

The Use of Tools of Data Mining to Decision Making in Engineering Education - A Systematic Mapping Study

Buenaño-Fernandez, Diego¹, Villegas-CH, William¹ and Luján-Mora, Sergio²

¹Facultad de Ingenierías y Ciencias Aplicadas, Universidad de las Américas, Quito Ecuador

²Departamento de Lenguajes y Sistemas de Informáticos, Universidad de Alicante

Correspondence to Buenaño-Fernandez, Diego(diego.buenano@udla.edu.ec)

Abstract

In recent years, there has been an increasing amount of theoretical and applied research that has focused on educational data mining. The learning analytics is a discipline that uses techniques, methods and algorithms that allow the user to discover and extract patterns in stored educational data, with the purpose of improving the teaching-learning process. However, there are many requirements related to the use of new technologies in teaching-learning processes that are practically unaddressed from the learning analytics. In an analysis of the literature, the existence of a systematic revision of the application of learning analytics in the field of engineering education is not evident. The study described in this article provides researchers with an overview of the progress made to date and identifies areas in which research is missing. To this end, a systematic mapping study has been carried out, oriented towards the classification of publications that focus on the type of research and the type of contribution. The results show a trend towards case study research that is mainly directed at software and computer science engineering. Furthermore, trends in the application of learning analytics are highlighted in the topics, such as student retention or dropout prediction, analysis of academic student data, student learning assessment and student behavior analysis. Although this systematic mapping study has focused on the application of learning analytics in engineering education, some of the results can also be applied to other educational areas.

Keywords Decision Making, Educational Data Mining, Engineering Education, Learning Analytics, Systematic Mapping.

I. INTRODUCTION

Learning analytics (LA) and educational data mining (EDM) are disciplines that analyze educational data about learners and their contexts. This analysis is done through a variety of methods, techniques and statistical tools, including machine learning and data mining (DM). The objective of LA is to provide an analysis of the data that exist in educational repositories such as learning management systems (LMS), for the purpose of understanding and optimizing learning and the environments in which it occurs [1]. This analysis can significantly support decision making in terms of planning on the part of teachers and managers of education institutions [2]. This topic has been dealt with in several articles, taking into account the different requirements of the stakeholders involved including teachers, instructors, students, course developers, researchers, educational institutions, and private training companies. Depending on the stakeholder's needs, different DM techniques have been applied [3]. These works relate to the field of education in general. However, in the field of engineering education, there is a need for a systematic identification of publications related to the importance of using LA and EDM in decision making.

The profession of engineering and, therefore, education in engineering, are fundamental pillars in the industrial and technological development of society [4]. Engineering education has some particular characteristics that must be taken into account when evaluating its effectiveness. For example, it has been proven that engineering students use certain learning styles that differ from students in other areas of study [5]. In the study carried out by [6], a trend it is observed in learning styles such as visual, active, inductive and sensing. Thus, the educational models that are applied in other areas cannot necessarily be replicated with the same logic with regard to engineering degree [5]. Additionally, although engineering

disciplines show similar patterns in all countries, there are differences in engineering practice across national cultures, especially in the way that the theoretical contributions of engineering science are used in the practice of engineering [4]. For example, in difficult abstract subjects such as mathematics, physics and simulation, it is recommended that a tutor teach these subjects rather than implementing constructivist models, where students develop their own knowledge. Learning in engineering requires developing dynamic structures for obtaining knowledge, aimed at solving practical problems [7]. In engineering, problem solving is of a different nature compared with other fields. Solving a design problem involves an extensive process of analysis, modeling, experimentation and execution, in which key decisions must be made [8]. The development of new products and methods, and the application of knowledge to new problems are everyday challenges for engineers [4]. On the other hand, it should be mentioned that the implementation of the study plans for engineering degree is complicated by the fact that its content, especially with regard to the practical aspect, is dynamic and is changing very fast [9]. In [1] it is proposed the creation of study sub-communities of LA with the aim of discussing and establishing round tables or workshops on specific topics in a particular area of study. This article serves as the basis for carrying out LA proposals in the area of engineering education.

A systematic mapping study is a means to categorize and summarize all the available research that is relevant to a specific subject area or to particular research questions in an unbiased manner [10]. The objective is to present a clear evaluation of a research topic, using a reliable, rigorous and auditable methodology. The systematic mapping study is considered by many to be a first and essential step in deciding in which specific areas to work in creating a more detailed systematic review [11], [12]. A systematic literature review and a systematic mapping study share certain commonalities, are complementary, and aim to identify research gaps. On the other hand, they are different to data analysis in terms of goals and approaches [13].

With respect to data analytics in the educational area (EDM and LA), in general, several surveys and systematic reviews have been published [3], [14], [15]. However, specifically in the area of engineering education, to the best of our knowledge, no empirical evidence-based studies have been reported that categorize and summarize studies that have been undertaken. Therefore, this is the first work conducting an analysis of academic papers reporting uses and results of LA in the area of engineering education.

It is important to note that there are several definitions of LA and that not all authors agree on them [16]. A widely-accepted approach is the one proposed by the Society for Learning Analytics Research¹ that defines LA as “...the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in which it occurs”. The final purpose of the LA contributes to the improvement of learning through the interpretation and contextualization of educational data.

