Educational Data Mining in European Union – Achievements and Challenges: A Systematic Literature Review

Corina Simionescu, Mirela Danubianu, Bogdănel Constantin Grădinaru, Marius Silviu Măciucă
Faculty of Electrical Engineering and Computer Science, Stefan cel Mare University of Suceava, Romania

***Abstract*—The quality of education is one of the pillars of sustainable development, as set out in "The 2030 Agenda for Sustainable Development", adopted by all United Nations Member States in 2015. Recent social and technological developments, as well as events such as the COVID-19 pandemic or conflicts in many parts of the world, have led to essential changes in the way education processes are carried out. In addition, they have made it possible to generate, collect and store large amounts of data related to these processes, data that can hide useful information for decisions that, in the medium or long term, can lead to a significant increase in the quality of education. Uncovering this information is the subject of Educational Data Mining. To understand the state-of-the-art reflected by recent developments, trends, theories, methodologies, and applications in this field, in the European Union, we considered it appropriate to conduct a systematic and critical literature review. Our paper aims to identify, analyze, and synthesize relevant information from these articles, both to build a foundation for further studies and to identify gaps or unexplored issues that can be addressed in future research. The analysis is based on research identified in three international databases recognized for content quality: Scopus, Science direct, and IEEEXplore.**

***Keywords*—Educational data mining; systematic literature review; European Union; Kitchenham methodology; data mining techniques**

I. Introduction

Nowadays, educational institutions generate, collect and store huge volumes of data from a variety of sources and processes. The use of computers, of internet or learning management systems (LMS) has triggered an exponential growth in the amount of data. Much of this is generated through online technologies, such as e-learning platforms, search engines, social networks, electronic communication, or through watching videos, etc., but also through direct collection from assessment processes or from monitoring students' behavior. Different types of data about users' online interactions, such as clicks, browsing preferences and behaviors, or about learning outcomes and student preferences are collected and stored. This avalanche of educational data provides opportunities for deeper understanding of learning processes, increasing the quality of education and improving teaching strategies.

To harness this wealth of data, researchers have used data mining techniques to extract "unexpected and valuable" patterns and knowledge [1]. Thus emerged the interdisciplinary field of Educational Data Mining (EDM), which focuses on obtaining hidden but potentially valuable information in the context of education and learning through specific data mining methods applied to data collected from these processes. It establishes a connection between two distinct fields: education, on the one hand, and computer science, specifically data mining, on the other [2] combining elements from artificial intelligence, machine learning, statistics, expert systems, databases, and visualization to investigate and optimize educational processes.

The International Educational Data Mining Society, which hosts and publishes the Journal of Educational Data Mining, offers the following definition of EDM: "Educational Data Mining is an emerging discipline concerned with developing methods for exploring the unique types of data that come from educational environments and using these methods to better understand students and the environments in which they learn" [3].

To identify the newest trends in the field, we conducted a Systematic Literature Review (SLR) in the EDM field. To reflect a state as close as possible to the present moment, we have chosen as our target publications from 2013 to 2023. It is well known that EU member countries have educational systems with specific local characteristics and there is no uniform standard. However, the application of EDM techniques on the data collected from these systems can lead to valuable information for decision making, and why not, to their optimization at EU level. This is why we have turned our attention to those papers that have authors affiliated to educational institutions in EU member countries or that use datasets collected from these institutions.

This analysis allows us to identify the challenges and opportunities associated with the use of data mining techniques in education, leading to the discovery of information that supports decisions that target more effective, personalized, and results-oriented learning. The research aims to provide a documented response regarding the data mining methods used in EDM, the level to which they have been used, their benefits and shortcomings.

In a preliminary survey of SLRs on EDM to which we had access, we found that although there are still such papers in the field, they address issues other than those proposed by us and do not reflect the current state of research. For example, the first such study was conducted by Romero and Ventura [4] and was updated in 2010 [5] and 2013 [6]. In these, 11 categories of tasks in EDM were identified including: data analysis and visualization, feedback to support instructors, recommendations for students, predicting student performance, modeling student behavior, grouping students, social network analysis, concept map development, and curriculum construction. Specific methods and techniques were presented for each of these.

In [7], a study that identified a set of educational functionalities, an approach to EDM, and two patterns describing EDM based on descriptive and predictive models was proposed. The computational techniques and not the applications were mainly analyzed. Other research is focused on the current state of application of different approaches in EDM. In [8] a systematic review of the literature addressing the use of clustering in EDM is given, and in [9] performance prediction based on machine learning techniques is addressed.

As a result, the novelty of our work lies in the fact that it provides an up-to-date overview of research and trends in EDM use across the European Union assesses the level of interest in the field and uncovers those aspects that may constitute new directions for research.

To carry out this analysis we considered the Kitchenham methodology [10]. This has become a well-known and respected approach in the academic community and has been applied in various research fields, from software engineering to public health.

Further, our paper is structured as follows: Section II introduces the concept of EDM, Section III discusses the methodology with all three stages of the Kitchenham methodology for conducting an SLR, Section IV provides a discussion on the reported results, Section V addresses some limitation and Section VI draws conclusions of the research.

II. Educational Data Mining

EDM process "converts raw data from educational systems into useful information with a potential positive impact on educational research and practice" [5].

It uses some of the core technologies in data mining to improve the quality of learning by modeling and discovering the correlation between learner's academic performance and learning behavior, teaching purpose and teaching strategy [8]. To achieve this goal, the following are mainly used: classification, association rules, clustering, regression, text mining, and web mining.

III. Working Methodology

Systematic literature review is an approach that, to be effective, must provide meaningful information as a foundation for decision-making. In this paper we have used the Kitchenham method [10] which states that the purpose of conducting a SLR is a broad review of the studies included in a particular field to recognize gaps in existing research for further investigation and to provide a thorough understanding of the field.

In accordance with the outlined guidelines, a systematic literature review consists in three distinct phases necessary for a formal research process, each of which containing specific steps and activities [10], as is presented in Fig. 1.

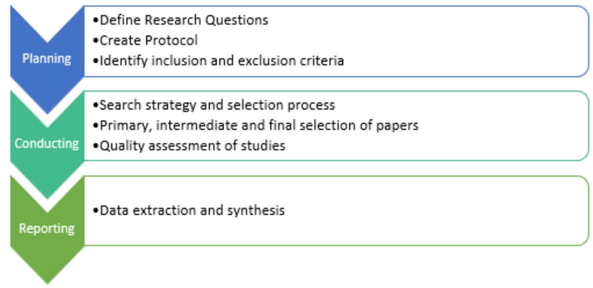

Fig. 1. Kitchenham methodology.

A. Planning the Review

We start by identifying the research questions followed by a description of the predefined used protocol. This includes information on the steps to be carried out in the review processes, such as:

1) Search strategy;
2) the paper selection strategy;
3) quality assessment;
4) data extraction and synthesis [11]
5) Additionally, a predefined review protocol reduces researchers bias [11]. In order to not duplicate the information, the protocol elements are described during the steps in which they are applied.

B. Purpose and Objectives of the Analysis, and Identification of Research Questions

The aim of our review is to assess the current state of research on EDM methods and techniques and their extent of use in the European Union.

The research has been focused on published works in the field of EDM, in the years 2013-2023, targeting countries belonging to the European Union. By identifying, then analyzing and synthesizing existing research we aim to gain an understanding of the trends of evolution, the methods and techniques used, and the results obtained from their application, thus facilitating an understanding of the current progress, opportunities and challenges in the field.

The objectives we propose are:

O1: To identify methods, techniques and algorithms used in EDM practice;

O2: To build a picture of EDM by analyzing and synthesizing published research from 2013-2023 on education systems in EU countries;

O3: To gain insight into the benefits of using data mining methods and techniques in education;

O4: To identify challenges, possible gaps and future research directions in EDM.

In line with the proposed goal, we formulated the following research questions:

RQ1: What is the extent to which EDM is implemented at the different levels of education systems in the EU?

RQ2: What is the trend of evolution of EDM research at EU level?

RQ3: To what extent are data mining techniques used in education?

All these lead to a documented answer to the following summary question: What is the current state of EDM research for education systems in EU countries?

C. Inclusion and Exclusion Criteria

To ensure that the literature under review fits the research aim, objectives and questions, we have established a set of rules in the form of inclusion and exclusion criteria. These allowed us to select relevant, quality and suitable papers to be considered for literature review in educational data mining.

1) Inclusion criteria

I1: Studies based on authors/data sets from the education system of European Union countries;

I2: Studies describing the use of data mining methods and techniques in the educational field of European Union countries;

I3: Papers that consistently address topics related to the purpose, objectives of the analysis and research questions.

2) Exclusion criteria

E1: Books and book chapters, book reviews, tutorials, errata, encyclopedias, editorials;

E2: Studies whose content could not be accessed (not Open access);

E3: Studies presenting a literature review covering aspects of Educational Data Mining.

E4: Papers written in other languages than English.

D. Metadata Used in the Analysis Process

An important aspect in achieving the proposed objectives is related to the design of the dataset to be collected. Beyond reading the articles, their analysis also requires the extraction of associated metadata values, which can be subject to different processing. We considered the following: authors, title of the paper, year of publication, abstract, paper length (in pages), DOI, document type, and keywords.

