A., Vidal-Silva, C., Mancilla, G., Tupac-Yupanqui, M. and Rubio, J.M. (2023), "Sustainable e-learning by data mining—successful results in a Chilean university", *Sustainability*, Vol. 15 No. 2, doi: [10.3390/su15020895](https://doi.org/10.3390/su15020895).
- Sghir, N., Adadi, A. and Lahmer, M. (2023), "Recent advances in predictive learning analytics: a decade systematic review (2012–2022)", *Education and Information Technologies*, Vol. 28 No. 7, pp. 8299-8333, doi: [10.1007/s10639-022-11536-0](https://doi.org/10.1007/s10639-022-11536-0).
- Shafiq, D.A., Marjani, M., Habeeb, R.A.A. and Asirvatham, D. (2022), "Student retention using educational data mining and predictive analytics: a systematic literature review", *In IEEE Access*, Vol. 10, pp. 72480-72503, doi: [10.1109/ACCESS.2022.3188767](https://doi.org/10.1109/ACCESS.2022.3188767).
- Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y. and Galligan, L. (2023), "Sentiment analysis and opinion mining on educational data: a survey", *Natural Language Processing Journal*, Vol. 2, p. 100003, doi: [10.1016/j.nlp.2022.100003](https://doi.org/10.1016/j.nlp.2022.100003).
- Shao, Z., Sun, H., Wang, X. and Sun, Z. (2020), "An optimized mining algorithm for analyzing students' learning degree based on dynamic data", *IEEE Access*, Vol. 8, pp. 113543-113556, doi: [10.1109/ACCESS.2020.3001749](https://doi.org/10.1109/ACCESS.2020.3001749).
- Shayakhmetova, B., Ayazhanov, S.S., Omarova, S., Ten, T., Drozd, V. and Medeubaeva, A. (2021), "Descriptive big data analytics in the field of education", *SIST 2021 – 2021 IEEE International Conference on Smart Information Systems and Technologies*, doi: [10.1109/SIST50301.2021.9465958](https://doi.org/10.1109/SIST50301.2021.9465958).
- Sokkhey, P., Penh, P. and Okazaki, C.T. (2020), "Developing web-based support systems for predicting poor-performing students using educational data mining techniques", *In IJACSA International Journal of Advanced Computer Science and Applications*, Vol. 11 No. 7, available at: [www.ijacsa.thesai.org](http://www.ijacsa.thesai.org)
- Sultana, S., Khan, S. and Abbas, M.A. (2017), "Predicting performance of electrical engineering students using cognitive and non-cognitive features for identification of potential dropouts", *International Journal of Electrical Engineering Education*, Vol. 54 No. 2, pp. 105-118, doi: [10.1177/0020720916688484](https://doi.org/10.1177/0020720916688484).
- Sun, X. (2023), "Evaluation on the application of data mining in the evaluation system of teaching ability of double teacher teachers", *International Conference on Network, Multimedia and Information Technology, NMITCON 2023*, doi: [10.1109/NMITCON58196.2023.10276049](https://doi.org/10.1109/NMITCON58196.2023.10276049).
- Tan, B. (2023), "Research on the prediction of the inauguration development direction of college students' entrepreneurship education based on educational data mining", *International Journal of Computational Intelligence Systems*, Vol. 16 No. 1, doi: [10.1007/s44196-023-00316-4](https://doi.org/10.1007/s44196-023-00316-4).
- Tang, H., Xing, W. and Pei, B. (2018), "Time really matters: understanding the temporal dimension of online learning using educational data mining", *Journal of Educational Computing Research*, Vol. 57 No. 5, pp. 1326-1347, doi: [10.1177/0735633118784705](https://doi.org/10.1177/0735633118784705).
- Tasnim, N., Paul, M.K. and Sattar, A.H.M.S. (2019), "Identification of drop out students using educational data mining", *2nd International Conference on Electrical, Computer and Communication Engineering, ECCE 2019, 1 April*, doi: [10.1109/ECACE.2019.8679385](https://doi.org/10.1109/ECACE.2019.8679385).

