Online-Journals.org

iJOE International Journal of Online and Biomedical Engineering

iJOE | eISSN: 2626-8493 | Vol. 19 No. 12 (2023) | OPEN ACCESS

https://doi.org/10.3991/ijoe.v19i12.40079

PAPER

Contributions of Data Mining to University Education, in the Context of the Covid-19 Pandemic: A Systematic Review of the Literature

Martín Díaz-Choque[1](✉), Omar Chamorro-Atalaya[1], Orlando Adrian Ortega-Galicio[1], José Antonio Arévalo-Tuesta[2], Elvira Cáceres-Cayllahua[3], Ronald Fernando Dávila-Laguna[4], Irma Esperanza Aybar-Bellido[5], Yina Betty Siguas-Jerónimo[5]

[1]Universidad Nacional Tecnológica de Lima Sur, Lima, Peru

[2]Universidad Nacional Federico Villarreal, Lima, Peru

[3]Universidad Nacional Agraria La Molina, Lima, Peru

[4]Universidad César Vallejo, Lima, Peru

[5]Universidad Privada San Juan Bautista, Ica, Peru

2014200248@untels.edu.pe

ABSTRACT

During the context of COVID-19, educational processes migrated to a strictly virtual scenario, so the quantity of information grew in such a way that techniques such as data mining or machine learning contributed to generating knowledge for decision-making. In this sense, it is relevant to define the state of the art of the contributions of data mining in the university environment, and from there, to see in perspective how these could be applied in scenarios of return to the face-to-face. In this sense, a systematic review of the literature is carried out, based on scientific evidence extracted from the Taylor & Francis, ERIC and Scopus databases. A qualitative content analysis approach and the PRISMA (Preferred Reporting Items for Systematic reviews and Meta-Analyses) statement were used to extract the findings published in scientific articles. The results were that educational data mining was applied to a greater extent in the field of "teaching", and it was focused on the search for patterns and predictive models to improve student performance, reduce student dropout, improve the student's quality of life, and teacher performance. In addition, as a resource for data extraction, university learning management systems (LMS) were used to a greater extent. It is concluded that tools such as data mining should be implemented as academic management policies, achieving a prospective on indicators linked to the improvement of student learning and performance.

KEYWORDS

data mining, higher education, COVID-19, systematic review

1 INTRODUCTION

In recent years, a large amount of data related to the different processes inherent to organizations have been accumulating [1]. In addition, it can be established that we currently live in a digitized world, which leads to the generation of massive data, so it does not necessarily imply an increase in knowledge [2–4]. The discovery of knowledge

Díaz-Choque, M., Chamorro-Atalaya, O., Ortega-Galicio, O.A., Arévalo-Tuesta, J.A., Cáceres-Cayllahua, E., Dávila-Laguna, R.F., Aybar-Bellido, I.E., Siguas-Jerónimo, Y.B. (2023). Contributions of Data Mining to University Education, in the Context of the Covid-19 Pandemic: A Systematic Review of the Literature. *International Journal of Online and Biomedical Engineering (iJOE)*, 19(12), pp. 16–33. https://doi.org/10.3991/ijoe.v19i12.40079

Article submitted 2023-04-01. Resubmitted 2023-06-01. Final acceptance 2023-06-01. Final version published as submitted by the authors.

based on stored data is an automatic process in which discovery and analysis converge [5, 6]. The discovery of related knowledge in databases continues to spread to almost all fields where large amounts of data are stored and processed [7]. Data mining is the process of exploring large amounts of data based on consistent procedures such as association rules or time sequences [8–10], with the purpose of establishing a systematic relationship between variables by detecting new subsets of data [11]. Thus, data mining concentrates its main purpose on extracting information about the relationship between variables from large amounts of data of different nature [12, 13].

As the Internet opened up new ways of communication, the education sector has embraced such technology and developed educational systems that collect data and are updated in real time [14]. Currently, there are tools that allow the extraction, storage, and analysis of information focused on educational processes [15, 16]. The application of data mining techniques and tools in various educational contexts is known as educational data mining [17–20] and aims to identify hidden patterns to use them in decision making [21–23]. Thus, educational data mining focuses on predicting student behavior in order to establish recommendations regarding the teaching-learning process, performance, activity management, among others [24–26]. For that purpose, educational data mining requires a set of activities aimed at preparing the input data, for which quantitative techniques are required [27, 28]. This is why it is important to apply specialized technologies such as educational data mining that allow these activities to be carried out through the integration of various disciplines such as statistics, artificial intelligence, machine learning, among others [29]. In this regard, Oviedo and Jiménez (2019) [30] point out that supervised techniques allow future data prediction tasks to be carried out, including classification and regression through the study of historical data [31], while unsupervised techniques allow describing current data, within which is found the grouping, association and selection of factors.

However, universities have the responsibility to understand the needs of their students and monitor their performance and satisfaction on an ongoing basis, in order to make continuous improvements in their processes in a timely manner [32–34]. Therefore, it is necessary that the members of the university board know the organizational dynamics and the behavior assumed by its members who intervene in it [35]. In this regard, in Pérez-Gutiérrez [36], it is established that it is relevant for higher education institutions to generate predictive or classification models that contribute to improving educational quality. Another relevant aspect of data mining in educational settings is that they are not only linked to the individual behavior of students, but rather seek to establish relationships between a group of students with demographic, economic, and administrative aspects [37]. Due to the importance of the value of knowledge that is generated with educational data mining, universities should focus their efforts on developing and applying this type of technology, since they turn out to be a fundamental pillar to achieve economic and social development, whose research in this field becomes relevant and important for society [38]. There are still universities that concentrate on a large amount of information; however, it is not being processed, leading to inadequate academic planning, a waste of institutional infrastructure, and a lack of knowledge about the real demand for the careers being offered [39].

Consequently, this article aims to carry out a systematic review on the contribution of data mining in the different areas of higher education, for which a search strategy based on Boolean integrators published in databases such as Taylor & Francis, ERIC, and Scopus were used. In order to achieve the objectives of the comprehensive review of the literature, the PRISMA statement (Preferred Reporting Items for Systematic reviews and Meta-Analyses) was used. Therefore, this article contains a section on similar studies, which allowed us to define the state of the art on systematic review studies about data mining in education. It also describes the methodological aspects considered in the systematic review, with which it was possible to extract the articles to be analyzed and synthesized, reducing any possibility of bias. The results section provides details about the findings based on the research questions (RQs). Furthermore, these results are discussed with other similar studies. Finally, the conclusions and limitations of the study are specified.

2 SIMILAR STUDIES

Molina and Utreras [40], carried out a systematic review of data mining applied to the classification of academic performance for a period from 2015 to 2018, in which, under a quantitative approach, they sought to identify which data mining techniques and methods are the best and most used. Likewise, in Panizzi [41] a systematic review of data mining applied to higher education, whose information search period was established between the years 2008 and 2018 was carried out, focusing on applied methodologies, used programming languages and what it intends to solve with data mining in universities. In the same line of study, Albreiji et al. [42] conducted a systematic review on data mining aimed at predicting student performance through machine learning techniques, focusing on aspects such as what predictive techniques are used in predicting academic performance, what solutions are reached and what is the production level in this field.