E. Conducting the Review

1) Search strategy and selection process: The first step in this process was to choose the appropriate international databases from which to select papers. For this purpose, we took into account three factors: the quality and international recognition of the database, the relevance for EDM, and the existence of advanced search tools. Accordingly, we chose three international databases: Scopus, Science direct, and IEEEXplore.

Our choice is motivated by the following considerations:

- Quality and international recognition: Scopus, ScienceDirect, and IEEEXplore are the benchmarks for the quality and international recognition of their content. They host peer-reviewed papers published by authors with expertise in the field, which makes the available information credible and reliable.
- Relevance for EDM: These databases host a significant number of scholarly articles and research papers in computer science, technology and education that are essential to EDM. This makes it possible to carry out a broad analysis and to identify trends in the field [12].
- Advanced search tools: all three platforms offer filters and advanced search tools, making it easy to identify and select relevant articles for deeper analysis [13].

Therefore, the choice of Scopus [14], ScienceDirect and IEEEXplore databases is justified to perform a comprehensive literature review, ensuring the selection of reliable and relevant sources. Access to the databases in this paper was provided through an institutional account provided by “Stefan cel Mare” University of Suceava, Romania.

2) The study selection process goes through three stages:

- Initial identification of published papers that could plausibly satisfy the search queries;
- Selection of candidate papers;
- Selection of final studies to analyze;
- Identification of potentially useful studies;
- This first step involved going through a systematic process of identifying and selecting papers that meet the purpose of our research. To find the best results for the research questions we used and tested various search strings.

In the beginning we used the string educational AND data AND mining as search expression in all metadata. As a result of the search, we obtained a very large number of articles in all three databases. For Scopus we obtained 99.731papers, for Science Direct, Elsevier 24.544 papers and for IEEExplore 8.937 works. We have found that many articles that comply with the search string do not actually relate to educational systems. As a result, we have restricted the search space to those papers with this string in the title. We have refined the search by adding a filter that only searches for papers published in the target interval, i.e., 2013-2023. Finally, we also considered the term school as a generic address targeting both university and pre-university environments. Table I shows the search strings in the four databases, constructed according to their specific rules.

TABLE I. SEARCH STRINGS TO IDENTIFY POTENTIALLY USEFUL STUDIES

Digital Library	Search string
Scopus	(ALL (educational AND data AND mining) AND TITLE (educational AND data AND mining) AND ALL (school)) AND PUBYEAR > 2012 AND PUBYEAR < 2024
Science direct, Elsevier	Educational AND data AND mining AND school Year:2013-2023 Title:educational AND data AND mining
IEEEXplore	("All Metadata":educational AND "All Metadata":data AND"All Metadata":mining) AND ("Document Title":educational AND ("Document Title":data AND "Document Title" :mining) AND ("All Metadata":school) Filters Applied: 2013 - 2023

A summary of the search results illustrating the evolution of the number of articles found initially and after the three refinements is presented in Table II.

TABLE II. POTENTIALLY USEFUL STUDIES

	Scopus	Science Direct	IEEE
Articles containing in all metadata the string ***educational AND data AND mining***	99.731	24.544	8.937
Articles published between 2013-2023	91.013	15.356	4.276
Articles having in title the string ***educational AND data AND mining***	636	28	183
Articles with the word ***school*** in all metadata	326	16	52

To ensure that the selected papers were useful for our research we scanned the titles, removed duplicates, considered the size of the paper in number of pages as a measure of consistency, reviewed papers under five pages and concluded that a paper of less than four pages did not provide sufficient information, and researched the affiliations of the authors and the language in which the paper was written. We applied additional filters in the search strings, where it was possible.

The structure of Scopus database allowed us to apply (extra) filters, unlike the other databases we searched. As a result, we easily applied filters that helped to reduce the number of items matching our criteria, such as:

- Language selection: we removed Portuguese, Chinese, Turkish, and we obtained 241 results.
- Country selection (we selected EU member countries).

(ALL (educational AND data AND mining) AND TITLE (educational AND data AND mining) AND ALL (school)) AND PUBYEAR > 2012 AND PUBYEAR < 2024 AND (LIMIT-TO (DOCTYPE , "ar") OR LIMIT-TO (DOCTYPE , "cp")) AND (LIMIT-TO (SRCTYPE , "j") OR LIMIT-TO (SRCTYPE , "p")) AND (LIMIT-TO (LANGUAGE , "English")) AND (LIMIT-TO (AFFILCOUNTRY , "Bulgaria") OR LIMIT-TO (AFFILCOUNTRY , "Cyprus") OR LIMIT-TO (AFFILCOUNTRY , "France") OR LIMIT-TO (AFFILCOUNTRY , "Finland") OR LIMIT-TO (AFFILCOUNTRY , "Germany") OR LIMIT-TO (AFFILCOUNTRY , "Greece") OR LIMIT-TO (AFFILCOUNTRY , "Ireland") OR LIMIT-TO (AFFILCOUNTRY , "Italy") OR LIMIT-TO (AFFILCOUNTRY , "Latvia") OR LIMIT-TO (AFFILCOUNTRY , "Lithuania") OR LIMIT-TO (AFFILCOUNTRY , "Netherlands") OR LIMIT-TO (AFFILCOUNTRY , "Poland") OR LIMIT-TO (AFFILCOUNTRY , "Romania") OR LIMIT-TO (AFFILCOUNTRY , "Slovakia") OR LIMIT-TO (AFFILCOUNTRY , "Spain") OR LIMIT-TO (AFFILCOUNTRY , "Sweden") OR LIMIT-TO (AFFILCOUNTRY , "Czech Republic") OR LIMIT-TO (AFFILCOUNTRY , "Croatia") OR LIMIT-TO (AFFILCOUNTRY , "Portugal"))

We obtained 43 results, three of which refer to the description of conference volumes. Consequently 40 papers remained for further analysis.

A scan of the titles revealed that some of them [15] [16] [17] [18] [19] [20] are themselves literature reviews that focus on different aspects of EDM. As a result, according to E3, they have been removed. Papers that do not meet criterion I3 and those papers that have been duplicated from other databases have also been removed. Table III summarize the studies removed, indicating the eligibility criteria and the reason for removal.

TABLE III. STUDIES REMOVED FROM SCOPUS

Reference	Removal reason	Comments
[15][16][17][18][19] [20]	E3	A Systematic Literature Review
[21]	I3	The study contains only 2 pages.
[22][23][24][25][26] [27][28]	-	Duplicates. Articles found in IEEExplore
[29][30][31]	-	Duplicates Articles found in Science Direct, Elsevier

For Science Direct, Elsevier we have removed four articles whose titles showed them to be reviews or surveys and one article referring to Brazil. Analyzing the affiliation of the authors, we also removed seven articles whose authors are all affiliated with institutions in countries outside the EU. Their details are shown in Table IV.

TABLE IV. STUDIES REMOVED FROM SCIENCE DIRECT AND ELSEVIER

Reference	Removal reason	Comments
[7][32][33] [34]	E3	Survey /A Systematic Literature Review
[35]	I2	The research concerns Brazil
[36][37] [38][39] [40][41] [42]	I1	The study does not meet criterion I1. Authors are affiliated to institutions in non-EU countries

Metadata analysis of the 52 results obtained in the IEEEXplore database led to the elimination of 45 papers, based on the failure to comply with the inclusion criteria or to comply with the exclusion criteria, as shown in Table V.

At this stage we obtained the following results: Scopus- 22 studies, Science Direct – 4 studies and IEEExplore 7 studies.

Final studies selection:

For the final selection of articles to be analyzed, we carried out a complete reading of the papers obtained in the previous stage, assessed their quality, and kept only those that fully met the inclusion/exclusion criteria.

TABLE V. STUDIES REMOVED FROM IEEEXPLORE

Reference	Removal reason	Comments
[43]	I2	The research concerns a school in South Africa
[44][45][46]	E3	A Systematic Literature Review
[11][47]	I2	The research concerns India
[48][49][50] [51][52][53][54][55][56]	I1	All authors are affiliated to institutions in China
[57] [58][59] [60][61] [62] [63][64][65] [66] [67] [68][79][80][81]	I1	All authors are affiliated to institutions in Ecuador, Thailand, India and Brazil
[69][70][71] [72] [73] [74] [75]	I1	All authors are affiliated to institutions in Japan, Israel, Serbia, Turkey and United Arab Emirates
[76]	I1	All authors are affiliated to institutions in Great Britain, and the paper was published in 2023, when it is out of EU
[77][78] [82] [83][84] [85]	I1	The study does not meet criterion I1. All authors are affiliated to institutions in Mexico, Argentina, South Africa and USA
[86]	I2	The research concerns Pakistan

According to criterion I1 we considered relevant for our research studies based on authors/data sets from the education system in the European Union countries. We motivate this choice by the fact that Romania - as a member of the European Union - shares many of the educational directives and regulations with other member states. This means that studies from these countries can better reflect the potential challenges, requirements, and opportunities that Romania might face or already faces. Education systems in EU countries tend to pursue a certain alignment with skills and knowledge standards, which can influence approaches to educational data mining, as they target comparable educational outcomes. Although cultural differences are present, education in the EU reflects also a certain cultural uniformity given by European values and principles. Therefore, research based on data from this region may be more relevant for Romania than research analyzing completely different educational systems, such as those in Asia or North America, for example.