- Torcate, A.S. and De Oliveira Rodrigues, C.M. (2021), "Educational data mining with learning analytics and unsupervised algorithms: analysis and diagnosis in basic education", *Proceedings – 2021 16th Latin American Conference on Learning Technologies, LACLO 2021*, pp. 67-74, doi: [10.1109/LACLO54177.2021.00014](https://doi.org/10.1109/LACLO54177.2021.00014).
- Trakunphutthirak, R. and Lee, V.C.S. (2021), "Application of educational data mining approach for student academic performance prediction using progressive temporal data", *Journal of Educational Computing Research*, Vol. 60 No. 3, pp. 742-776, doi: [10.1177/07356331211048777](https://doi.org/10.1177/07356331211048777).
- Tsiakmaki, M., Kostopoulos, G., Kotsiantis, S. and Ragos, O. (2020), "Implementing autoML in educational data mining for prediction tasks", *Applied Sciences*, Vol. 10 No. 1, doi: [10.3390/app10010090](https://doi.org/10.3390/app10010090).
- Wang, G.X. (2018), "Mining analysis for college students' physical performance and fitness tests", *Proceedings – 2018 International Conference on Virtual Reality and Intelligent Systems, ICVRIS 2018*, pp. 191-195, doi: [10.1109/ICVRIS.2018.00054](https://doi.org/10.1109/ICVRIS.2018.00054)
- Wang, H. (2023), "Teaching quality monitoring and evaluation using 6G internet of things communication and data mining", *International Journal of System Assurance Engineering and Management*, Vol. 14 No. 1, pp. 120-127, doi: [10.1007/s13198-021-01206-8](https://doi.org/10.1007/s13198-021-01206-8).
- Wang, J., Xie, Z., Li, Y., Song, Y., Yan, J., Bai, W., Zhou, T. and Qin, J. (2020), "Relationship between health status and physical fitness of college students from South China: an empirical study by data mining approach", *IEEE Access*, Vol. 8, pp. 67466-67473, doi: [10.1109/ACCESS.2020.2986039](https://doi.org/10.1109/ACCESS.2020.2986039).
- Widyaningsih, Y., Fitriani, N. and Sarwinda, D. (2019), "A semi-supervised learning approach for predicting student' s performance: First-Year", *2019 12th International Conference on Information and Communication Technology and System (ICTS)*, pp. 291-295.
- Wongvorachan, T., He, S. and Bulut, O. (2023), "A comparison of undersampling, oversampling, and SMOTE methods for dealing with imbalanced classification in educational data mining", *Information*, Vol. 14 No. 1, doi: [10.3390/info14010054](https://doi.org/10.3390/info14010054).
- Wu, F. (2022a), "Intelligent application of data mining model in Chinese international education", *Mathematical Problems in Engineering*, Vol. 2022, pp. 1-7, doi: [10.1155/2022/9171551](https://doi.org/10.1155/2022/9171551).
- Wu, J. (2022b), "Intelligent classroom learning model of college English based on data mining technology in mobile edge computing environment", *Wireless Communications and Mobile Computing*, Vol. 2022, doi: [10.1155/2022/4777195](https://doi.org/10.1155/2022/4777195).
- Wu, X. (2022c), "The collection and utilization of web resources for teaching world history based on data mining technology", *Journal of Environmental and Public Health*, Vol. 2022, doi: [10.1155/2022/9124952](https://doi.org/10.1155/2022/9124952).
- Xilin, Z. and Honglian, Y. (2021), "Design of data mining and evaluation system for college students' mental health", *Proceedings – 2021 13th International Conference on Measuring Technology and Mechatronics Automation, ICMTMA 2021*, pp. 575-578, doi: [10.1109/ICMTMA52658.2021.00132](https://doi.org/10.1109/ICMTMA52658.2021.00132)
- Xing, W. and Du, D. (2019), "Dropout prediction in MOOCs: using deep learning for personalized intervention", *Journal of Educational Computing Research*, Vol. 57 No. 3, pp. 547-570, doi: [10.1177/0735633118757015](https://doi.org/10.1177/0735633118757015).
- Yağcı, M. (2022), "Educational data mining: prediction of students' academic performance using machine learning algorithms", *Smart Learning Environments*, Vol. 9 No. 1, doi: [10.1186/s40561-022-00192-z](https://doi.org/10.1186/s40561-022-00192-z).
- Yakubu, M.N. and Abubakar, A.M. (2021), "Applying machine learning approach to predict students' performance in higher educational institutions", *Kybernetes*, doi: [10.1108/K-12-2020-0865](https://doi.org/10.1108/K-12-2020-0865).
- Yang, L. (2021), "Data mining based college English teaching assistant expert system", *Proceedings – 2021 6th International Conference on Smart Grid and Electrical Automation, ICSGEA 2021*, pp. 280-283, doi: [10.1109/ICSGEA53208.2021.00068](https://doi.org/10.1109/ICSGEA53208.2021.00068).
- Yang, T., Chen, B., Wang, W. and Li, S. (2022), "Research on feedback service for teaching based on educational data mining", *Proceedings – 2022 International Conference on Machine Learning and Knowledge Engineering, MLKE 2022*, pp. 306-309, doi: [10.1109/MLKE55170.2022.00065](https://doi.org/10.1109/MLKE55170.2022.00065).

- 
- Yürüm, O.R., Taşkaya-Temizel, T. and Yıldırım, S. (2023), “The use of video clickstream data to predict university students’ test performance: a comprehensive educational data mining approach”, *Education and Information Technologies*, Vol. 28 No. 5, pp. 5209-5240, doi: [10.1007/s10639-022-11403-y](https://doi.org/10.1007/s10639-022-11403-y).
- Zárate, L., Rodrigues, M.W., Dias, S.M., Nobre, C. and Song, M. (2023), “SciBR-M: a method to map the evolution of scientific interest – a case study in educational data mining”, *Library Hi Tech*, Vol. 41 No. 6, pp. 1694-1724, doi: [10.1108/LHT-04-2022-0222](https://doi.org/10.1108/LHT-04-2022-0222).
- Zhang, W. and Qin, S. (2018), “A brief analysis of the key technologies and applications of educational data mining on online learning platform”, *IEEE 3rd International Conference on Big Data Analysis, ICBDA 2018*, 83-86, doi: [10.1109/ICBDA.2018.8367655](https://doi.org/10.1109/ICBDA.2018.8367655)
- Zhao, X. (2022), “Leveraging data mining technique to enhancing online education and its efficiency study”, *Mathematical Problems in Engineering*, Vol. 2022, doi: [10.1155/2022/4946619](https://doi.org/10.1155/2022/4946619).
- Zhou, X. and Wu, X. (2022), “Teaching mode based on educational big data mining and digital twins”, *Computational Intelligence and Neuroscience*, Vol. 2022, doi: [10.1155/2022/9071944](https://doi.org/10.1155/2022/9071944).

### Further reading

- Hussain, S. and Khan, M.Q. (2021), “Student-performulator: predicting students’ academic performance at secondary and intermediate level using machine learning”, *Annals of Data Science, MI*, doi: [10.1007/s40745-021-00341-0](https://doi.org/10.1007/s40745-021-00341-0).

### Corresponding author

Inusah Fuseini can be contacted at: [finusah@uds.edu.gh](mailto:finusah@uds.edu.gh)