An additional term or concept used in EDM is provided by Academic Analytics. In [17] it is defined it as the application of business intelligence tools to learning. Academic analytics aims to present an academic information report. These reports are focused solely for the benefit of educational institutions. In the other hand, LA is more focused on data analysis for the making decisions related to students, teachers and the learning process itself [18].

Another analytic possibility that we should consider is EDM, which exploits machine-learning, statistical and data-mining algorithms in terms of different types of educational data. The main objective is to analyze these types of data in order to resolve educational research issues [19]. LA also shares many attributes with EDM [20]. Thus, whether EDM is a subset of LA or is a different discipline with some points in common, is still an open debate for many researches [16].

The objective of the study presented in this article is to categorize and summarize publications related to LA in the field of engineering education, providing a basis for reflection on the part of members of this academic community. The content of this systematic mapping is organized as follows. Section II covers the description of the research method used in this study. Section III reports on the analysis and results obtained. Section IV presents a detailed discussion of these results. Finally, Section V presents the conclusions and suggestions for future work.

II. RESEARCH METHOD

The systematic mapping approach of this work uses a methodology that has been applied in others works [13], [21], [22], [23], [24]. It is based on four stages: feasibility, planning, classification scheme and the presentation of results, as shown in Fig. 1.

A. FEASIBILITY

In this phase, an initial analysis was carried out to determine the possibility of carrying out the systematic mapping. This analysis was made from the execution of the following tasks:

- Determine if systematic mappings, surveys or systematic reviews of literature on the subject of study have been previously carried out.
- Define initial search string.
- Apply the initial search string to the indexes and scientific databases, in order to know a first approximation of the number of previous studies with respect to the subject to be investigated.

¹ <http://www.solaresearch.org/mission/about/>

- Identify the keywords used in the previous works, with the aim of having an idea for refining the search strings.
- Determine the candidate search string.

The first search of previous systematic mappings, surveys and systematic reviews on LA and EDM in the specific area of engineering education was negative, no similar works were found, which makes this proposal unique and pioneering. Based on the articles identified and that make explicit reference to the use of LA and EDM in engineering education, initial search string was developed and tested. According to this string, we proceeded to search exclusively on the titles of the publications; however, this considerably limited the number of articles obtained since they did not reach more than ten in each database. With these results, it was tried to search the main keywords (LA and EDM) in the titles of the publications while the secondary terms (Decision making, Engineering education) were searched in the abstract of the documents. In this way, it was possible to obtain enough documents that truly contribute with the proposed mapping.

B. PLANNING

In the planning stage, all the relevant decisions are taken to carry out the systematic mapping study. In this stage, we propose a protocol for the search and collection of articles based on three tasks: Establishment of the research question, definition of the search strategy and selection of primary studies, and study selection - criteria for inclusion / exclusion.

It is recommended to record the information at all stages of the process including planning. It is also noteworthy that the mapping process is iterative and may need revisions. On the other hand, in this stage, it is important to define the tools for recording data such as spreadsheets and reference management software.

1) ESTABLISHMENT OF THE RESEARCH QUESTION

According to the objectives of this study, the research questions focused on categorizing and structuring the publications relating to LA and EDM in the field of engineering education. For this, four research questions were identified, as shown in Table 1.

These research questions will help to characterize the current knowledge about LA and EDM in the field of engineering education, in order to identify gaps which would allow the identification of specific areas to allow a more detailed investigation to be conducted.

2) DEFINITION OF THE SEARCH STRATEGY AND SELECTION OF PRIMARY STUDIES

The search string used as the basis for obtaining the relevant papers took into account the existence of terms related to LA, which were then identified and grouped into sets. These keywords were identified through test searches and references in the sample documents. These terms were: “Educational Data Mining” and “Knowledge Discovery in Databases”. Therefore, the initial search string was as follows:

(“Learning Analytics” OR “Educational Data Mining”) OR “Knowledge Discovery in Databases” AND (“Engineering”) AND (“Decision Making”).

This string was applied to documents published between January 2012 and December 2017, the latter being the date on which the search was conducted. This range was chosen because, from 2012 onwards, the number of publications began to grow exponentially. The scientific databases that were searched were ACM Digital Library, Google Scholar, IEEE Xplore, Scopus, ScienceDirect and Web of Science. These sources ensured that quality publications were included within the research domain. For each database, a specific search string was configured according to the syntax used by each one. The factors described below guided the selection of the digital databases with which systematic mapping was carried out: (1) number of documents indexed, (2) frequency of update, and (3) options to execute advanced queries. For the study, six digital library and abstract and citation database were selected, which met the proposed factors. Table 2 shows the search strings for each database.

The search string was manually refined according the features of the search engine of each database. In all cases it was necessary to use the advanced search option.

FIGURE 1 Phases of systematic mapping.

TABLE 1 Research questions and motivations of the study

[Insert Table 1]

TABLE 2 Search string for each digital library and abstract and citation database

[Insert Table 2]

3) STUDY SELECTION - CRITERIA FOR INCLUSION / EXCLUSION

Inclusion criteria are the characteristics that prospective studies must have to be accepted in the systematic mapping study, whereas exclusion criteria are those characteristics that reject prospective studies. On the one hand, inclusion criteria used are language of publication, relevance of the journal or conference and that is specific to the topic to be considered [13]. On the other hand, exclusion criteria are studies that are not relevant when it comes to answering the research questions and that do not meet basic quality criteria defined in previous studies [23].