At the time of the study, the UK was no longer a member of the European Union, imposing a different legislative and regulatory framework from the EU. Therefore, practices and policy on data mining in education may differ significantly. EU member countries are governed by common regulations and directives, including in the area of data protection (GDPR), which influence how information can be collected and analyzed in the education sector. This gives a more consistent basis of comparability between them, unlike the UK which may now have its own rules.

A full examination of the 22 articles previously obtained in Scopus database revealed that some of them do not meet the inclusion criteria or align with the exclusion criteria. Table VI presents the status of the articles removed in the current stage and the reasons for their removal.

TABLE VI. STATUS OF THE ARTICLES REMOVED FROM SCOPUS

Reference	Removal reason	Comments
[87][88] [89]	E3	A Systematic Literature Review
[90]	I1	The research is based on a dataset from a school in Iraq. The study has seven authors: five are affiliated with educational institutions in Iraq, one author from Hungary, and another from Germany.
[91]	E2	The study is not full accessible
[92]	I3	The paper provides a review of EDM and compares existing techniques, defines the concept of explainability and reviews recent advances in explainable artificial intelligence, assesses the current state of explainability in modern EDM approaches, discusses the multidimensional requirement for educational data mining, highlighting limitations of prediction accuracy metrics, and proposes integrating explainability into future EDM techniques to fill accuracy gaps.
[93]	I1	UK is no longer in the EU
[94][95]	E2	The study is not full accessible
[96]	I2	The authors use the mystery method to explore educational data. This method does not belong to the field of EDM.
[97]	I2	The study has 5 authors, 4 authors from Ecuador, one author from Spain. Although we have gone through the study very carefully, the authors do not specify the country of the educational institution from which the dataset was collected. At the end of the paper thanks are given to the institution Escuela Politécnica Nacional, from which we infer that the research is based on a dataset from a non-EU country.

At the end of the candidate studies search stage in Science Direct, Elsevier 4 research articles remained.

Considering author affiliation, we found some studies that have team authors affiliated to institutions in both EU and non-EU countries. In these cases, we read the whole paper to identify the origin of the underlying dataset. For example, in [98], the author group consists of 3 researchers affiliated with a university in Pakistan and one researcher affiliated with a university in Germany. Near the end of the paper, the authors note that the research was based on a dataset collected from a university in Pakistan, which does not meet inclusion criterion I1.

The article [31] has two authors, one affiliated with a university in Africa, the other with a university in France. Although it passed the four filters initially applied to all studies, we consider it not relevant to our research, because it assesses the affective states and behavior of people who use serious collaborative crisis management games to improve their quality of life.

A surprising situation we encountered in [29]. The paper has four authors: Three authors affiliated with universities in EU member countries (Cyprus, The Netherlands, Finland) and one author affiliated with a university in Australia. In their study the authors research two datasets: a dataset collected from EU member countries and a dataset collected from Australia. We considered that the paper fits the specified criteria. Table VII summarizes the articles removed and the reason for their removal.

TABLE VII. STATUS OF THE ARTICLES REMOVED FROM SCIENCE DIRECT, ELSEVIER

Reference	Removal reason	Comments
[98]	I1	Data collected from Pakistan
[31]	I3	The topic is outside the scope of our research

Finally, we studied the seven articles obtained in the previous step in IEEExplore.

The aim stated in [23] is to identify the influence of thermal conditions in classrooms on learning, to facilitate specific measures to improve these conditions. The research is motivated by the fact that the expense of ensuring energy efficiency in schools represents a very large amount of money allocated from the budget of educational institutions. The authors carried out an analysis of data collected from the GAIA platform to investigate the condition of school buildings in Europe. The GAIA platform deployed a pilot IoT infrastructure in three countries (Greece, Italy, Sweden), monitoring 18 school buildings in real time for electricity consumption and indoor and outdoor environmental conditions. Data collected over 2 years is examined to assess the indoor conditions of classrooms and provide insight into the functioning of these educational buildings. This is the first initiative to establish a quantitative comparison between different buildings and classrooms. The analysis presented here can serve as a tool for school managers and building administrators, helping them to quickly identify classrooms that do not meet standards for indoor environmental conditions and take specific actions to improve them. Data was collected in the cloud using IoT devices. Cloud delivery was not always done correctly due to the instability of the wi-fi connection. Although in the title the authors refer to the use of data mining techniques in the research, they did not describe any of the techniques used, which led to the study being eliminated, according to I2.

We also found an article that is a variant of an analyzed one, and a work that is in line with the exclusion criterion E3. Finally, we removed three more items as shown in Table VIII.

TABLE VIII. PAPERS REMOVED FROM CANDIDATE STUDIES IN IEEEXPLORE

Reference	Removal reason (inclusion/exclusion criterion)	Comments
[23]	I2	The article does not meet the objectives of the review
[24]	E2	The article is a variant of the study [22]. It cannot be accessed because it is not open access.
[25]	E3	The study explores the challenges and opportunities related to the analytical processing of big data generated and stored in higher education institutions. One of its chapters contains a brief overview of the most commonly used EDM techniques in educational research, with the authors stating that the most commonly used are classification techniques.

At the end of each of the three selection rounds the results were as follows:

TABLE IX. NUMBER OF STUDIES SELECTED FOR FINAL REVIEW

Digital Repository	Initial identification	Candidate selection	Final selection
Scopus	326	22	11
Science direct, Elsevier	16	4	2
IEEExplore	52	7	4

F. Reporting

As it is presented in Table IX the final review covers 17 articles. Table X presents a summary of these papers and highlights for each of them, the research objectives, the country whose educational system is targeted, data mining techniques and tools (applications) used, year of publication and the education level addressed in the research.

TABLE X. OVERVIEW OF SELECTED STUDY ANALYSES

Reference	*Objectives*	*Country*	*Methods/techniques*	*Year*	*Educational level*
[22]	Development of a game portal to help first graders to solve their homework, with the aim of analyzing the learning process.	Slovakia	**Classification** (Fuzzy Decision Trees), **Association Rules**	2013	Preuniversity
[26]	Strategies and algorithms for handling and managing missing data, in the context that if each student has control over their data and decides what data to make available for analysis, there is a possibility that this phenomenon will alter the quality of the models. Some classification algorithms were evaluated, simulating missing values on data sets collected from students.	Portugal	**Classification** (Support Vector Machine, Neural Network, Decision Trees, Random Forrest)	2018	Preuniversity
[27]	Exploring logistic regression, support vector machine, k-nearest neighbors, and random forest techniques for predicting student success in solving real-world operational scenarios.	Italy	**Prediction** (Logistic regression), **Classification** (Random Forest, SVM, k-NN)	2021	Preuniversity
[28]	The use of EDM to develop a digital educational resource (DER) for scientific education in primary schools. The paper evaluates the impact of the proposed learning approach on students, focusing on the development of science skills and self-regulated learning.	Portugal	**Prediction**, **Association rules**	2017	Preuniversity
[29]	Understanding students' learning, behavior, and experiences in computer-supported classroom activities.	Cyprus, Olanda, Finland	**Association rules**	2017	University

		Australia			
[30]	Identifying factors associated with the effectiveness of secondary schools using data collected from the Spanish PISA 2015 sample.	Spain	**Classification** (Decision Trees)	2020	Preuniversity
[99]	Understanding the performance of interactions between students with different learning styles and computer-based tools in the problem-solving process by EDM techniques.	Cyprus	**Clustering** (K-means), **Association rules**	2013	University
[100]	Presents the importance of mining log data provided by LMS in problem-based learning (PBL) training, with the aim of improving this method for learning practical skills in healthcare.	Finland	**Visualization**, **Clustering**	2014	University
[101]	Designing a machine learning-based framework to improve the performance of data mining tasks and analyzing the effectiveness of this framework in extracting information related to student performance in a real case study.	Romania	**Prediction** (Regression), Classification	2022	University
[102]	Assessing the usefulness of the Bayesian Profile Regression model for identifying students more likely to drop out of school. By considering students' performance, motivation and resilience, this technique allows the profiling of students at higher risk of school failure	Italy	**Prediction** (Bayesian Profile Regression)	2018	University
[103]	Presents the application of data mining techniques on educational data from event logs downloaded from an e-learning environment.	Croatia	**Clustering**, **Classification** (Decision Trees)	2020	University
[104]	Identifies the effectiveness of classical data mining algorithms vs. autoML in predicting students' early-stage failure, grades, and dropout.	Greece	**Classification** (Naïve Bayes, Decision Tree, Random Forest), **Prediction** (M5Rules)	2020	University
[105]	The paper examines students' ideas about procedural understanding and identifies its core ideas that are essential for understanding scientific inquiry. Two studies are conducted to identify students' conceptions of different steps of inquiry and to present quantitative information related to the criteria of quality in science.	Germany	**Visualization**	2023	Preuniversity
[106]	A complete EDM process, seen as a combination of a data warehouse (DW) specifically designed for educational purposes and data pipelines, whose benefit would be repeatability and adaptability to the specific needs of the educational system is proposed. The functionality of the project is tested by producing Dashboards containing information on online activity, academic performance, and social interaction of students.	Greece	**Clustering**, **Visualization**	2023	University
[107]	The concept of Augmented Intelligence method (AUI) in EDM is introduced. When applied in cycles, the AUI method generates new knowledge from the educational context. The method has been tested by generating several adjustable decision tree models.	Spain	**Classification** (Decision Trees (ID3))	2019	Preuniversity
[108]	A semi-automated method for generating educational competency maps from multiple-choice question repositories using Bayesian structural learning and data mining techniques is proposed.	Spain	**Classification** (Bayesian techniques, k-NN) **Association rule mining**	2019	University
[109]	Study on the quality of life of teachers and their perception on the education system by mining data collected from the opinions of those who are directly involved in this system	Romania	**Classification** (Decision Trees, C4.5)	2013	Preuniversity

IV. DISCUSSIONS

Based on data summarized above, some interesting aspects about the current level of involvement of EDM in increasing the quality of the educational process in European Union can be revealed.