Other studies, as the one detailed by Liñan and Peréz [43], carry out a systematic review of educational data mining and learning analysis, focusing on the exhaustive review of the available literature, similar and different aspects, and evolution over time, in the context of teaching MOOC (Massive Open Online Course) courses. Thus, also focusing on a specific scenario in university education such as self-regulated learning [44], they carry out a review of the data mining system applied to that particular scenario, for which its study focuses on identifying what techniques and algorithms are the most used and which are the most optimal in the existing literature. Another study closely related to data mining is the one developed by Nunn et al. [45], in which they carried out a systematic review of the methods used for learning analytics in university institutions, focusing on the scientific findings from 2000 to 2016, and whose objective was to investigate the methods, benefits and challenges used in higher education.

3 METHODOLOGY

3.1 Type of study

The systematic review developed in this article is based on a qualitative study, in which it identifies, selects, evaluates and synthesizes results and findings of scientific articles regarding the contribution of data mining in the different areas of higher education, for which it was established to formulate three research questions in order to carry out and achieve the objectives of the review [46], which are shown below:

- RQ1: In which area of higher education was data mining applied in the context of COVID-19?

- RQ2: What data extraction sources were used in higher education for the application of data mining in the Context of COVID-19?
- RQ3: What is the contribution of data mining in the different areas of higher education?

In addition, to structure the systematic review procedure, the PRISMA statement was used in this article, which provides guidance on how to report the use of automation tools at various steps of the review process, such as searching, reference selection, data collection, evaluation and synthesis of the study, reducing the bias in the search for results that do not fit the subject under study [47].

3.2 Search strategy

Regarding the search strategy, descriptors related to the variables under study and their dimensions were used, which are interrelated with the research questions, written in Spanish and English. Table 1 shows the search strategy required by each database, in which results were obtained by defining the search equation through the combination of Boolean integrators.

Table 1. Search equation through Boolean indicators

Database	Search Equation
Taylor & Francis	[[All: "minería de datos educativos"] OR [All: "educational data mining"]] AND [[All: "educación superior"] OR [All: "higher education"] OR [All: "universidad"] OR [All: "university"] OR [All: "ámbito universitario"] OR [All: "university environment"]]
SCOPUS	((TITLE (minería AND de AND datos AND educativos) OR TITLE (educational AND data AND mining))) AND ((TITLE (educación AND superior) OR TITLE (higher AND education) OR TITLE (universidad) OR TITLE (university) OR TITLE (ámbito AND universitario) OR TITLE-ABS-KEY (university AND environment)))
ERIC	(("minería de datos educativos" OR "Educational Data Mining") AND ("Educación superior" OR "Higher education" OR "Universidad" OR "University" OR "Ámbito universitario" OR "University environment"))

3.3 Inclusion and exclusion criteria

The inclusion (IC) and exclusion (EC) criteria are shown in Table 2 with which the eligibility of scientific articles is sought, avoiding the removal of biases at the time of identification and selection of bibliographic references. The purpose of defining the inclusion and exclusion criteria is to restrict the search to scientific articles that guarantee the answer to the research questions, taking into account the relevant scenario and context for the systematic review [48].

Table 2. Inclusion and exclusion criteria

Inclusion Criteria	Exclusion Criteria
IC1: Graded and peer-reviewed articles	EC1: Articles not submitted for peer-review
IC2: Full-text articles available	EC2: Articles that only allow access to their abstract.
IC3: References defined only as scientific	EC3: Congress, Books and Records.
IC4: Articles published from 2020 to 2022	EC4: Articles published before 2020

3.4 Data extraction

Figure 1 shows the flowchart taken as a reference from the PRISMA statement, referenced in Aljohani and Chandran [49], which describes the process of identification, projection, eligibility and inclusion of the research found in the Taylor & Francis, Eric, and Scopus databases. The process begins with the identification of the articles found through the search equation described in the previous section, in which 654 scientific articles were found on 1st January, 2023. Duplicate articles were then excluded from the databases with which the identified articles were found. It was reduced to 528. Subsequently, those articles whose abstracts were not related to the research questions were filtered, reducing them to 198 articles that were projected to be eligible. In the third stage, the inclusion and exclusion criteria were applied, reducing them to 35 articles. Finally, a final review of the total content of each of the articles was carried out, reducing the references included in the systematic review to 16 scientific articles.

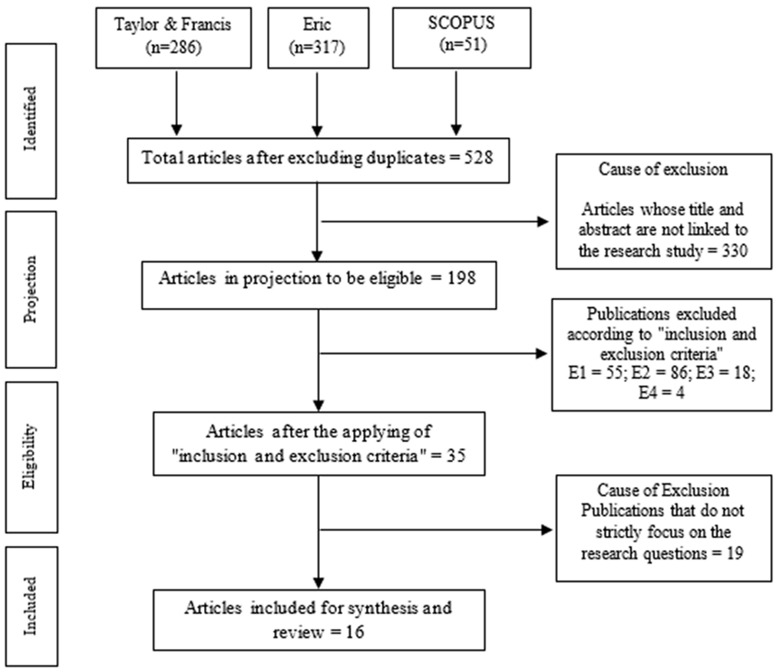

Fig. 1. Flow diagram obtained from the application of the PRISMA sentence

3.5 Quality assessment of the articles included in the systematic review

In order to reduce the risk of bias in the selection of scientific articles considered for the content analysis and synthesis phase, we proceeded to use the instrument validated by [50], in which it proposes four criteria to assess quality (CAQ) of the articles included in the systematic review. After being adapted to the subject of study, they are expressed as follows: the article provides details of the contributions of educational data mining (CAQ1), if the article presents methodological coherence (CAQ2), if the article includes a clear argument (CAQ3), and if the article contributes to the field of study (CAQ4). In Table 3, the quality assessment results expressed as a percentage are shown. The scores for each criterion can take the values of 1, 3, and 5, representing a low, regular, and high rating, respectively. The minimum percentage for an article to be considered of low quality is 50%.