The inclusion and exclusion criteria, which have been defined for the present work, are objective, easy to verify and do not require further interpretation. The criteria have the function of guaranteeing that the selected articles comply with some parameters that have been evaluated in previous publications [13], [21], [22].

In addition, in order to ensure that only relevant studies were entered into the pool of publications that would ultimately be analyzed, prior to inclusion the researchers evaluated each identified publication. In this evaluation it was verified that the articles complied with the following criteria:

- Articles published in peer-reviewed journals and international conference proceedings.
- Studies that showed the use of LA and EDM oriented to decision making exclusively in the area of engineering education.
- Articles published in the English language.

In addition, publications that met at least one of the following conditions were excluded:

- Articles of less than five pages.
- Articles overlapped by another paper from the same authors presenting more complete results. In those cases, we only kept the extended version.
- Duplicate articles from the same study in different sources.

FIGURE 2 Systematic mapping study process.

After the initial search in which 3,732 studies were identified, the inclusion and exclusion criteria were applied. In a first run of this process, the number of documents was reduced by 84 % (3,118). This allowed us to refine the initial base and keep relevant articles related to the proposed mapping. Of the 614 articles selected, we proceeded to read the abstract, the keywords and the introduction, leaving 518 documents after this process. These documents were completely analyzed (full-text scanning) leaving a total of 56 articles. Finally, a quality evaluation was applied to these articles, after which there were 52 articles. The details of this process are shown in Fig. 2.

Once the full-text reading was done, it was agreed that 16 articles should be excluded. In addition to the general inclusion and exclusion criteria, it was considered important to assess the quality of the primary studies. This quality assessment was performed by applying the following questions to each of the selected studies [25]:

- Does the study present a method and/or technique for assessing the quality of DM techniques that are implemented in educational settings?
- Has the study been published in a relevant journal or conference? (e.g. journal ranked in JCR or SJR, conference ranked in CORE²)
- Has the study been cited by other authors?

After all the steps of this phase were completed, the final number of articles that passed to the classification and data extraction phase was 52.

C. CLASSIFICATION SCHEME - DATA EXTRACTION STRATEGY

According to the classification scheme suggested by [26], the data extraction strategy is based on the exploration of certain parts of each article (abstract, introduction and results). Only in those cases where the information is not clear or is not well structured, a complete reading of the paper was performed. For our study, we followed the systematic process shown in Fig. 3 This is based on the execution of key wording, which is a strategy to reduce the time needed to develop the classification scheme and to ensure that the scheme consider the existing studies [26].

The data extraction strategy involved dividing each research question into more specific criteria. In this way, a procedure was defined through which a set of specific terms was established for each research question, thus facilitating the extraction

² <http://www.core.edu.au/conference-portal>

and categorization of documents [24]. Table 5 shows the division of the research questions into criteria and concepts that aim to support the data extraction and categorization of publications [25].

If there were doubts as to the classification of a specific publication, the scheme was refined whereby concepts were added, modified or eliminated. In such a situation, the previously classified publication had to be evaluated once more to validate their classification within the new scheme. This procedure was very useful as it led to a stable classification scheme [23]. The procedure is detailed in Fig. 4. To document the data extraction process, we used a concept matrix, the same one that was done in Microsoft Excel. According to [27], we can find two types of systematic reviews of literature, revisions focused on concepts and revisions focused on authors. In the first approach, the concepts determine the organizational framework of the review. While in the second approach the authors are the ones who determine the organizational framework of the review. This approach fails to synthesize the literature. A matrix of concepts is built up throughout the systematic review process. The structure is similar to the one shown in Fig. 5. For the proposed mapping, some concepts were defined along with the research questions; however, as the article review was carried out, these concepts were refined.

FIGURE 3 Systematic classification schema (based on [26]).

FIGURE 4 Procedure of classification scheme.

FIGURE 5. Concept matrix scheme (based on [27]).

Table 3 Division of research questions into criteria and concepts

[Insert Table 3]

III. RESULTS OF THE MAPPING

To evaluate and interpret the findings with regard to the collected studies we used non-statistical methods. According to the search strategy and the classification scheme followed, most of the identified case studies were experimental or exploratory in nature. Some of these were evaluation studies, while others were empirical studies [28],[29],[30]. In addition, application areas within engineering differed from study to study, but most of them focused on science, technology, and mathematics [15]. Table 4 shows the total number of relevant studies distributed per year and by source. As can be seen in this table, a constant increase of articles is detected in the analyzed period, especially from the year 2015. The data that is shown in the year 2017 can be increased because many articles have not yet been published in the selected databases. In particular, those articles that were presented in the last quarter of 2017.

The overall results are based on the review of the selected primary studies which have been classified in relation to each research questions. The next step was to perform a detailed analysis using the final documentary corpus, through a descriptive synthesis of the criteria and concepts laid out in Table 3. This meant answering each of the research questions raised in the research protocol. Table 5 shows the summary of the results obtained in the study. These results were organized according to the defined criteria and the several responses identified in each publication. In addition, the number of publications regarding each response option and the respective references are presented.