- Data mining techniques have been used mainly at the academic level. 53% of the papers refer to universities while only 47% address problems in the preuniversity environment (Fig. 2).
- Educational Data Mining not only facilitates the efficient discovery of useful patterns and knowledge but supports strategic decision-making in this vital area. Based on the data presented in Table II, we can state that worldwide interest in EDM research is high and justified by the increasing availability of educational data, but also by the development of technologies. In the second selection stage of studies, we noted the interest and upward trend for EDM research in countries such as China, India, Japan.

The number of studies selected in the final phase reflects a medium interest of researchers affiliated to educational institutions in EU member countries (Fig. 3).

This proves the need to continue our research on educational datasets. The aim is to transform data into useful knowledge to contribute to the improvement of educational processes.

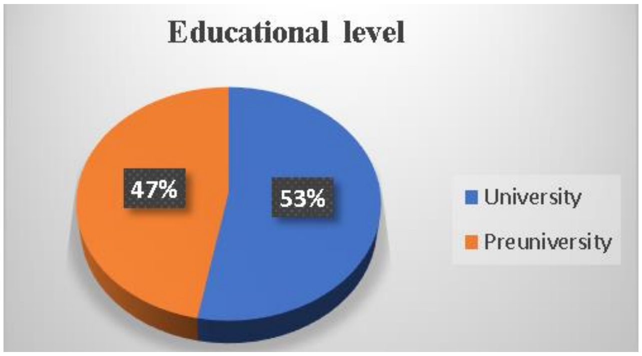

Fig. 2. Education level at which EDM was used.

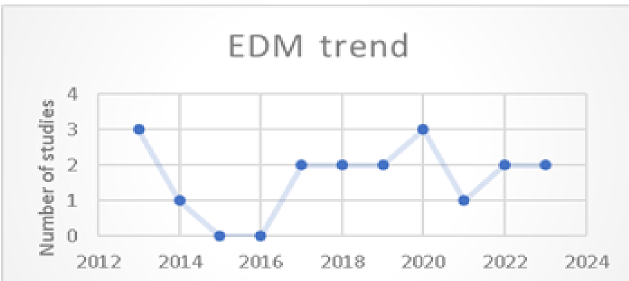

Fig. 3. Trends in EDM research in EU member countries, 2013-2023.

It is necessary to mention that our research includes papers published until September 2023, so it is possible that the number of papers will increase until December 2023.

- EDM uses different methods to extract information. In the reviewed papers we found that classification is most used (Fig. 4), especially decision trees which allowed the development of predictive models, followed by regression as a predictive technique (Fig. 5). To discover relationships between different variables or events in the data, researchers used association rules. Clustering techniques were applied to identify groups with similar needs to develop personalized teaching strategies.

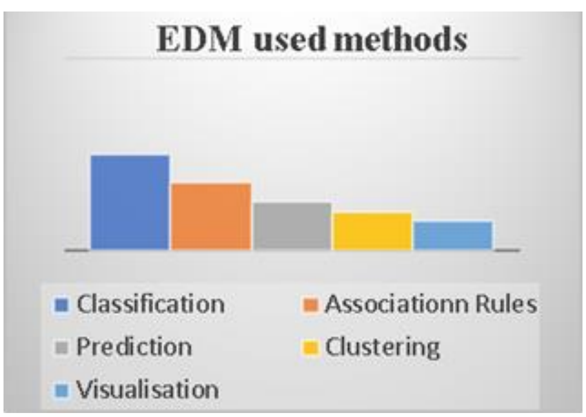

Fig. 4. Data mining methods used in the final selection of studies to be reviewed.

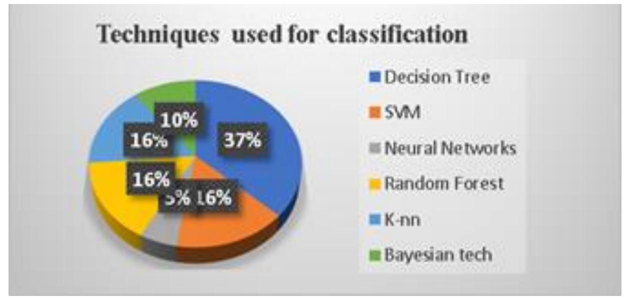

Fig. 5. Data mining methods used for the final selection of studies to be reviewed.

- The current relatively low level of EDM research interest in EU is highlighted by the fact that the 17 articles analyzed come from only 10 of the 28 Member States, as shown in Fig. 6.

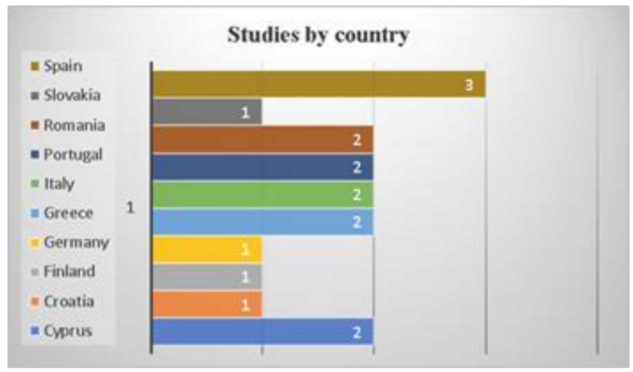

Fig. 6. EU member countries contributing with studies to our research.

Datasets and tools:

In most cases, source datasets were used. These were collected from students and teachers as responses to questionnaires, or by directly retrieving the necessary data from the LMSs used. Some of the sources of these datasets are listed below (Table XI).

TABLE XI. SUMMARY OF ORIGINAL DATASETS SOURCES

Reference	Country	Source of dataset
[26]	Portugal	649 students from secondary schools
[27]	Italy	197 students from secondary schools
[30]	Spain	31236 students from 896 secondary schools
[99]	Cyprus	101 students
[100]	Finland	116 students at Medical School of Tampere Moodle log
[101]	Romania	original dataset collected in Babeș-Bolyai University for three years, for a Computer Science discipline
[102]	Italy	data collected through an online questionnaire completed by 561 undergraduate students of an Italian university
[103]	Croatia	185 students – 59.605 records -from a university in Croatia
[104]	Greece	data collected from learning platforms in Aristotle University of Salonic

[105]	Germany	Study 1: 47 students Study 2: 64 students from a secondary school
[106]	Greece	Hellenic Open University
[108]	Spain	archives of multiple-choice test answers for 12 exams between 2004-2006
[109]	Romania	105 teachers from schools in Cluj Napoca

From the provided data concerning the used tools, data presented in Table XII, and in Fig. 7 it is evident that the most used environments for designing and executing data mining processes were Weka and RapidMiner.

Finally, we propose a brief discussion of the insights gained from this research which aims to highlight the current state of research, achievements, and challenges that induce the need for further directions of work.

TABLE XII. SUMMARY OF RESULTS FROM THE SELECTED STUDIES ANALYSIS

Reference	Tools /applications
[29]	Statistica, Data Miner
[30]	HLM 7
[99]	Model-It ®
[100]	SPY US
[101]	IntelliDaM
[102]	Freel R package, PReMiuM
[103]	RapidMiner
[104]	Weka
[106]	KNIME
[107]	Weka
[108]	Python, R, GNU Octave
[109]	Rapid Miner, Weka

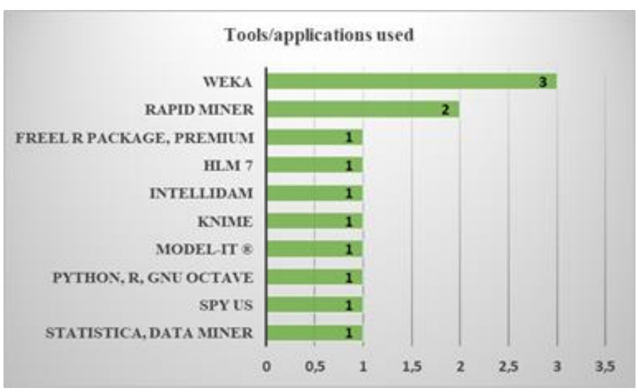

Fig. 7. Most used tools in studies of our research.

The methods used most frequently in extracting patterns from the data in the studies reviewed are classification followed by association rule discovery.