Table 3. Inclusion and exclusion criteria

Reference	CAQ1	CAQ2	CAQ3	CAQ4	Total
[24]	3	3	5	3	70%
[36]	5	5	5	3	90%
[44]	3	3	5	5	80%
[51]	5	3	5	3	80%
[52]	5	3	5	5	90%
[53]	3	3	5	5	80%
[54]	5	5	5	3	90%
[55]	3	5	3	5	80%
[56]	5	5	3	5	90%
[57]	5	3	5	5	90%
[58]	3	3	3	5	70%
[59]	5	3	5	5	80%
[60]	5	3	3	5	80%
[61]	5	5	3	3	80%
[62]	3	5	5	3	80%
[63]	5	3	5	5	90%

4 RESULTS

Having applied the PRISMA methodology and proceeded to identify the scientific publications included in the systematic review, 16 scientific articles were found. Figure 2 shows the distribution of the scientific articles included in the systematic review process during COVID-19 in years 2020 to 2022, which are linked to the research questions, and were obtained from the Scopus, ERIC, and Taylor & Francisco databases.

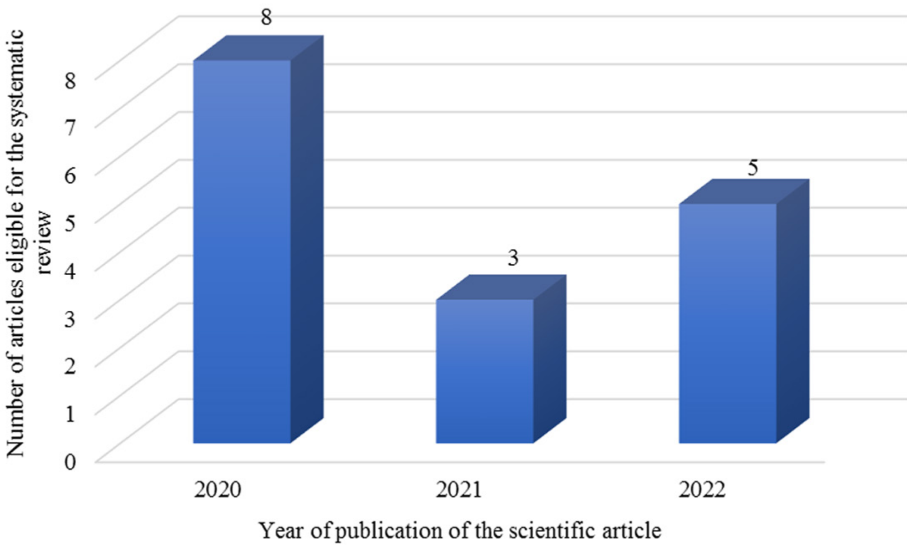

Fig. 2. Distribution of articles included for the systematic review by year of publication

4.1 RQ1: In which areas of higher education was data mining applied in the context of COVID-19?

Regarding the identification of the data mining application fields in higher education, a review of the scientific findings based on the "missions" developed by a university was carried out; in this regard in [64–66], it is established that the prospective conditions of any society, as well as the demands of its development at all levels, require dynamic relationships between universities and their defined missions as "teaching", "research", and "social projection".

Regarding the first research question, the 16 articles reviewed are related to the field of university "teaching". This field is developed in two modalities, which are virtual education and face-to-face education. Regarding virtual education, the academic performance has as a variable a value of 37.5% of the reviewed articles; in face-to-face education, it has two types of variables, the academic performance with a value of 37.5% of the articles reviewed and the student dropout with a value of 25% of the articles reviewed. Figure 3 shows the percentage distribution of each study variable with respect to the two teaching modalities.

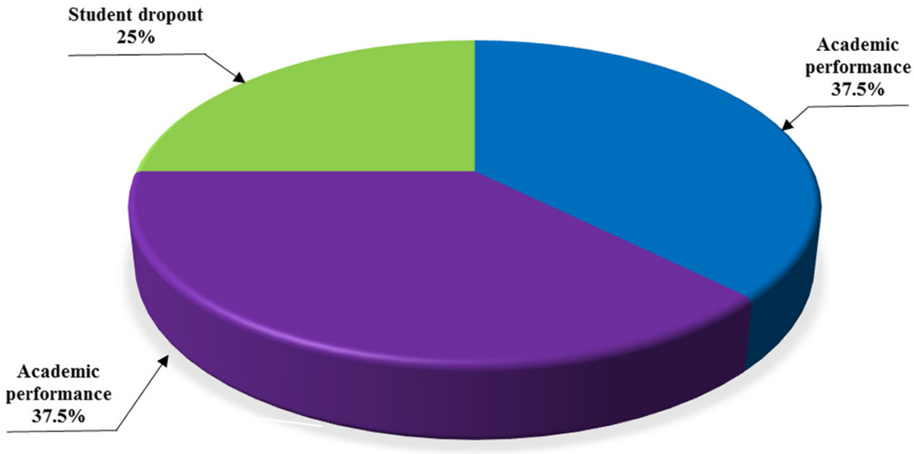

Fig. 3. Percentage distribution of articles reviewed by study variables regarding the teaching modality

By virtue of the above, Table 4 provides details of the scope of application that is related to "teaching"; as well as the two teaching modalities as "virtual education" and "face-to-face education"; and finally, the variables under study that are related to "Academic Performance" and "Student Dropout"; based on systematic review research articles.

Table 4. Findings on teaching modalities and study variables

Ambit of Application	Teaching Modality	Variable Under Study	Reference
Teaching	Virtual teaching	Academic performance	[51–56]
	Face-to-face teaching	Academic performance	[44], [57–61]
		Student desertion	[24], [36], [62], [63]

Based on the findings of the systematic review, a crosstab analysis was performed between the teaching modality and the variable under study, with respect to the teaching field. In reference to Table 5, it is shown that, from the articles reviewed to a greater extent, 37.5% apply data mining in face-to-face teaching, taking the academic performance as the study variable. Likewise, it is possible to identify that 37.5% apply data mining in virtual teaching modality, taking the academic performance as a variable. On the other hand, no research was found on the application of data mining in virtual teaching modality focused on the student dropout variable.

Table 5. Result of the cross-tab analysis between the teaching modality and the variables under study

Teaching Modality	Variable Under Study		Total
	Academic Performance	Student Dropout	
Virtual teaching	37.5%	0%	37.5%
Face-to-face teaching	37.5%	25%	62.5%
Total	75%	25%	100.0%

4.2 RQ2: What data extraction sources were used in higher education for the application of data mining in the Context of Covid-19?

Regarding the second research question, three sources of data extraction were found. One of them has a value of 68.75% of the reviewed articles in reference to the "University Management System registration". On the other hand, 18.75% of the articles viewed refer to "Registration in Google forms", and finally, 12.5% of the articles viewed refer to "Data accumulated in learning platforms". Figure 4 shows the percentage distribution of data extraction sources in higher education from the reviewed articles.