TABLE 4 Summary of relevant studies per year

[Insert Table 4]

TABLE 5 Mapping of studies per criteria and identified answer

[Insert Table 5]

A. THE MAIN SOURCES OF DATA TO WORK WITH LA AND EDM (RQ1)

Systematic mapping results indicate that, in 18 of the 52 articles analyzed (34 %), the log files or records of the LMS of the institution in question are used as data sources. However, many of the LA applications have been conducted with information external to the LMS such as institutional databases, student achievement academic records (29 %), student admissions records or databases created specifically for this purpose (25 %). The articles according to the type of data source used are shown in Table 5. This distribution makes it clear that greater emphasis needs to be placed on the application of LA or EDM to non-traditional data sources, such as unstructured data. These types of data require different structures and formats, since they come from different sources such as web usage logs, click density maps, tweets, Facebook chats, text messages and videos.

FIGURE 6 Areas of applications of the LA or EDM.

When discussing to specific databases, we refer to articles in which it was observed that the data come from questionnaires, student admission records, e-portfolios or by using game-based learning (GBL) [12], [31], [32], [33], [34]. These methods of teaching are aimed at improving student learning by adapting to their reality.

B. METHODS OF LA AND EDM MOST COMMONLY USED IN THE FIELD OF ENGINEERING EDUCATION (RQ2)

An important parameter that was analyzed in the mapping study has to do with the methods or techniques of DM adopted by the authors to analyze the data collected. In the field of engineering education, and according to the proposed mapping, the methods and techniques most commonly used within LA and EDM are analysis and visualization (19 %), association rules, clustering and classification (25 %), techniques of preprocessing and filtering (10 %), statistical techniques (7 %) and not specific method or technique (17 %). Among the analyzed publications, it is observed that there is a significant percentage of studies that deal with a specific job of developing applications or frameworks oriented to the LA or EDM. Therefore, the application of a DM method or technique cannot be identified in all publications [33], [35], [36], [37].

C. AREAS OF APPLICATION OF LA AND EDM RESULTS IN ENGINEERING EDUCATION (RQ3)

It is possible to identify several applications or tasks in educational environments in which LA or EDM has been executed. For example, [38], [39], [40] suggest some key areas of application for EDM: improving student models, improving domain models and studying the pedagogical support provided by learning software. In [41] it is suggest the following EDM subjects/tasks:

- Applications dealing with the assessment of the student's learning performance.
- Applications that provide course adaptation and learning recommendations based on the student's learning behavior.
- Approaches dealing with the evaluation of learning material and educational web-based courses.
- Applications that involve feedback to both teacher and students in terms of e-learning courses and
- Developments for the detection of atypical student learning behavior.

Fig. 6 shows a summary of the number of articles by each area of application with regard to LA or EDM, sorted by year, in the area of engineering education. In the analyzed works, it is observed that the main application of LA and EDM, 19 publications (36 %) is improvement in terms of the academic development of the students. This application is directly related to the student's participation in the educational process such as in terms of the analysis of learning styles or learning strategies, the analysis and improvement of student skills, and the motivation trends of students, among others. A second application area of LA and EDM, identified in this study with 6 publications (10 %), is the prediction of student attrition. In engineering education, this is a critical aspect that must be taken into account to mitigate its effect on the educational outcomes. Another important area of application has to do with plans to improve education such as the analysis of traditional teaching methods, types of feedback in online teaching models, monitoring and improving the quality of teaching, among others. In the study, 7 publications related to this area were identified (10 %). The improvement in the entire teaching-learning process is an area in which special emphasis is placed on the articles analyzed, here 16 publications (31 %) have been found. For example, in [42], [43], [44], [45] specific works are identified in which the application of LA and EDM has focused on the improvement of teaching-learning processes.

FIGURE 7 Engineering degree that most frequently apply LA and EDM studies.

D. ENGINEERING CAREER MOST FREQUENTLY APPLY LA AND EDM STUDIES (RQ4)

A significant percentage of the publications analyzed (30 %), do not specify having carried out a study of a specific engineering career. On the other hand, it is observed in Fig. 7 that 60 % of the publications correspond to the all studies related a degree in the area of information and communication technologies (ICT).

IV. DISCUSSION

Nowadays, there is a wide variety of educational environments such as e-learning systems, hypermedia educational systems, learning object repositories, social network learning, educational games environments, ubiquitous computing environments, etc. Each system provides data with heterogeneous structures and formats, and with different forms of storage and retrieval. It is for this reason that it is necessary to identify the data sources of each system/environment, with the objective of applying

specific mechanisms related to preprocessing and filtering. In the results of this systematic mapping study, there is a marked trend (approximately 63 % of the works analysed) using two sources of data (RQ1): files or logs of LMS and institutional databases. Generally, higher education institutions maintain distributed information systems, which complicates the application of LA and EDM skills. In this sense, it is necessary to deepen the design and implementation of data warehouse projects in educational environments [46], [35]. These projects allow the user to consolidate different heterogeneous data sources into a single repository with regard to which they can start the DM process.