- Classification was used to understand the learning process of primary school children aiming to identify individual and group patterns and to design appropriate educational resources to maintain their motivation and attention [22]. In [27], classification was used to predict student performance in solving problems raised by real scenarios and in [30] to identify factors associated with the effectiveness of secondary schools. Other topics related to the educational process were also covered. In [109] a study of the quality of life of teachers in the pre-university environment and their perception on the education system was carried out, exploring a dataset containing the opinions of 105 teachers in Cluj-Napoca (Romania). Beyond finding patterns directly related to educational activity, research has also been targeted on the usefulness of exploring data from the logs of the LMSs used (mainly Moodle) [100] [103] or the effect of introducing of Augmented Intelligence method in EDM [107].
- Association rules as rule-based statements, that aim to find interesting relationships between data items in large datasets, were used in [108] to discover the relationships between answers to different test questions in order to be able to infer from them the relationships between competences in an educational area. In [28] a subtype of these, called Causal Data Mining, was used to find causal relationships between different events. In [29] association rules were a useful method for finding relationships between learners' use of simulation and their performance in this context. Another variant of association rules, namely Sequence Association, has been used for the purpose of referencing an immediately following action as a function of a previous one [99].

At the same time, the reviewed studies revealed a variety of challenges, including:

- Data quality: the quality of educational data can vary and can be influenced by human error, incorrect input or different data sources that are not standardized. Ensuring data quality is essential for accurate results.
- Confidentiality and ethics: Educational data may contain sensitive information about students and teachers. It is crucial to protect the confidentiality of this data and to respect ethical standards regarding its collection, storage, and use. In [26] the issue of confidentiality and transparency is addressed, and the idea that each student should have control over the data they make available in the dataset is stated. It is found, however, that this approach results in datasets with a lot of missing data, which can seriously distort the quality of the acquired information.
- Generalizing results: Some patterns or findings may be specific to one context or group of learners and may be difficult to generalize to wider or other educational situations.
- Evaluation and validation: Rigorous evaluation of models and methods is essential to ensure that the results obtained are valid and reliable. However, there are cases where model validation is still subjective, with no metrics being used to assess the results obtained [108].

- Prediction vs. interpretation: Another issue is the balance between creating accurate predictive models and understanding the reasons behind these predictions. To have a meaningful impact on education, it is important to be able to explain why and how certain conclusions are reached.

V. Limitations

The limitations that may affect the relevance and applicability of the results in the diverse educational context of EU member countries are multiple.

One of the main barriers is the costs associated with obtaining access to the literature. We consider this to be a significant barrier to in-depth analysis of the field of Educational Data Mining (EDM). In our approach, our ability to consult published materials was limited by the resources available through the University "Ștefan cel Mare" of Suceava.

On the other hand, the accessibility and quality of data can differ substantially between EU countries. While some countries may have extensive and well-organized digital resources, others may still be in the early stages of developing the infrastructure necessary for effective research. Thus, the literature may disproportionately focus on research from countries with more advanced infrastructure, providing an incomplete picture.

The applicability of survey findings across the EU can also be difficult due to socio-economic and educational differences. It is important to put the results in context to ensure correct interpretations and to avoid inappropriate extrapolations.

We believe that a careful and reflexive approach is essential in any effort to reduce limitations and maximize the potential of EDM in a unified, yet diverse, European educational area.

VI. Conclusions

The originality of our research lies in providing an updated perspective on the state of research and current dynamics in the field of Educational Data Mining (EDM) within the European Union. Our work aims to assess the level of integration of EDM in the educational systems of EU Member States, exploring the extent of use of these innovative technologies at different levels of education. We believe that the level of interest shown in EDM by the academic and educational communities in EU member countries can be improved. On the other hand, we believe that to contribute to quality improvement in education through the use of EDM techniques, open access to EDM research is imperative.

Regarding the impact, the research emphasizes the transformative potential of EDM in tailoring educational experiences to meet the individual needs of students, thus enhancing the personalization of learning. It also highlights the essential role of EDM in predicting students' academic performances, which allows educators to provide timely support to students at risk of poor performance. Moreover, the study showcases the broader applications of EDM in refining educational strategies and addressing systemic issues, such as reducing school dropout rates and addressing the root causes of poor performance among students and teachers.

This approach advances the understanding of the role of EDM in education and also sets the "scene" for future innovations and improvements in the educational landscape.

Within the analytical approach, the paper is structured around some essential research questions. The first question investigates the extent to which EDM practices are embedded in the educational mechanisms of the European Union, revealing the level of penetration and acceptance of these methods among educational institutions.

The fact that for the period under review, i.e. 2013-2023, only two of the articles are from Romania, draws our attention to a modest presence in the literature in the field at national level. This highlights an untapped potential and the opportunity to further explore the contributions that EDM can bring to Romanian pre-university education, where the adoption of such analytical technologies could catalyze significant advances in the adaptation and personalization of the educational process. We thus justify our approach to use EDM techniques with the aim of improving the quality of pre-university education in Romania.

Research on the use of EDM in the European Union over the last decade has focused mainly on higher education. One of the reasons could be that at this level LMSs are more widely used, allowing relatively easy data collection.

Interest in applying EDM methods in EU member countries is low. This is reflected on the one hand, by the small number of papers that met our selection criteria, to which authors from less than half of these countries contributed, and on the other hand, by the trend of published research in the considered interval (Fig. 3).

The analysis shows that understanding learning and the educational process through data mining techniques can provide deep and detailed insight into how students learn, behave, interact, and progress. Summarizing our research, it can be stated that EDM techniques contribute to:

- Personalizing learning by identifying individual learning patterns for each student or for groups of students with similar characteristics. By data mining one can discover the learning preferences and rhythms of everyone or group. This makes it easier to personalize content and teaching methods, ensuring that each student/group gets the support they need;

- Predicting performance by developing predictive models to predict student academic performance. This can help teaching staff to early identify students who may be struggling and to intervene with additional measures;

- Improving educational processes by using patterns uncovered in data about the effectiveness of teaching methods, learning materials or educational resources. In this way, educational institutions can optimize their strategies to improve students' growth and learning experience;

- Structural problems identification by detecting and combating high dropout rates, and factors contributing to low student or teacher performance. The insights hidden in data can help the development of educational policies and the efficient allocation of resources.

In 2020 the European Parliament conducted a study on “rethinking education in the digital age” in which it stressed the need for learning content personalization, and for the monitoring and control of learners' behavior, which overlaps very well with the objectives of EDM. As until now efforts in this area have been modest, it opens up wide perspectives for research to address issues related to:

- the widespread use of EDM methods and techniques applied to real data, from all sources of educational data, generated by individuals or groups of individuals in institutional frameworks, the results of which form the basis for quality decisions;
- confidentiality and ethics of data collection and analysis, protocols for developing appropriate procedures to preserve data and protect student privacy;
- the use of multimodal data mining methods, knowing that the data collected can be of different types (images, text, structured data), and a multimodal approach can better capture the subtleties.

Future research directions discussed in the material include:

- Integration of EDM Tools into E-Learning: Advancing the incorporation of Educational Data Mining (EDM) components into e-learning systems and tools for designing e-learning courses.
- Cross-Domain Application: Exploring the application of the semi-automatic generation of competency maps across various educational domains beyond the initially tested field.
- Methodology Refinement: Further developing and refining the presented methodology based on additional experiments and feedback, aiming to establish a more robust EDM process and software platform.
- Automation and Teacher Support: Enhancing the level of automation in the generation of competency maps and educational recommendations, thereby providing more significant support to teachers and learners.
- These directions aim to expand the scope and efficacy of EDM in improving educational outcomes and tailoring learning experiences.

Thus, it highlights the need for a continuous effort in the use and improvement of educational tools and strategies through data-driven perspectives, with the goal of supporting educators and enhancing learning outcomes.

REFERENCES

[1] M. Danubianu, ”A data preprocessing framework for students' outcome prediction by data mining techniques”, 2015 19th International Conference on System Theory, Control and Computing (ICSTCC), Cheile Gradistei, Romania, 2015, pp. 836-841, doi: 10.1109/ICSTCC.2015.7321398.

[2] Feng, G., & Fan, M. ”Research on learning behavior patterns from the perspective of educational data mining: Evaluation, prediction and visualization. Expert Systems with Applications”, 121555, 2023.

[3] Harikumar, Smitha, ”A Study on Educational Data Mining.” International Journal of Computer Trends and Technology. vol 8. pp. 90-95, 2014, 10.14445/22312803/IJCTT-V8P117.

[4] C. Romero and S. Ventura, “Educational data mining: a survey from 1995 to 2005”, Expert Systems with Applications, vol. 33, no. 1, pp. 135–146, 2007.

[5] C. Romero and S. Ventura, “Educational data mining: A review of the state of the art”, IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews, vol. 40, no. 6, pp. 601–618, 2010.

[6] C. Romero and S. Ventura, “Data mining in education”, Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol. 3, no. 1, pp. 12–27, 2013.

[7] Pena-Ayala, “Educational data mining: A survey and a data mining-based analysis of recent works”, Expert Systems with Applications, vol. 41, no. 4, pp. 1432–1462, 2014.

[8] Dutt, A., Ismail, M. A., & Herawan, T. ”A systematic review on educational data mining.” Ieee Access, 5, 15991-16005, 2017.

[9] Albreiki, B., Zaki, N., & Alashwal, H. ”A systematic literature review of student’ performance prediction using machine learning techniques.” Education Sciences, 11(9), 552, 2021.

[10] Kitchenham, Barbara & Charters, Stuart. ”Guidelines for performing Systematic Literature Reviews in Software Engineering.” 2, 2007.