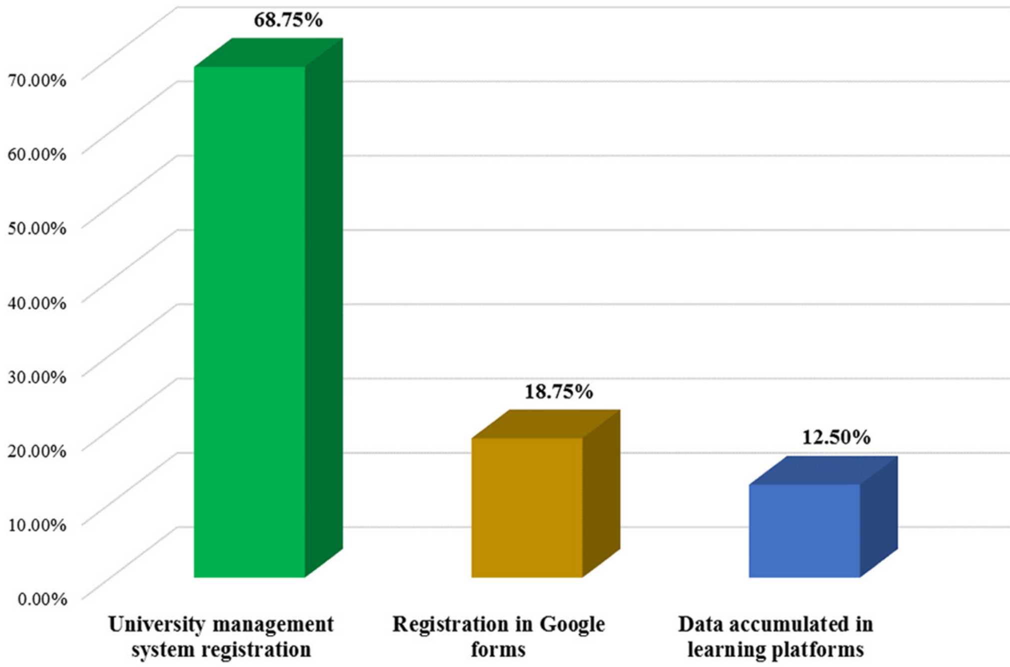

Fig. 4. Percentage distribution of reviewed articles from data extraction sources in university education

On the other hand, a data analysis was obtained from the 16 scientific articles, where 25% of the research articles have a data analysis in "Student Demographic, Academic Performance Data", and the same percentage in "Student academic data"; 37.5% of the research articles have a data analysis in "Data about the student, Their Performance, Self-assessment"; 6.25% of the research articles have a data analysis in "Teacher data regarding the use of Moodle platform tools", and the same percentage in "Data of individuals, Use of tools for students". Figure 5 shows the percentage distribution of the analysis data of the reviewed articles.

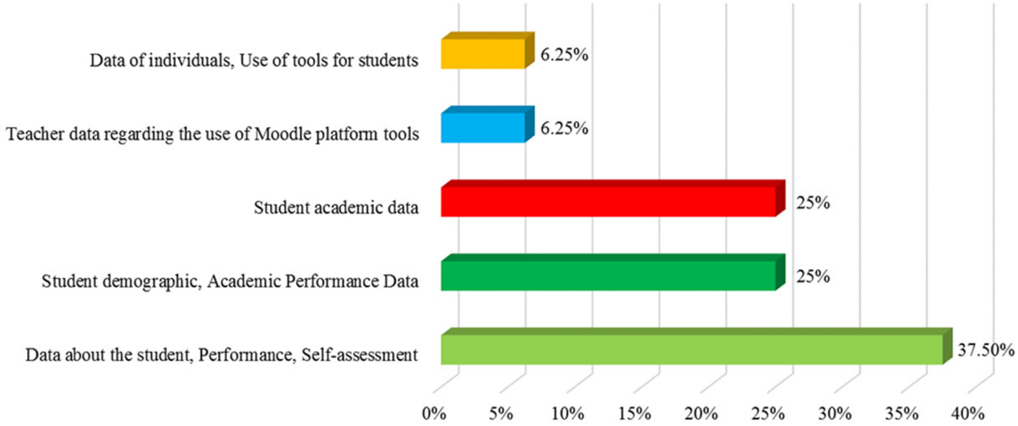

Fig. 5. Percentage distribution of the articles reviewed regarding the data extraction in higher education

By virtue of the above, Table 6 provides details about the three sources of data extraction that are related to the "University management system Registration", the "Registration in Google forms", and the "Data accumulated in learning platforms". Finally, the analysis data refers to the "Student Demographic Data, Academic Performance Data", "Data about the student, Performance, self-assessment", "Student academic data", "Teacher data regarding the use of Moodle platform tools", "Data of individuals, use of tools for students", based on the reviewed articles.

Table 6. Findings on the sources of data extraction and analysis used in higher education

Data Extraction Source	Data Under Analysis	Reference
University Management System Registration	• Student demographic • Academic Performance Data	[36], [44], [57], [63]
	• Data about the student • Student performance • Student Self-assessment	[51], [55], [56]
	• Student academic data	[59], [60], [62]
	• Teacher data regarding the use of Moodle platform tools	[53]
Registration in Google forms	• Data about the student • Student performance • Student self-assessment	[24], [54], [61]
Data accumulated in learning platforms	• Student academic data	[52]
	• Data of individuals • Use of tools for students	[58]

Based on the findings of the systematic review, a cross-tabulation analysis was carried out between data extraction and data analysis with respect to the data extraction sources. With reference to Table 7, it is shown that from the articles reviewed, 25% applied data mining to extract the demographic data of the students and academic performance data, taking the University Management System registration as analysis data. Likewise, 6.3% applied data mining to extract the teacher data regarding the use of the Moodle platform tools, taking the University management system registration as analysis data. On the other hand, no research was shown on the application of data mining in the extraction of student demographic and academic performance data, analyzing data from the registration in Google forms and data accumulated in learning platforms.

Table 7. Result of the cross-tabulation analysis between data extraction and data analysis

Data Extraction Source	Data Under Analysis			Total
	University Management System Registration	Registration in Google Forms	Data Accumulated in Learning Platforms	
• Student demographic • Academic performance data	25.0%	0.0%	0%	25.0%
• Data about the student • Performance • Self-assessment	18.8%	18.8%	0%	37.5%
• Student academic data	18.8%	0.0%	6.3%	25.0%
• Teacher data regarding the use of Moodle platform tools	6.3%	0.0%	0%	6.3%
• Data of individuals • Use of tools for students	0.0%	0.0%	6.3%	6.3%
Total	68.8%	18.8%	12.5%	100.0%

4.3 RQ3: What is the contribution of data mining in the different areas of higher education?

Regarding this third research question, in order to obtain greater detail from the exploration of the systematic review findings and results, Table 8 shows the contributions achieved by each article reviewed, which consist of those that are mostly linked to pension alert systems focused on student performance.