According to an IBM study mentioned in [47], 80 % of organizations handle unstructured data that emerge from a variety of formats such as text, videos, audio, diagrams, images and combinations of two or more formats. Only 13 publications out of a total of 52 (25 %) have oriented their study in terms of non-traditional data sources such as structured questionnaire databases [29], [30], [48], [49], educational games databases [32], [33], [34], [36] e-portfolio databases [50]. A special approach is needed for the analysis of the databases generated from the application of the GBL concept. A special approach is needed for the analysis of the databases generated from the application of the GBL concept. Increasingly, educational games are more popular in teaching processes in higher education, as they have proven to be effective and active learning tools. Serious games have the ability to involve the student, making their learning experience a dynamic activity where the student needs to make decisions and learn from their effects. They offer different levels of difficulty and allow us to adapt the content to the user performance [32].

According to [34], the findings revealed that the use of computer games is linked to a range of perceptual, cognitive, behavioral, affective and motivational impacts and outcomes. In this sense, it is necessary to evaluate these data sources and to consider the possibility of applying LA and EDM techniques in order to evaluate the development of learning skills. On the other hand, it is also important to consider the existence of other popular learning environments such as cognitive tutors, mobile and ubiquitous environments, massive open online courses and social learning platforms [15]. The constant growth of these learning environments generates different types of data, which require special attention in their preprocessing phase. In many cases, it is possible to find unstructured data. As these learning environments continue to gain space and achieve prominence in higher education, it becomes necessary to analyze the underlying problems that emerge from the application of technology in the teaching-learning process. An example is presented in [51] in which a method of meta-analysis is established that allows the evaluation of learning in relation to the different characteristics of the technology used in a case study.

In relation to the application of LA and EDM methods and techniques (RQ2), the mapping study shows that the most commonly-used approaches in the field of engineering education are discovery using models and visualization (10 publications) (19 %) followed by association rules, clustering and classification (13 publications) (25 %). Studies show that visualization techniques are closely related to the implementation of dashboard [28], [52], [53], [54]. This technique is especially relevant for engineering students because of their familiarity with statistical numerical environments. This technique allows students to reflect on their own academic performance and allows teachers to make decisions based on the graphical observation of student performance. Four publications (8 %) present particular techniques that have not been considered in previous studies in the form of mining software repositories [55], multimodal LA [56], text mining [55] and machine learning methods [43]. On the other hand, the application of preprocessing and filtering techniques [57], [58] is of special relevance due to the variety of data sources that are currently available in terms of DM.

Around 17 % of the publications analyzed do not relate to the use of a specific method or technique of DM. This is due to the fact that the case studies focused on aspects such as the development of tools to support EDM management [59], [60], [61], the development of specific purpose systems [35], [39], the application of game-based teaching-learning methodologies [33] and framework development [36], [37]. This is justified because most of the publications analyzed are related to software engineering and computer engineering. In this sense, the results of these studies (frameworks, applications, methodologies) can be applied in LA and EDM research that are proposed in other engineering degree.

The classification of publications according to RQ3 is illustrated in Table 5. As can be seen in this table, 3 % of the studies investigate issues related to improving the academic development of students. In this group of articles, the authors focus on topics such as the types of learning activities that engineering students prefer, their learning styles, how to learn in complex learning environments, and learning engagement [30], [62], [63]. This research will enable students to improve the quality of their learning experience and to foster self-regulated learning. Some 31 % of the studies emphasize the application of LA and EDM in order to improve the teaching-learning process [45] through mapping students' learning processes [43], evaluating and providing feedback in real time to improve a course [64], proposing teaching/learning tools that allow students to engage in learning activities [61], and choosing an appropriate teaching methodology based on knowing the students [48].

Eight of the studies (15 %) considered are aimed at predicting the student dropout rate and assess general competencies in engineering [33], [50], [65]. Additionally, two studies focus on the evaluation of generic competences in engineering [33], [66]. These two aspects are intrinsically related, in that the permanence of students in a career has to do with their understanding of the learning outcomes or competencies they must develop throughout their academic training. LA and EDM have much to contribute in these matters. On the one hand, the analysis of the academic data relating to students will allow academic institutions to validate the achievement of certain learning outcomes. On the other hand, the analysis of the

causes of student dropout will enable key decisions to be made in the teaching-learning process. An area of special interest is the application of the results obtained from LA and EDM in the improvement of educational plans. About 15 % of the articles reviewed focus on this topic. Within this group of articles there are studies related to teacher evaluation data [49]. Authors also explore, identify and evaluate various factors as indicators of performance for academic prediction purposes [60], [65].

In relation to the RQ4, the degree in which LA and EDM techniques are most frequently applied are engineering related to ICT (42 % of publications), followed by the mathematics and physics degree (13 % of publications). These results allow us to identify an important research gap with regard to which special emphasis should be placed in terms of future research. Degrees such as environmental engineering, biotechnology engineering, mechanical engineering, civil engineering or agroindustry engineering should be taken into account in order to analyze the behavior of LA and EDM methods and techniques on teaching-learning processes in which students, teachers and educational authorities are involved. This research challenge consists of validating the functionality of the methods and techniques of LA and EDM in this group of degrees, or alternatively studying the characteristics of these degree in order to apply techniques that allow a strengthening of the learning assessment process. This work can be done through the creation of sub-communities of study in which we can discuss and establish standards for more particular aspects of LA [1].