[11] S. Pulakhandam și N. Patil, ”Recomandation of Optimal Locations for Government Financed Educational Institutes in Urban India Using a Hybrid Data Mining Technique”, 2015 Second International Conference on Advances in Computing and Communication Engineering, Dehradun, India, 2015, pp. 560 -567, doi: 10.1109/ICACCE.2015.140.

[12] Abu-Rumman, Ayman. ”Scopus Content Coverage Guide”, January 2019.

[13] A. Valente, M. Holanda, A. M. Mariano, R. Furuta and D. Da Silva, ”Analysis of Academic Databases for Literature Review in the Computer Science Education Field”, 2022 IEEE Frontiers in Education Conference (FIE), Uppsala, Sweden, pp. 1-7, 2022, doi: 10.1109/FIE56618.2022.9962393.

[14] Jeroen Baas, Michiel Schotten, Andrew Plume, Grégoire Côté, Reza Karimi, ”Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies”. Quantitative Science Studies 2020; vol 1 (1): pp. 377–386. doi: https://doi.org/10.1162/qss_a_00019.

[15] Okewu, Emmanuel & Adewole, Phillip & Misra, Sanjay & Maskeliunas, Rytis & Damaševičius, Robertas. ”Artificial Neural Networks for Educational Data Mining in Higher Education: A Systematic Literature Review. Applied Artificial Intelligence.” 35. 983-1021. 10.1080/08839514.2021.1922847, 2021.

[16] Bošnjaković, Natalija & Đurđević Babić, Ivana. ”Systematic Review on Educational Data Mining in Educational Gamification.” Technology, Knowledge and Learning. 1-18. 10.1007/s10758-023-09686-2., 2023.

[17] Charitopoulos, Angelos & Rangoussi, Maria & Koulouriotis, Dimitrios. ”On the Use of Soft Computing Methods in Educational Data Mining and Learning Analytics Research: a Review of Years 2010–2018.” International Journal of Artificial Intelligence in Education. 30. 10.1007/s40593-020-00200-8, 2020.

[18] Agrusti, F., Bonavolontà, G., & Mezzini, M. ”University Dropout Prediction through Educational Data Mining Techniques: A Systematic Review.” Journal of E-Learning and Knowledge Society, vol 15(3), 161-182. https://doi.org/10.20368/1971-8829/1135017, 2019.

[19] Papamitsiou, Zacharoula & Economides, Anastasios. ”Learning Analytics and Educational Data Mining in Practice: A Systematic Literature Review of Empirical Evidence.” Educational Technology & Society. 17. 49-64, 2014.

[20] Ihantola, Petri & Hellas, Arto & Butler, Matthew & Börstler, Jürgen & Edwards, Stephen & Isohanni, Essi & Korhonen, Ari & Petersen, Andrew & Rivers, Kelly & Rubio, Miguel & Sheard, Judy & Skupas, Bronius & Spacco, Jaime & Szabo, Claudia & Toll, Daniel. ”Educational Data Mining and Learning Analytics in Programming: Literature Review and Case Studies.” 10.1145/2858796.2858798., 2015.

[21] Jormanainen, I., & Sutinen, E. ”An open approach for learning educational data mining.”, In Proceedings of the 13th Koli Calling International Conference on Computing Education Research (pp. 203-204)., November, 2013.

[22] Zaitseva, Elena & Vitaly, Levashenko & Kostolny, Jozef & Kvassay, Miroslav. ”A multi-valued decision diagram for estimation of multi-state system.”, IEEE EuroCon 2013. 645-650. 10.1109/EUROCON.2013.6625049., 2013.

[23] N. Zhu, A. Anagnostopoulos și I. Chatzigiannakis, „On Mining IoT Data for Evaluating the Operation of Public Educational Buildings”, 2018 IEEE International Conference on Pervasive Computing and Communications Workshops (PerCom Workshops), Atena, Grecia, doi: 10.1109/PERCOMW.2018.8480226., pp. 278-283, 2018.

[24] Vitaly, Levashenko & Zaitseva, Elena & Kostolny, Jozef & Kvassay, Miroslav. ”Educational portal with data mining support based on modern technologies.”, 1-6. 10.1109/ICETA.2015.7558490., 2015.

[25] Stefanova, Kamelia & Kabakchieva, Dorina. ”Educational data mining perspectives within university big data environment.”, 10.1109/ICE.2017.8279898., pp. 264-270, 2017.

[26] A. Askinadze and S. Conrad, ”Respecting Data Privacy in Educational Data Mining: An Approach to the Transparent Handling of Student Data and Dealing with the Resulting Missing Value Problem”, 2018 IEEE 27th International Conference on Enabling Technologies: Infrastructure for Collaborative Enterprises (WETICE), Paris, France, doi: 10.1109/WETICE.2018.00037., pp. 160-164, 2018.

[27] Scaradozzi D., Cesaretti L., Screpanti L., Mangina E., ”Identification and Assessment of Educational Experiences: Utilizing Data Mining with Robotics”, IEEE Robotics and Automation Magazine, DOI: 10.1109/MRA.2021.3108942, vol 28 (4), pp. 103 – 113, 2021.

[28] Tavares, Rita & Vieira, Rui & Pedro, Luis. ”A preliminary proposal of a conceptual Educational Data Mining framework for Science Education Scientific competences development and self-regulated learning”, 2017.

[29] Angeli, Charoula & Howard, Sarah & Ma, Jun & Yang, Jie & Kirschner, Paul. ”Data mining in educational technology classroom research: Can it make a contribution?”, Computers & Education. 113. 10.1016/j.compedu.2017.05.021., 2017.

[30] Martínez Abad, Fernando & Gamazo, Adriana & Conde, María José. ”Educational Data Mining: Identification of factors associated with school effectiveness in PISA assessment.”, Studies in Educational Evaluation. 66. 100875. 10.1016/j.stueduc.2020.100875., 2020.

[31] Daoudi, Ibtissem & Chebil, Raoudha & Tranvouez, Erwan & Lejouad Chaari, Wided & Espinasse, Bernard. ”Improving Learners’ Assessment and Evaluation in Crisis Management Serious Games: An Emotion-based Educational Data Mining Approach.”, Entertainment Computing. 38. 10.1016/j.entcom.2021.100428., 2021.

[32] Shaik, T., Tao, X., Dann, C., Xie, H., Li, Y., & Galligan, L. ”Sentiment analysis and opinion mining on educational data: A survey. Natural Language Processing Journal”, 100003., 2022.

[33] Aulakh, K., Roul, R. K., & Kaushal, M. ”E-learning enhancement through Educational Data Mining with Covid-19 outbreak period in backdrop: A review.”, International Journal of Educational Development, 102814., 2023.

[34] Dol, S. M., & Jawandhiya, P. M., ”Classification Technique and its Combination with Clustering and Association Rule Mining in Educational Data Mining—A survey.”, Engineering Applications of Artificial Intelligence, 122, 106071., 2023.

[35] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G., ”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., 2019.

[36] Silva Filho, R. L. C., Brito, K., & Adeodato, P. J. L., ”A data mining framework for reporting trends in the predictive contribution of factors related to educational achievement.”, Expert Systems with Applications, 221, 119729., 2023.

[37] Injadat, M., Moubayed, A., Nassif, A. B., & Shami, A. ”Systematic ensemble model selection approach for educational data mining.”, Knowledge-Based Systems, 200, 105992, 2020.

[38] Adekitan, A. I., & Salau, O. ”The impact of engineering students' performance in the first three years on their graduation result using educational data mining.”, Heliyon, vol 5(2)., 2019.

[39] Lemay, D. J., Baek, C., & Doleck, T., ”Comparison of learning analytics and educational data mining: A topic modeling approach.”, Computers and Education: Artificial Intelligence, 2, 100016., 2021.

[40] Xing, W., Guo, R., Petakovic, E., & Goggins, S., ”Participation-based student final performance prediction model through interpretable Genetic Programming: Integrating learning analytics, educational data mining and theory.”, Computers in human behavior, vol 47, pp. 168-181, 2015.

[41] Gobert, J. D., Kim, Y. J., Sao Pedro, M. A., Kennedy, M., & Betts, C. G., ”Using educational data mining to assess students’ skills at designing and conducting experiments within a complex systems microworld.”, Thinking Skills and Creativity, vol 18, pp. 81-90, 2015.

[42] Cabrera, D. F. M., & Zareipour, H., ”Data association mining for identifying lighting energy waste patterns in educational institutes.”, Energy and Buildings, vol 62, pp. 210-216, 2013.

[43] Ramaphosa, Khokhoni & Zuva, Tranos & Kwuimi, Raoul. ”Educational Data Mining to Improve Learner Performance in Gauteng Primary Schools.”, 1-6. 10.1109/ICABCD.2018.8465478., 2018.

[44] Shafiq, Dalia & Marjani, Mohsen & Ariyaluran Habeeb, Riyaz Ahamed & Asirvatham, David, ”Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review.”, IEEE Access. 10. 10.1109/ACCESS.2022.3188767., 2022.

[45] Khanna, Leena & Singh, Shailendra & Alam, Mansaf. ”Educational data mining and its role in determining factors affecting students academic performance: A systematic review.”, 1-7. 10.1109/IICIP.2016.7975354., 2016.

[46] AI Al-Alawi, MAA Alfateh and AM Alrayes, ”Educational Data Mining Utilization to Support the Admission Process in Higher Education Institutions: A Systematic Literature Review”, 2023 International Conference On Cyber Management And Engineering (CyMaEn), Bangkok, Thailand, doi: 10.1109/CyMaEn57228.2023.10051077., pp. 332-339, 2023.