Table 8. Findings on the contribution of data mining in higher education

Reference	Variable Under Study	Contribution to Research
[51]	Academic performance	Through data mining, they developed an interim alert system based on student performance in a mixed online environment through the prediction of a group of high-risk students.
[52]		Through data mining, they accurately predicted the student performance, for that purpose, they used the learnings in courses and the classifications of the different students.
[53]		Through data mining, they identified behavior patterns of teachers regarding the use of the resources and the Moodle platform tools and the performance of students during the semester.
[54]		Through data mining, they predicted the student performance in the online physics course, so they proposed to replicate it for future online courses.
[55]		Using data mining, they predicted the student performance in courses of each semester. With this prediction, the importance of the prerequisite courses is stressed to the students.
[56]		They used educational data mining methods that predicted the student academic performance and dropout rate, with the aim of assessing students' mastery of specific skills.
[57]		Through data mining, they found that the age of admission has a negative effect on the completion of studies, while the performance of the first semester contributes positively to graduation.
[58]		Through data mining, they determined that the use of Moodle learning management system (LMS) resources and the student performance, are significantly related, so they are useful for strategic academic planning purposes with LMS data in the university.
[59]		Through data mining you may monitor the use of learning management system (LMS) resources by students and the university staff, thus predicting the academic performance of students.
[60]		Through data mining, they discovered patterns in the student learning, thus predicting the student academic performance. In addition, the student performance for a given discipline can be improved by including additional characteristics in the learning process that represent their previous courses results.
[61]		Data mining is a promising research area in education, extracting useful information from students without any prior hypotheses.
[44]		Using data mining, they predicted student demographics, where the lowest-risk students are those who live on campus, while the two highest-risk groups are commuters. Similarly, the performance of the students was assessed.
[24]	Student desertion	Through data mining, they found 27 factors involved in student dropout, but three of them are the most relevant factors, which are the lack of guidance, the lack of an adequate student environment, and the lack of an adequate follow-up. In addition, seven patterns such as the student environment, and an insufficient financial support, among others. Therefore, student dropout depends on multiple factors.
[36]		Data mining in the education field is increasing, due to the prediction of the student needs. It also predicts student dropout rates. Student dropouts are due to a composite data set, such as student demographics and transcript recording at different points in their careers. In addition, the performance of the engineering courses is related to the physics and mathematics courses.
[62]	Student dropout	Through data mining, they determined that the most important attributes in student dropout are absences, last semester completed, percentage of failed subjects, credits completed, and study program.
[63]		Through data mining, they predicted those factors that affect the dropout rate of engineering students. One of the dropout factors is related to the credits in which a student enrolls and the relationship between the credits passed and those in which the student has enrolled.

5 DISCUSSION

In relation to the fields and in what the contribution of data mining in higher education consists of, it was identified from the review of the literature that these are found in the teaching field, mainly focused on virtual and face-to-face modes. In this regard, Panizzi [41] concludes that from the systematic review it was possible to determine that data mining in higher education solves problems related to teaching and the student by focusing on the student academic performance, the dropout and the strategic educational quality. Although it is true that there are coincidences in the results, there is one aspect in which they do not indicate that it is about the application of data mining based on the efficient use of the resources of the learning management system. This can be understood since the aforementioned research addresses its study until 2019. However, due to Covid-19, and whenever there is greater use of virtual tools for online teaching and learning, data mining has made a great contribution to the solution of problems related to the optimization of the use of the LMS resources. In another similar research, Molina and Utreras [40] mention that the issues addressed by educational data mining from the perspective of teaching are mainly focused on solving problems related to academic performance and dropout, through predictive models, as well as web-based didactic media. However, it has not been indicated that issues such as monitoring the teacher performance are addressed, which is a very relevant aspect in the university system, even more so when it is linked to the student performance. Another study that is also important to take into account is the one developed by Albreiji et al. [42]. Considering that its study timeframe goes until 2021, it indicates that the contributions of data mining are focused on predicting the risk and student dropout. This is possibly quite limited or provides a reduced appreciation of its production as part of its systematic review process due to the fact that its research questions narrow the scope of the search, as the data extraction contributes to many other aspects of higher education.

Likewise, Liñan and Pérez [43] point out that in the future, educational data mining and learning analytics will be applied on a massive scale in online learning environments, and mainly in MOOC courses. Today, and according to the results found in this systematic review, I was able to demonstrate that, as indicated by [51–53], [55, 56], data mining is applied in online teaching environments, establishing predictive patterns and models related to its application and performance; of course, as a relevant factor to achieve better performance in the student and teacher.

In this regard, Panizzi [41], as a result of his systematic review, agrees with the results of this research, since it establishes that the most used algorithms in educational data mining are several, highlighting the Random Forest, Linear Regression, and SVM algorithm. However, Molina and Utreras [40] point out that in a review from 2013 to 2018, about the most used algorithms in data mining with respect to education, it can be observed that the decision trees are the most used (43.64%) for these cases because it has several solutions for the same problem. It also allows us to analyze the possible consequences of making a decision. After this method, the most used are the Bayesian Networks. However, this research is only limited to data mining applications in areas related to academic performance, and not to teacher performance, or patterns of use of learning management systems, so there may be understandable differences about the different types of algorithms used in data mining in higher education.

6 CONCLUSION

Once the systematic review has been carried out, it is concluded that the fields of application of educational data mining in higher education are focused solely on the field of "teaching", both in the face-to-face and virtual modality, and the purpose of the applications focuses on finding patterns and predictive models of the student performance, dropout and abandonment, teacher performance, and the use of the learning management system (LMS) resources. By virtue of the foregoing, the data that is extracted for this purpose is linked to data on academic performance, demographic conditions, LMS data, and text data of the student and teacher. Based on this state of the art obtained from the systematic review of educational data mining in the field of higher education, it is important that future studies focus on exploring whether universities count within their academic management privacy policies and regulations with the implementation of tools linked to data mining, from its different models and algorithms, since what is found in the systematic review are cases in which no one mentions that it is based on an academic or prospective management policy of the University.

This systematic review was strictly focused on the analysis of the scientific evidence of data mining in the university academic field and during the pandemic; however, it is specified that data mining also requires computer tools for the extraction, storage, processing and visualization of results, so this study did not extract information on these aspects. Therefore, future studies can take these findings as a starting point and give an approximation on what computer tools are necessary to carry out the implementation of data mining in the educational environment.

7 ACKNOWLEDGMENT

Gratitude and appreciation to the National Technological University of South Lima (Universidad Nacional Tecnológica de Lima Sur) for supporting the development of this research.

8 REFERENCES

[1] M. Diógenes "Integration of data mining and blended learning to address the university educational process", M. S. thesis, National University of the South, Bahía Blanca, Argentina, 2021. https://repositoriodigital.uns.edu.ar/handle/123456789/6172

[2] D. Chamorro-Sangoquiza, A. Vargas-Muñoz, and A. Umaquinga-Criollo, "Comparative study of data mining techniques to reveal patterns of academic performance in secondary education", *Iberian Journal of Information Systems*, vol. 32, no. 1, pp. 455–468, 2020. http://repositorio.espe.edu.ec/xmlui/handle/21000/21822

[3] B. Paaben, B. Jeffris, K. Yacef, J. McBroom, and I. Koprinska, "Mapping python programs to vectors using recursive neural encodings", *Journal of Educational Data Mining*, vol. 13, no. 3, pp. 1–35, 2021. https://doi.org/10.5281/zenodo.5634224

[4] V. S. Gómez, "Learning Analytics: Methodological proposal and case study in non-university education", MS. Thesis, Faculty of Information Technology, University of Valladolid, Valladolid, Spain, 2018. https://uvadoc.uva.es/bitstream/handle/10324/33352/TFM-G959.pdf?sequence=1

[5] S. Timarán, "Application in data mining of the detection of academic performance patterns in saber pro tests", Editorial University of Nariño, 2021.