V. CONCLUSIONS AND FURTHER WORK

Previous reviews of the research literature concerning LA and EDM have provided a meaningful understanding of the conceptual basis of this domain. However, these studies have not engaged in an analysis of the results of research in the area of engineering education. In the systematic mapping developed in this work, we have searched the literature and collected representative articles of real cases with real data, in terms of both the LA domain and EDM in engineering environments. The analysis of selected case studies and their results reveal the potential of this emerging field of educational research. Along with the opportunities that have emerged, a series of gaps have been identified that require the immediate attention of researchers. From the systematic mapping performed, Table 6 details some topics that have been identified as opportunities and trends in terms of LA. These subjects have been implemented basically in engineering degrees related to information and communication technologies. As a result, gaps and, at the same time, research opportunities that need to be channeled to other engineering degrees can be identified.

TABLE 6 Opportunities and trends for development of LA or EDM in engineering education

[Insert Table 6]

An analysis parameter with regard to the application of LA and EDM that was not observed in this systematic mapping study is related to the study of affective factors that influence learning outcomes. The affective factors deserve special consideration in engineering education environments, where the levels of tension due to the academic load are high, thus affecting the academic performance of the students. The area of education in mathematics has been tested with the application of various types of smart learning technologies [33], [34]. One of the objectives of implementing these technologies in the educational field is to include the affective and behavioral characteristics of the student in the teaching-learning process. Due to the characteristics of these technologies, the analysis of the results of their implementation must be carried out through the application of LA and EDM techniques.

Another crucial issue in LA and EDM research is the application of DM methods and techniques through on-line data visualization tools that allow the user to identify, in a timely manner, students at risk of dropping-out by informing them periodically about their progress, and even comparing them with their peers. On the other hand, the prediction of drop-out and retention are also key issues for LA and EDM research. In order to predict student, drop-out in the early stages, the authors have observed that the use of classification algorithms is one of the most viable solutions for the process. Additionally, the literature analyzed shows that there are other methods such as social network analysis and discovery with models that could be used. These methods or techniques are a trend within DM in educational settings, and therefore it is an important challenge to study and apply them to education in engineering. A future work may be to contrast the results of this article with the systematic mapping of other areas, and thus shed light on the discourse of LA and EDM in different fields.

REFERENCES

- [1] C. Romero and S. Ventura, "Educational data mining: a review of the state of the art," *IEEE Trans. Syst. man, Cybern. - Part c Appl. Rev.*, vol. 40, no. 6, pp. 601–618, 2010.
- [2] U. Jørgensen, "Historical accounts of engineering education," *Rethink. Eng. Educ. CDIO Approach*, pp. 216–240, 2007.
- [3] P. Rosati, "Specific differences and similarities in the learning preferences of engineering students," in *Proceedings - Frontiers in Education Conference Designing the Future of Science and Engineering Education*, 1999, p. 12C1/17-12C1/22.
- [4] R. M. Felder and R. Brent, "Understanding Student Differences," *J. Eng. Educ.*, vol. 94, no. 1, pp. 57–72, 2005.
- [5] J. C. Perrenet, P. A. J. Bouhuijs, and J. G. M. M. Smits, "The Suitability of Problem-based Learning for Engineering Education: Theory and

practice," *Teach. High. Educ.*, vol. 5, no. 3, pp. 345–358, Jul. 2000.