[47] Bijoy, Md. Hasan Imam & Pramanik, Anik & Rahman, Md & Hasan, Mehedi & Akhi, Sumiya & Rahman, Md. Mahbubur. ”MKRF Stacking-Voting: A Data Mining Technique for Predicting Educational Satisfaction Level of Bangladeshis Student During Pandemic.”, 10.1109/I2CT54291.2022.9824357., 2022.

[48] Fang Hai-guang, Wang Xiao-chun, Hou Wei-feng and Chu Yun-hai, ”Research on educational data mining of digital learning process for elementary school”, 2014 9th International Conference on Computer Science & Education, Vancouver, BC, Canada, doi: 10.1109/ICCSE.2014.6926582, pp. 849-854, 2014.

[49] Gao, Xiaopeng & Ruan, Shuai & Wang, Xuejiao & Ji, Shufan., ”Mining Relations between Courses and Research Directions from Educational Data.”, 815-818. 10.1109/HPCC-CSS-ICESS.2015.167., 2015.

[50] Y. Wang, L. Xu, Q. Wang, H. Lv and Y. Zhang, ”Educational Data Mining and Learning Analysis System Based on Python”, 2022 12th International Conference on Information Technology in Medicine and Education (ITME), Xiamen, China, doi: 10.1109/ITME56794.2022.00122., pp. 559-563, 2022.

[51] T. Yang, B. Chen, W. Wang and S. Li, ”Research on Feedback Service for Teaching Based on Educational Data Mining”, 2022 International Conference on Machine Learning and Knowledge Engineering (MLKE), Guilin, China, doi: 10.1109/MLKE55170.2022.00065., pp. 306-309, 2022.

[52] B. Guo, R. Zhang, G. Xu, C. Shi and L. Yang, ”Predicting Students Performance in Educational Data Mining”, 2015 International Symposium on Educational Technology (ISET), Wuhan, China, doi: 10.1109/ISET.2015.33., pp. 125-128, 2015.

[53] T. Zeng, ”The research and practice of a five-sided educational data mining framework”, 2017 IEEE 3rd Information Technology and Mechatronics Engineering Conference (ITOEC), Chongqing, China, doi: 10.1109/ITOEC.2017.8122514, pp. 1050-1053, 2017.

[54] N. Cheng, L. Wang, R. Fei, W. Li and B. Wang, ”Workflow Model Mining Based On Educational Management Data Logs”, 2019 Chinese

Control And Decision Conference (CCDC), Nanchang, China, doi: 10.1109/CCDC.2019.8832988., pp. 5450-5455, 2019.

[55] W. Han, D. Jun, G. Xiaopeng and L. Kangxu, ”Supporting quality teaching using educational data mining based on OpenEdX platform”, 2017 IEEE Frontiers in Education Conference (FIE), Indianapolis, IN, USA, doi: 10.1109/FIE.2017.8190730., pp. 1-7, 2017.

[56] G. Feng, M. Fan and Y. Chen, ”Analysis and Prediction of Students’ Academic Performance Based on Educational Data Mining”, in IEEE Access, doi: 10.1109/ACCESS.2022.3151652., vol. 10, pp. 19558-19571, 2022.

[57] J. P. Zaldumbide Proaño and V. C. Párraga Villamar, ”Systematic Mapping Study of Literature on Educational Data Mining to Determine Factors That Affect School Performance”, 2018 International Conference on Information Systems and Computer Science (INCISCOS), Quito, Ecuador, doi: 10.1109/INCISCOS.2018.00042., pp. 239-245, 2018.

[58] Kerdprasop, Nittaya & Kerdprasop, Kittisak. ”Educational attainment trend analysis with the visual data mining tool.”, 10.1109/UMEDIA.2015.7297451., pp. 180-185, 2015.

[59] Rojanavasu, Pornthep. ”Educational Data Analytics using Association Rule Mining and Classification.”, 10.1109/ECTI-NCON.2019.8692274., pp. 142-145, 2019.

[60] P. Nuankaew, D. Teeraputon, W. Nuankaew, K. Phanniphong, S. Imwut and S. Bussaman, ”Perception and Attitude Toward Self-Regulated Learning in Educational Data Mining”, 2019 6th International Conference on Technical Education (ICTechEd6), Bangkok, Thailand, doi: 10.1109/ICTechEd6.2019.8790875., pp. 1-5, 2019.

[61] P. Nuankaew, P. Nasa-Ngium and W. S. Nuankaew, ”Self-Regulated Learning Styles in Hybrid Learning Using Educational Data Mining Analysis”, 2022 26th International Computer Science and Engineering Conference (ICSEC), Sakon Nakhon, Thailand, doi: 10.1109/ICSEC56337.2022.10049322., pp. 208-212, 2022.

[62] Hegde, Vinayak & H S, Sushma. ”A Framework to Analyze Performance of Student's in Programming Language Using Educational Data Mining”, 10.1109/ICCIC.2017.8524244., pp. 1-4, 2018.

[63] Ritambhara and S. N. Singh, ”Creativity: Mining of Innovative Thinking Using Educational Data”, 2023 International Conference on Disruptive Technologies (ICDT), Greater Noida, India, doi: 10.1109/ICDT57929.2023.10150690., pp. 445-449, 2023.

[64] S. Srivastava, S. Karigar, R. Khanna and R. Agarwal, ”Educational Data Mining: Classifier Comparison for the Course Selection Process”, 2018 International Conference on Smart Computing and Electronic Enterprise (ICSCEE), Shah Alam, Malaysia, doi: 10.1109/ICSCEE.2018.8538434., pp. 1-5, 2018.

[65] V. R. L. B. da Silva, F. de Albuquerque Silva and V. Burégio, ”Characterizing Educational Data Mining”, 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), Coimbra, Portugal, doi: 10.23919/CISTI.2019.8760815., pp. 1-5, 2019.

[66] Santos, Kelly & Menezes, Angelo & Carvalho, Andre & Montesco, Carlos. ”Supervised Learning in the Context of Educational Data Mining to Avoid University Students Dropout” 10.1109/ICALT.2019.00068., pp. 207-208 2019.

[67] E. M. Queiroga et al., ”Experimenting Learning Analytics and Educational Data Mining in different educational contexts and levels”, 2022 XVII Latin American Conference on Learning Technologies (LACLO), Armenia, Colombia, doi: 10.1109/LACLO56648.2022.10013478., 2022, pp. 1-9.

[68] A. S. Torcate and C. M. de Oliveira Rodrigues, ”Educational Data Mining with Learning Analytics and Unsupervised Algorithms: Analysis and Diagnosis in Basic Education”, 2021 XVI Latin American Conference on Learning Technologies (LACLO), Arequipa, Peru, doi: 10.1109/LACLO54177.2021.00014., pp. 67-74, 2021.

[69] Abdar, Moloud & Zomorodi, Mariam & Zhou, Xujuan. ”An Ensemble-Based Decision Tree Approach for Educational Data Mining”, 10.1109/BESC.2018.8697318., pp. 126-129, 2018.

[70] O. Sukhbaatar, K. Ogata and T. Usagawa, ”Mining Educational Data to Predict Academic Dropouts: a Case Study in Blended Learning Course”, TENCON 2018 - 2018 IEEE Region 10 Conference, Jeju, Korea (South), doi: 10.1109/TENCON.2018.8650138., pp. 2205-2208, 2018.

[71] M. Abdar, N. Y. Yen and J. C. Hung, ”Educational Data Mining Based on Multi-objective Weighted Voting Ensemble Classifier”, 2017 International Conference on Computational Science and Computational Intelligence (CSCI), Las Vegas, NV, USA, doi: 10.1109/CSCI.2017.192., pp. 357-362,2017.

[72] P. Galván, ”Educational evaluation and prediction of school performance through data mining and genetic algorithms”, 2016 Future Technologies Conference (FTC), San Francisco, CA, USA, doi: 10.1109/FTC.2016.7821617., pp. 245-249, 2016.

[73] G. Dimić, D. Rančić, O. P. Rančić and P. Spalević, ”Descriptive Statistical Analysis in the Process of Educational Data Mining”, 2019 14th International Conference on Advanced Technologies, Systems and Services in Telecommunications (TELSIKS), Nis, Serbia, doi: 10.1109/TELSIKS46999.2019.9002177., pp. 388-391, 2019.

[74] F. Ünal and D. Birant, ”Educational Data Mining Using Semi-Supervised Ordinal Classification”, 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Ankara, Turkey, doi: 10.1109/HORA52670.2021.9461278., pp. 1-5, 2021.

[75] N. Eleyan, M. Al Akasheh, E. F. Malik and O. Hujran, ”Predicting Student Performance Using Educational Data Mining”, 2022 Ninth International Conference on Social Networks Analysis, Management and Security (SNAMS), Milan, Italy, doi: 10.1109/SNAMS58071.2022.10062500., pp. 1-7, 2022.

[76] K. Okoye, S. Islam, U. Naeem and S. Hosseini, ”Semantic Data Engineering for Intelligent Educational Learning Systems through Process Mining”, 2023 Future of Educational Innovation-Workshop Series Data in Action, Monterrey, Mexico, doi: 10.1109/IEEECONF56852.2023.10105072., pp. 1-6, 2023.