[6] P. E. C. Chacón, A. Zapata, V. Menéndez, and P. Canto, "Analysis of the performance of university professors in the use of MOODLE through data mining techniques: Proposals of training needs", *Distance Education Journal*, no. 58, pp. 1–41, 2018. https://revistas.um.es/red/article/view/351411

[7] J. Campo-Ávila, G. Ramos-Jiménez, R. Morales-Bueno, and M. Baena-García, "Educational data mining for personalized prediction of academic performance in 2017", In Proc. International Convention 2017 Central University Marta Abreu de Las Villas, 2017, pp. 1–10. https://riuma.uma.es/xmlui/bitstream/handle/10630/15477/cipi2017talf.pdf?sequence=1&isAllowed=y

[8] P. Arevalo-Marin, Y. Cabrera-Piedra, M. Cabrera-Sarango, and J. Caicedo-Rodríguez, "Educational data mining to identify the relationship between intelligence quotient, styles of learning, emotional intelligence and intelligences multiple engineering students", *Iberian Journal of Information Systems and Technologies*, no. 17, pp. 48–63, 2019. https://www.proquest.com/openview/a78cfaa62708fd241076ce89e18cfa6a/1?pq-origsite=gscholar&cbl=1006393

[9] S. Haroa, R. Pazmiño, M. Conde, and F. Peñalvo, "Data mining to uncover trends in classification of degree works", In Proc. Congress of Science and Technology ESPE, 2018, pp. 125–128. https://knowledgesociety.usal.es/sites/default/files/evidencias/silvia-rub%C3%A9n11rev.pdf

[10] S. Aisa, A. Akhriana, A. Ahyuna, A. Irmayana, I. Irmawati, and N. Aini, "An analysis of students' learning interest in programming language based on data mining with fuzzy c-means method", *International Journal of Online and Biomedical Engineering*, vol. 17, no. 9, pp. 4–20, 2021. https://doi.org/10.3991/ijoe.v17i09.24595

[11] F. B. Alves, and D. A. Lima, "Use of classification for the analysis and mining of data in the teaching-learning tool Google classroom", *New Ideas in Educational Computing*, vol. 14, pp. 589–594, 2018. https://doi.org/10.6084/m9.figshare.14994423

[12] F. Martínez-Abad, and J. P. Hernández-Ramos, "Data mining techniques with free software for factor detection associated with performance", *Journal of Studies and Experiences in Education*, vol. 2, no. 2, pp. 134–146, 2018. https://www.redalyc.org/journal/2431/243156768012/html/

[13] J. Pincay-Ponce, J. Herrera-Tapia, J. Terranova-Ruiz, and M. Cruz-Felipe, "Educational data mining: Incidence of factors socioeconomic factors in school achievement", *Iberian Journal of Information Systems and Technologies*, no. 49, pp. 654–667, 2022. https://www.proquest.com/openview/30ab36bec36b2b523c7352c3824731ea/1.pdf?pqorigsite=gscholar&cbl=1006393

[14] C. Vergara, A. Oviedo, C. Carmona, G. Vélez, and I. Amón, "Learning styles and data mining: A study preliminary in the university context", *Engineering and Innovation Journal*, vol. 6, no. 2, pp. 32–37, 2018. https://doi.org/10.21897/23460466.1742

[15] J. Carrillo, "Big data-analytics of learning and mining of data applied in the University", *Journal of Production, Science and Research,* vol. 2, no. 8, pp. 39–54, 2018. https://doi.org/10.29018/issn.2588-1000vol2iss8.2018pp29-54

[16] K. Kyritsi, V. Zorkadis, E. Stavropoulos, and V. Verykios, "The pursuit of patterns in educational data mining as a threat to student privacy", *Journal of Interactive Media in Education*, vol. 1, no. 2, pp. 1–10, 2019. https://doi.org/10.5334/jime.502

[17] R. Timarán-Pereira, A. Hidalgo-Troya, and F. Vidal-Alegría, "A look at academic performance in the saber pro tests for engineering students from educational data mining", *Iberian Journal of Information Systems and Technologies*, no. 38, pp. 29–42, 2020. https://dialnet.unirioja.es/servlet/articulo?codigo=8574384

[18] X. Zhou, and X. Wu, "Teaching mode based on educational big data mining and digital twins", *Hindawi Journal*, vol. 2022, pp. 1–13. https://doi.org/10.1155/2022/9071944

[19] S. Khan and S. Alqahtani, "Big data application and its impact on education," *International Journal of Emerging Technologies in Learning*, vol. 15, no. 17, pp. 36–46. https://doi.org/10.3991/ijet.v15i17.14459

[20] I. Jugo, B. Kovačić, and V. Slavuj, "Increasing the adaptivity of an intelligent tutoring system with educational data mining: A system overview," *International Journal of Emerging Technologies in Learning*, vol. 11, no. 3, pp. 67–70. https://doi.org/10.3991/ijet.v11i03.5103

[21] D. Zapata, "Method for the detection of students at risk of desertion, based on a design of metrics and a data mining technique", M. S. thesis, Department of Computation and Decision Sciences, National University of Colombia, Medellín, Colombia, 2021. https://repositorio.unal.edu.co/handle/unal/80615

[22] S. Bianco, S. Martíns, H. Amatriain, and H. Merlino, "Automation proposal for mining projects educational data", In Proc. XXV Argentine Congress of Computer Science (CACIC), 2019, pp. 366–375. http://sedici.unlp.edu.ar/handle/10915/90359

[23] O. Chamorro-Atalaya, G. Morales-Romero, Y. Meza-Chaupis, E. Auqui-Ramos, J. Ramos-Cruz, C. León-Velarde, I. Aybar-Bellido, "Quadratic vector support machine algorithm, applied to prediction of university student satisfaction, "*Indonesian Journal of Electrical Engineering and Computer Science*, vol. 27, no. 1, pp. 139–148, 2022. https://doi.org/10.11591/ijeecs.v27.i1.pp139-148

[24] A. Urbina-Nájera, J. Camino-Hampshrire, and R. C. Barboza, "University dropout: Prevention patterns through the application of educational data mining", *Journal of Educational Research, Assessment and Evaluation*, vol. 26, no. 1, pp. 1–19, 2020. https://doi.org/10.7203/relieve.26.1.16061

[25] O. Castrillón-Gómez, W. Sarache, and S. Ruiz-Herrera, "Prediction of the main variables that lead to the student dropout through mining techniques data", *University Training Journal*, vol. 13, no. 6, pp. 217–228, 2020. https://doi.org/10.4067/S0718-5006 2020000600217