- [6] R. Felder, "Engineering Education: a Tale of two paradigms," in *Proceedings of 2nd. International Conference on Geotechnical Engineering Education*, 2012, pp. 9–14.
- [7] X. Ochoa, A. Hershkovitz, A. Wise, and S. Knight, "Towards a Convergent Development of Learning Analytics," *J. Learn. Anal.*, vol. 4, no. 3, pp. 1–6, 2017.
- [8] B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature Reviews in Software Engineering," *Keele Univ. Durham Univ. Jt. Rep.*, vol. 2, p. 1051, 2007.
- [9] E. Martínez-Cámara, M. Á. García Cumbreiras, M. T. Martín Valdivia, and L. A. Ureña López, "Detección de la polaridad de tweets en español," 2011.
- [10] M.-C. Li and C.-C. Tsai, "Game-Based Learning in Science Education: A Review of Relevant Research," *J. Sci. Educ. Technol.*, vol. 22, no. 6, pp. 877–898, 2013.
- [11] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf. Softw. Technol.*, vol. 64, no. 1, pp. 1–18, 2015.
- [12] J. T. Avella, M. Kebritchi, S. Nunn, and T. Kanai, "Learning Analytics Methods, Benefits, and Challenges in Higher Education: A Systematic Literature Review," *Online Learn. J.*, vol. 20, no. 2, pp. 13–29, 2016.
- [13] Z. Papamitsiou and A. Economides, "Learning Analytics and Educational Data Mining in Practice : A Systematic Literature Review of Empirical Evidence," *J. Educ. Technol. Soc.*, vol. 17, no. 4, pp. 49–64, 2014.
- [14] M. A. Conde, F. J. Garcia-Penalvo, D.-A. Gomez-Aguilar, and R. Theron, "Exploring Software Engineering Subjects by Using Visual Learning Analytics Techniques," *IEEE Rev. Iberoam. Tecnol. del Aprendiz.*, vol. 10, no. 4, pp. 242–252, 2015.
- [15] P. J. Goldstein and R. N. Katz, "Academic Analytics : The Uses Of Management Information And Technology In Higher Education," *Educ. Q.*, vol. 8, no. December, p. 113, 2005.
- [16] G. Siemens and P. Long, "Penetrating the Fog: Analytics in Learning and Education," *Educ. Rev.*, vol. 46, no. 1, pp. 31–40, 2011.
- [17] T. Barnes, M. Desmarais, C. Romero, and S. Ventura, "Educational Data Mining 2009," in *Proceedings of the 2nd International Conference on Educational Data Mining (EDM)*, 2009, pp. 1–338.
- [18] A. Van Barneveld, K. Arnold, and J. Campbell, "Analytics in higher education: Establishing a common language," *Educ. Learn. Initiat.*, no. January, pp. 1–11, 2012.
- [19] A. Fernandez, E. Insfran, and S. Abrahão, "Usability evaluation methods for the web: A systematic mapping study," *Inf. Softw. Technol.*, vol. 53, no. 8, pp. 789–817, 2011.
- [20] V. Cosentino, J. L. Canovas Izquierdo, and J. Cabot, "A Systematic Mapping Study of Software Development With GitHub," *IEEE Access*, vol. 5, pp. 7173–7192, 2017.
- [21] P. Cedillo, A. Fernandez, E. Insfran, and S. Abrahão, "Quality of Web Mashups: A Systematic Mapping Study," in *Current Trends in Web Engineering SE - 8*, vol. 8295, Denmark: Springer, Cham, 2013, pp. 66–78.
- [22] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic mapping studies in software engineering," in *Proceedings of the 12th international conference on Evaluation and Assessment in Software Engineering (EASE)*, 2008, pp. 68–77.
- [23] D. Macchi and M. Solari, "Mapeo Sistemático de la Literatura sobre la Adopción de Inspecciones de Software," in *Proceedings of conferencia Latinoamericana de Informática (CLEI)*, 2012, pp. 1–8.
- [24] R. Wendler, "The maturity of maturity model research: A systematic mapping study," *Inf. Softw. Technol.*, vol. 54, no. 12, pp. 1317–1339, 2012.
- [25] J. Webster and R. T. Watson, "Analyzing the Past to Prepare for the Future : Writing a Literature Review," *MIS Q.*, vol. 26, no. 2, pp. xiii–xxi, 2016.
- [26] J. L. Santos, K. Verbert, S. Govaerts, and E. Duval, "Goal-oriented visualizations of activity tool tracking. A case study with engineering students," in *Proceedings of 2nd International Conference on Learning Analytics and Knowledge (LAK)*, 2012, pp. 143–152.
- [27] K. Abhyankar and S. Ganapathy, "Technology-Enhanced Learning Analytics System Design for Engineering Education," *Int. J. Inf. Educ. Technol.*, vol. 4, no. 4, pp. 345–350, 2014.
- [28] N. Pratheesh and T. Devi, "Sensation of learning analytics top prevail the software engineering education," in *Proceedings of International Conference on Advanced Computing & Communication Systems (ICACCS)*, 2013, pp. 845–850.
- [29] C.-U. Lei, K. L. Man, and T. O. Ting, "Using Learning Analytics to Analyze Writing Skills of Students: A Case Study in a Technological Common Core Curriculum Course," *Int. J. Comput. Sci.*, vol. 41, no. 3, pp. 41–45, 2014.
- [30] N. Pratheesh and T. Devi, "Necessity of learning analytics in software engineering education," *J. Eng. Sci. Technol.*, vol. 10, no. 3, pp. 269–281, 2015.
- [31] V. Tam, E. Y. Lam, S. T. Fung, W. W. T. Fok, and A. H. K. Yuen, "Enhancing educational data mining techniques on online educational resources with a semi-supervised learning approach," in *Proceedings of IEEE International Conference on Teaching, Assessment, and Learning for Engineering (TALE)*, 2015, pp. 203–206.
- [32] O. Moscoso-Zea, Andres-Sampedro, and S. Lujan-Mora, "Datawarehouse design for educational data mining," in *Proceedings of 15th International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2016, pp. 1–6.
- [33] I. Menchaca, M. Guenaga, and J. Solabarrieta, "Using learning analytics to assess project management skills on engineering degree courses," in *Proceedings of the Fourth International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM)*, 2016, pp. 369–376.
- [34] Y. Vanessa Niet, V. G. Diaz, and C. E. Montenegro, "Academic decision making model for higher education institutions using learning analytics," in *Proceedings of 4th International Symposium on Computational and Business Intelligence (ISCBI)*, 2016, pp. 27–32.
- [35] A. U. Khasanah, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 215, no. 2, pp. 1–7.
- [36] W. Lie, L. Hai-yan, C. Biao, and Z. Yuan, "Application of Educational Data Mining on Analysis of Students' Online Learning Behavior," in *International Conference on Image, Vision and Computing*, 2017, pp. 1011–1015.
- [37] J. Jovanović, D. Gašević, S. Dawson, A. Pardo, and N. Mirriahi, "Learning analytics to unveil learning strategies in a flipped classroom," *Internet High. Educ.*, vol. 33, no. 1, pp. 74–85, 2017.
- [38] R. Lumbantoruan, R. Sitorus, N. Siagian, and C. Elsa, "Formative Assessment and Learning Analytics in Informatics Diploma Program," *Adv. Sci. Lett.*, vol. 23, no. 5, pp. 4472–4477, 2017.
- [39] A. Vahldick, A. J. Mendes, and M. J. Marcelino, "Learning analytics model in a casual serious game for computer programming learning," *Lect. Notes Inst. Comput. Sci. Soc. Telecommun. Eng. LNICST*, vol. 176, no. 6, pp. 36–44, 2017.
- [40] A. R. Cano, B. Fernández-Manjón, and Á. J. García-Tejedor, "GLAID: Designing a game learning analytics model to analyze the learning process in users with intellectual disabilities," in *International Conference on Serious Games, Interaction and Simulation*, 2017, pp. 45–52.
- [41] M. Khair, C. El Moucary, and W. Zakhem, "Creating an Educational Roadmap for Engineering Students via an Optimal and Iterative Yearly Regression Tree using Data Mining," in *Proceedings of the International Conference on Knowledge Engineering and Ontology Development*