[77] E. Mondragon-Estrada and C. Camacho-Zuñiga, ”Undergraduate's Perspective on Being an Effective Online Student During Lockdown due to COVID-19 Pandemic: An Educational Data Mining Study”, 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, Monterrey, Mexico, doi: 10.1109/IEEECONF53024.2021.9733773., pp. 1-5, 2021.

[78] S. M. De Oca, M. Villada-Balbuena and C. Camacho-Zuñiga, ”Professors' Concerns after the Shift from Face-to-face to Online Teaching amid COVID-19 Contingency: An Educational Data Mining analysis”, 2021 Machine Learning-Driven Digital Technologies for Educational Innovation Workshop, Monterrey, Mexico, doi: 10.1109/IEEECONF53024.2021.9733778., pp. 1-5, 2021.

[79] A. A. Mehta and N. J. Buch, ”Depth and breadth of educational data mining: Researchers' point of view”, 2016 10th International Conference on Intelligent Systems and Control (ISCO), Coimbatore, India, doi: 10.1109/ISCO.2016.7727075., pp. 1-6, 2016.

[80] K. Kaur and O. Dahiya, ”Role of Educational Data Mining and Learning Analytics Techniques Used for Predictive Modeling”, 2023 3rd International Conference on Innovative Practices in Technology and Management (ICIPTM), Uttar Pradesh, India, doi: 10.1109/ICIPTM57143.2023.10117779., pp. 1-6, 2023.

[81] D. G. Chandra and A. C. Raman, „Educational Data Mining on Learning Management Systems Using SCORM”, 2014 Fourth International Conference on Communication Systems and Network Technologies, Bhopal, India, doi: 10.1109/CSNT.2014.91., pp. 362-368, 2014.

[82] G. N. Rangone, G. A. Montejano, A. G. Garis, C. A. Pizarro and W. R. Molina, ”An Educational Data Mining Model based on Auto Machine Learning and Interpretable Machine Learning”, 2022 IEEE Global Conference on Computing, Power and Communication Technologies (GlobConPT), New Delhi, India, doi: 10.1109/GlobConPT57482.2022.9938243., pp. 1-6, 2022.

[83] J. Mandlazi, A. Jadhav and R. Ajoodha, ”Educational data mining: using knowledge tracing as a tool for student success”, 2021 3rd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Windhoek, Namibia, doi: 10.1109/IMITEC52926.2021.9714608., pp. 1-7, 2021.

[84] N. Ndou, R. Ajoodha and A. Jadhav, ”Educational Data-mining to Determine Student Success at Higher Education Institutions”, 2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC), Kimberley, South Africa, doi: 10.1109/IMITEC50163.2020.9334139., pp. 1-8, 2020.

[85] C. Baek and T. Doleck, ”Educational Data Mining: ”A Bibliometric Analysis of an Emerging Field”, in IEEE Access, doi: 10.1109/ACCESS.2022.3160457., vol. 10, pp. 31289-31296, 2022.

[86] Ullah, M. R., Shahzad, S. K., & Naqvi, M. R. ”Challenges and Opportunities for Educational Data Mining in Pakistan”. In 2019 International Conference on Engineering and Emerging Technologies (ICEET). IEEE., pp. 1-6, February, 2019.

[87] Švábenský, Valdemar & Vykopal, Jan & Celeda, Pavel & Kraus, Lydia. ”Applications of educational data mining and learning analytics on data from cybersecurity training”, Education and Information Technologies. 27. 10.1007/s10639-022-11093-6., 2022.

[88] Vahdat, Mehrnoosh & Oneto, Luca & Ghio, Alessandro & Anguita, Davide & Funk, Mathias & Rauterberg, Matthias. ”Advances in learning analytics and educational data mining”, 2015.

[89] Ahrens, Andreas & Gruenwald, Norbert & Zascerinska, Jelena & Melnikova, Julija. ”A Novel Design of the Pre-Processing Stage of Data Mining for Educational Purposes”, Balkan Region Conference on Engineering and Business Education. 10.2478/cplbu-2020-0042., vol 1. pp. 353-361, 2019.

[90] Najm, I. & Mohammed Dahr, Jasim & Khalaf, Alaa & Hashim, Ali & Akeel, Wid & Kamel, Mohammed B. & Humadi, Aqeel. ”OLAP Mining with Educational Data Mart to Predict Students’ Performance. Informatica. 46. 10.31449/inf.v46i5.3853., 2022.

[91] Chytas, K., Tsolakidis, A., Triperina, E. and Skourlas, C. (), ”Educational data mining in the academic setting: employing the data produced by blended learning to ameliorate the learning process”, Data Technologies and Applications,. https://doi.org/10.1108/DTA-06-2022-0252, vol 57, No. 3, pp. 366-384, 2023.

[92] Tousside, Basile & Dama, Yashwanth & Frochte, Jörg. ”Towards Explainability in Modern Educational Data Mining: A Survey”, 10.5220/0011529400003335, pp. 212-220. 2022.

[93] Moodley, Raymond & Chiclana, Francisco & Carter, Jenny & Caraffini, Fabio. ”Using Data Mining in Educational Administration - A Case Study on Improving School Attendance”, Applied Sciences. 10.3390/app10093116., 2020.

[94] Gómez-Rey, Pilar & Fernández-Navarro, Francisco & Barberà, Elena. ”Ordinal regression by a gravitational model in the field of educational data mining”, Expert Systems. 33. n/a-n/a. 10.1111/exsy.12138., 2015.

[95] Santos, Olga C. & G. Boticario, Jesus. ”User Centred Design and Educational Data Mining support during the Recommendations Elicitation Process in Social Online Learning Environments”, Expert Systems. ??. ??. 10.1111/exsy.12041., 2014.

[96] Benninghaus, Jens & Mühling, Andreas & Kremer, Kerstin & Sprenger, Sandra. ”Complexity in Education for Sustainable Consumption—An Educational Data Mining Approach using Mysteries. Sustainability”, 11. 722. 10.3390/su11030722., 2019.

[97] Peñafiel, Myriam & Vásquez, Maria & Vásquez, Diego & Zaldumbide, Juan & Luján-Mora, Sergio. ”Data Mining and Opinion Mining: A Tool in Educational Context”, 10.1145/3274250.3274263, pp. 74-78, 2018.

[98] Asif, Raheela & Merceron, Agathe & Abbas, Dr-Syed & Haider, Najmi. ”Analyzing undergraduate students' performance using educational data mining”, Computers & Education. 113. 10.1016/j.compedu.2017.05.007., 2017.

[99] Angeli, Charoula. ”Using educational data mining methods to assess field-dependent and field-independent learners’ complex problem solving”, Educational Technology Research and Development. 61. 10.1007/s11423-013-9298-1., 2013.

[100] Walldén, Sari & Mäkinen, Erkki. ”Educational Data Mining and Problem-Based Learning”, Informatics in Education. vol 13, pp. 141 – 156, 2014.

[101] Czibula, Gabriela & Ciubotariu, George & Maier, Mariana & Lisei, Hannelore. ”IntelliDaM: A Machine Learning-Based Framework for Enhancing the Performance of Decision-Making Processes. A Case Study for Educational Data Mining”, IEEE Access. 10. 1-1. 10.1109/ACCESS.2022.3195531, 2022.

[102] Sarra, A., Fontanella, L., & Zio, S., ”Identifying Students at Risk of Academic Failure Within the Educational Data Mining Framework”, Social Indicators Research, vol 146(2), pp. 41-60, 2018.

[103] Križanić, Snježana. ”Educational data mining using cluster analysis and decision tree technique: A case study”, International Journal of Engineering Business Management. 12. 184797902090867. 10.1177/1847979020908675., 2020.

[104] Tsiakmaki M, Kostopoulos G, Kotsiantis S, Ragos O., ”Implementing AutoML in Educational Data Mining for Prediction Tasks”, Applied Sciences, https://doi.org/10.3390/app10010090, vol 10(1):90, 2020.

[105] Julia C. Arnold, Andreas Mühling & Kerstin Kremer, ”Exploring core ideas of procedural understanding in scientific inquiry using educational data mining”, Research in Science & Technological Education, DOI: 10.1080/02635143.2021.1909552, vol 41:1, pp. 372-392, 2023.

[106] Tsoni, Rozita & Garani, Georgia & Verykios, Vassilios. ”Data pipelines for educational data mining in distance education”, Interactive Learning Environments, 10.1080/10494820.2022.2160466., pp. 1-14, 2023.

[107] Toivonen, Tapani & Jormanainen, Ilkka. ”Evolution of Decision Tree Classifiers in Open Ended Educational Data Mining”, TEEM'19: Proceedings of the Seventh International Conference on Technological Ecosystems for Enhancing Multiculturality, 10.1145/3362789.3362880., pp. 290-296, 2019.

[108] Alfonso, David & Riesco, Angeles & Pickin, Simon. ”Semi-Automatic Generation of Competency Maps Based on Educational Data Mining”, International Journal of Computational Intelligence Systems. 12. 10.2991/ijcis.d.190627.001., 2019.

[109] Haisan, Angel-Alex & Breşfelean, V. ”A Data Mining Examination on the Romanian Educational System - teachers Viewpoint”, International Journal of Mathematical Models and Methods in Applied Sciences, vol 7, pp. 277-285., 2013.