[26] O. Chamorro-Atalaya, O. Ortega-Galicio, G. Morales-Romero, D. Villar-Valenzuela, Y. Meza-Chaupis, C. Leon-Velarde, and L. Quevedo-Sánchez, "Vector support machine algorithm applied to the improvement of satisfaction levels in the acquisition of professional skills," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 26, no. 1, pp. 597–604, 2022. https://doi.org/10.11591/ijeecs.v26.i1.pp597-604

[27] H. Cajo, and B. Geovanny, "Data mining in information systems learning management in education University", *Virtual Campus Journal*, vol. 7, no. 2, pp. 115–128, 2018. https://dialnet.unirioja.es/servlet/articulo?codigo=6681865

[28] K. Jones, A. Rubel, and E. LeClere, "A matter of trust: Higher education institutions as information fiduciaries in an age of educational data mining and learning analytics", *Journal of the Association for Information Science and Technology*, vol. 71, no. 10, pp. 1227–1241, 2019. https://doi.org/10.1002/asi.24327

[29] A. Aquino, G. Molero-Castillo, and R. Rojano, "Towards a new data mining process user-centric", *Educational Tracks Journal*, vol. 36, no. 114, pp. 272–291, 2015. https://pistaseducativas.celaya.tecnm.mx/index.php/pistas/article/view/303/0

[30] A. Oviedo, and J. Jiménez, "Educational data mining: Analysis of the performance of engineering students in the saber-pro tests", *Polytechnic Journal*, vol. 15, no. 29, pp. 128–140, 2019. https://doi.org/10.33571/rpolitec.v15n29a10

[31] V. A. Domínguez., A. E. M. Cervantes, and A. T. A. Bravo, "Prototype of web application for the prediction of the School performance through data mining educational", *Humanities, Technology and Science of the National Polytechnic Institute,* vol. 26, pp. 1–7, 2022. https://www.revistaelectronica-ipn.org/ResourcesFiles/Contenido/27/TECNOLOGIA_27_001055.pdf

[32] E. Ayala, R. López, and V. Menéndez, "Predictive models of academic risk in computing careers with educational data mining", *Distance Education Journal*, vol. 21, no. 66, pp. 1–36, 2021. https://doi.org/10.6018/red.463561

[33] J. Berens, K. Schneider, S. Gortz, S. Oster, and J. Burghoff, "Early detection of students at risk-predicting student dropouts using administrative student data from German universities and machine learning methods", *Journal of Educational Data Mining*, vol. 11, no. 3, pp. 1–41, 2019. https://doi.org/10.5281/zenodo.3594771

[34] O. F. C. Atalaya, T. N. D. Leyva, D. Y. A. Santillan, J. I. C. Bedriñana, D. M. B. Pichilingue, and A. Roman-Gonzalez, "Satisfaction of the graduate for the continuous improvement of educational quality in UNTELS," *Advances in Science, Technology and Engineering Systems*, vol. 4, no. 5, pp. 151–157, 2019. https://doi.org/10.25046/aj040520

[35] A. Oviedo, and G. Almendrales, "Study on learning styles through data mining as support for academic management in educational institutions", *Iberian Journal of Information Systems and Technologies*, no. 29, pp. 1–13, 2018. https://dialnet.unirioja.es/servlet/articulo?codigo=6798991

[36] B. R. Pérez-Gutiérrez, "Comparison of data mining techniques to identify signs of student desertion, based on academic performance", *UIS Engineering Journal*, vol. 19, no. 1, pp. 193–204, 2020. https://doi.org/10.18273/revuin.v19n1-2020018

[37] G. Mancilla-Vela, P. Leal-Gatica, A. Sánchez-Ortiz, and C. Vidal-Silva, "Factors associated with the success of students in the modality of online learning: An analysis in data mining", *University Training Journal*, vol. 13, no. 6, pp. 23–36, 2020. https://www.scielo.cl/scielo.php?pid=S0718-50062020000600023&script=sciar ttext_plus&tlng=es; https://doi.org/10.4067/S0718-50062020000600023

[38] D. Labañino, and L. Soto, "Data mining for knowledge management in Latin American Universities", *EasyChair*, no. 4502, pp. 1–17, 2020. https://easychair.org/publications/preprint/cxQX

[39] M. U. Fassler, A. Cisneros, G. Samaniego, and E. Barba, "Educational data mining for decision-making in the leveling and admission unit at the Ecuadorian university level", *Summits Journal*, vol. 4, no. 2, pp. 55–67, 2019. https://dialnet.unirioja.es/servlet/articulo?codigo=6836545

[40] S. K. C. Molina, and V. K. T. Utretas, "Data mining applied to the classification of academic performance", MS. Thesis, Professional School of Informatics, Higher Polytechnic Agricultural School of Manabí, Calceta, Ecuador, 2018. https://repositorio.espam.edu.ec/xmlui/handle/42000/862

[41] M. Panizzi, "Establishment of the state of the arte on educational data mining in the higher level: A systematic mapping study", *Journal of Scientific Research of the University of Moron*, vol. 2, no. 4, pp. 51–60, 2019. https://repositorio.unimoron.edu.ar/bitstream/10.34073/164/1/Establecimiento%20del%20estado%20del%20arte%20sobre%20la%20Miner%C3%ADa%20de%20Datos%20Educacional%20en%20el%20Nivel%20Superior%20-%20Un%20Estudio%20de%20Mapeo%20Sistem%C3%A1tico.pdf

[42] B. Albreiji, N. Zaki, and H. Alashwal, "A systematic literature review of student' performance prediction using machine learning techniques", *Education Sciences Journal*, vol. 11, no. 9, pp. 1–27, 2021. https://doi.org/10.3390/educsci11090552

[43] L. C. Liñan, and A. J. Peréz, "Educational data mining and learning analytics: Differences, similarities, and time evolution", *Universities and Knowledge Society Journal*, vol. 12, no. 3, pp. 98–112, 2015. https://doi.org/10.7238/rusc.v12i3.2515

[44] E. Araka, R. Oboko, E. Maina, and R. Gitonga, "Using educational data mining techniques to identify profiles in self-regulated learning: An empirical evaluation", *International Review of Research in Open and Distributed Learning*, vol. 23, no. 1, pp. 131–162, 2022. https://doi.org/10.19173/irrodl.v22i4.5401

[45] S. Nunn, J. T. Avella, T. Kanai, and M. Kebritchi, "Learning analytics methods, benefits, and challenges in higher education: A systematic literature review", *Online learning Journal*, vol. 20, no. 2, pp. 13–29, 2016. https://olj.onlinelearningconsortium.org/index.php/olj/article/view/790; https://doi.org/10.24059/olj.v20i2.790

[46] O. Moscoso-Zea, P. Saa, and S. Luján-Mora, "Evaluation of algorithms to predict graduation rate in higher education institutions by applying educational data mining", *Journal of Engineering Education*, vol. 24, no. 1, pp. 4–13, 2019. https://doi.org/10.1080/22054952.2019.1601063
[47] J. Pardal-Refoyo, and B. Pardal-Peláez, "Annotations to structure a systematic review", *Annotations to Structure a Systematic Review*, vol. 11, no. 2, pp. 155–160, 2020. https://scielo.isciii.es/pdf/orl/v11n2/2444-7986-orl-11-02-155.pdf
[48] L. Subramainan, and M. Mahmoud, "A systematic review on students' engagement in classroom: Indicators, challenges and computational", *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 1, pp. 105–115, 2020. https://doi.org/10.14569/IJACSA.2020.0110113
[49] N. Aljohani, and D. Chandran, "The adoption of mobile health applications by patients in developing countries: A systematic review", *International Journal of Advanced Computer Science and Applications*, vol. 12, no. 4, pp. 16–21, 2021. https://doi.org/10.14569/IJACSA.2021.0120403
[50] C. L. Ávila, and S. B. Gómez, "Gamification in higher education: A systematic review," *Engineering Journal University of Medellín*, vol. 16, no. 31, pp. 97–124, 2017. https://doi.org/10.22395/rium.v16n31a5
[51] H. Hui-Chun, L. I-Fan, L. Che-Tien, and S. Yu-S, "Applying educational data mining to explore students' learning patterns in the flipped learning approach for coding education", *Journal Symmetry*, vol. 12, no. 2, pp. 1–14, 2020. https://doi.org/10.3390/sym12020213
[52] L. F. Daghestani, L. F. Ibrahim, R. S. Al-Towirgi, and H. A. Salman, "Adapting gamified learning systems using educational data mining techniques", *Computer Applications in Engineering Education*, vol. 28, no. 3, pp. 1–22, 2020. https://doi.org/10.1002/cae.22227
[53] J. Calderon-Valenzuela, K. Payihuanca-Mamani, and N. Bedregal-Alpaca, "Educational data mining to identify the patterns of use made by the university professors of the moodle platform", *International Journal of Advanced Computer Science and Applications*, vol. 13, no. 1, pp. 321–328, 2022. https://doi.org/10.14569/IJACSA.2022.0130140
[54] M. Taub, A. M. Banzon, T. Zhang, and Z. Chen, "Tracking changes in students' online self-regulated learning behaviors and achievement goals using trace clustering and process mining", *Frontiers in Psychology*, vol. 13, pp. 1–22, 2022. https://doi.org/10.3389/fpsyg.2022.813514
[55] Y. Zhang, Y. Yun, R. An, J. Cui, H. Dai, and X. Shang, "Educational data mining techniques for student performance prediction: Method review and comparison analysis", *Frontiers in Psychology*, vol. 12, pp. 1–19, 2021. https://doi.org/10.3389/fpsyg.2021.698490
[56] M. Tuncay, C. Borekei, and S. Demirel, "Quality assurance in distance education through data mining", *International Journal of Technology in Education and Science*, vol. 6, no. 3, pp. 443–457, 2022. https://doi.org/10.46328/ijtes.396
[57] M. A. Prada, M. Domínguez, J. Lopez, and P. Alves, "Educational data mining for tutoring support in higher education: A web-based tool case study in engineering degrees", *Journal IEEE Access*, vol. 8, pp. 212818–212836, 2020. https://doi.org/10.1109/ACCESS.2020.3040858
[58] E. U. Okike, and M. Mogorosi, "Educational data mining for monitoring and improving academic performance at university levels", *International Journal of Advanced Computer Science and Applications*, vol. 11, no. 11, pp. 570–581, 2020. https://doi.org/10.14569/IJACSA.2020.0111171
[59] N. Z. Salih, and W. Khalaf, "Prediction of student's performance through educational data mining techniques", *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1708–1715, 2021. https://doi.org/10.11591/ijeecs.v22.i3.pp1708-1715

[60] G. Czibula, G. Ciubotariu, M.-I. Maier, and H. Lisei, "IntelliDaM: A machine learning-based framework for enhancing the performance of decision-making processes. A case study for educational data mining", *IEEE Access*, vol. 10, pp. 80651–80666, 2022. https://doi.org/10.1109/ACCESS.2022.3195531

[61] S. Kingir, B. Gok, and A. Selman, "Exploring relations among pre-service science teachers' motivational beliefs, learning strategies and constructivist learning environment perceptions through unsupervised data mining", *Journal of Baltic Science Education*, vol. 19, no. 5, pp. 804–823, 2020. https://eric.ed.gov/?id=EJ1271116; https://doi.org/10.33225/jbse/20.19.804

[62] A. Urbina-Nájera, A. Téllez-Velásquez, and R. C. Barbosa, "Patterns to identify dropout university students with educational data mining", *Electronic Journal of Educational Research*, vol. 23, no. 29, pp. 1–15, 2021. https://doi.org/10.24320/redie.2021.23.e29.3918

[63] J. Zárate-Valderrama, N. Bedregal-Alpaca, and V. Cornejo-Aparicio, "Classification models to recognize dropout patterns in university students", *Ingeniare Chilean Engineering Journal*, vol. 29, no. 1, pp. 168–177, 2020. https://doi.org/10.4067/S0718-33052021000100168

[64] G. T. Bohorquez, "University extension in a public university", *Culture, Education and Society*, vol. 11, no. 2, pp. 221–236, 2020. https://dialnet.unirioja.es/servlet/articulo?codigo=7823 617; https://doi.org/10.17981/cultedusoc.11.2.2020.14

[65] M. Ruiz-Corbella, and E. López-Gómez, "The role of the university in 21st Century: Understanding its origin in order to project its future", *High School Journal*, vol. 48, no. 189, pp. 1–19, 2019. https://www.scielo.org.mx/pdf/resu/v48n189/0185-2760-resu-48-189-1.pdf; https://doi.org/10.36857/resu.2019.189.612

[66] J. E. D. Nader, A. P. Panunzio, and M. I. Hernández, "Scientific research: A university function to be considered in the ecuadorian context", *Edumecentro Journal*, vol. 10, no. 4, pp. 166–179, 2018. https://dialnet.unirioja.es/servlet/articulo?codigo=6689173

9 AUTHORS

Martín Díaz-Choque is an electrical mechanical engineer from the National Technological University of Lima Sur. He is a researcher recognized by CONCYTEC (National Council of Science, Technology and Technological Innovation).

Omar Chamorro-Atalaya is a research professor at the Faculty of Engineering and Management of the National Technological University of Lima Sur (UNTELS).

Orlando Adrian Ortega-Galicio is a senior professor at the Faculty of Engineering and Management at the National Technological University of South Lima (UNTELS).

José Antonio Arévalo-Tuesta is a research professor at the Faculty of Economic Sciences of the Federico Villarreal National University (UNFV).

Elvira Cáceres-Cayllahua is a professor attached to the academic department of Business Management, Faculty of Economics and Planning of the Universidad Nacional Agraria La Molina (UNALM).

Ronald Fernando Dávila-Laguna is a professor attached to the research area of the Cesar Vallejo University (UCV). He has a doctorate in administration and a master's degree in Social Development Management.

Irma Esperanza Aybar-Bellido is a professor at the Faculty of Health Sciences of the San Juan Bautista Private University (UPSJB).

Yina Betty Siguas-Jerónimo is a professor at the Faculty of Health Sciences of the San Juan Bautista Private University (UPSJB).