- (KEOD), 2012, pp. 43–52.
- [42] G. Robles and J. González-Barahona, “Mining student repositories to gain learning analytics: An experience report,” in *Proceedings of Global Engineering Education Conference (EDUCON)*, 2013, pp. 1249–1254.
- [43] M. Berland, T. Martin, T. Benton, C. Petrick Smith, and D. Davis, “Using Learning Analytics to Understand the Learning Pathways of Novice Programmers,” *J. Learn. Sci.*, vol. 22, no. 4, pp. 564–599, 2013.
- [44] C. Malliarakis, M. Satratzemi, and S. Xinogalos, “Integrating learning analytics in an educational MMORPG for computer programming,” in *Proceedings - IEEE 14th International Conference on Advanced Learning Technologies (ICALT)*, 2014, pp. 233–237.
- [45] P. Blikstein, M. Worsley, C. Piech, M. Sahami, S. Cooper, and D. Koller, “Programming Pluralism: Using Learning Analytics to Detect Patterns in the Learning of Computer Programming,” *J. Learn. Sci.*, vol. 23, no. 4, pp. 561–599, 2014.
- [46] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, D. Leony, and C. Delgado Kloos, “ALAS-KA: A learning analytics extension for better understanding the learning process in the Khan Academy platform,” *Comput. Human Behav.*, vol. 47, no. 1, pp. 139–148, 2015.
- [47] M. Berland, D. Davis, and C. P. Smith, “AMOEBa: Designing for collaboration in computer science classrooms through live learning analytics,” *Int. J. Comput. Collab. Learn.*, vol. 10, no. 4, pp. 425–447, 2015.
- [48] N. Buniyamin, U. Bin Mat, and P. M. Arshad, “Educational data mining for prediction and classification of engineering students achievement,” in *IEEE 7th International Conference on Engineering Education (ICEED)*, 2015, pp. 49–53.
- [49] Y. H. Jiang, S. S. Javaad, and L. Golab, “Data mining of undergraduate course evaluations,” *Informatics Educ.*, vol. 15, no. 3, pp. 85–102, 2016.
- [50] J. M. Spector, “The potential of smart technologies for learning and instruction,” *Int. J. Smart Technol. Learn.*, vol. 1, no. 1, pp. 21–32, 2016.
- [51] B. B. Jayasingh, “A Data Mining Approach to Inquiry Based Inductive Learning Practice in Engineering Education,” in *Proceedings of IEEE 6th International Conference on Advanced Computing (IACC)*, 2016, pp. 845–850.
- [52] F. M. Almutairi, N. D. Sidiropoulos, and G. Karypis, “Context-Aware Recommendation-Based Learning Analytics Using Tensor and Coupled Matrix Factorization,” *IEEE J. Sel. Top. Signal Process.*, vol. 11, no. 5, pp. 729–741, 2017.
- [53] J. Baalsrud Hauge, M. Kalverkamp, F. Bellotti, R. Berta, A. De Gloria, and G. Barabino, “Requirements on learning analytics for facilitated and non facilitated games,” in *IEEE Global Engineering Education Conference (EDUCON)*, 2014, no. April, pp. 1126–1132.
- [54] A. Vozniuk, M. J. Rodriguez-Triana, A. Holzer, S. Govaerts, D. Sandoz, and D. Gillet, “Contextual learning analytics apps to create awareness in blended inquiry learning,” in *Proceedings of International Conference on Information Technology Based Higher Education and Training (ITHET)*, 2015, pp. 1–5.
- [55] L. Echeverria, A. Benitez, S. Buendia, R. Cobos, and M. Morales, “Using a learning analytics manager for monitoring of the collaborative learning activities and students’ motivation into the Moodle system,” in *Proceedings of IEEE 11th Colombian Computing Conference (CCC)*, 2016, pp. 1–8.
- [56] F. Buendía and J. Benlloch-Dualde, “Learning Analytics Sources,” in *Proceedings of International Symposium on Computers in Education (SIEE)*, 2016, pp. 1–6.
- [57] D. Buenaño-Fernández and S. Lujan-Mora, “Comparison of applications for educational data mining in Engineering Education,” in *IEEE World Engineering Education Conference (EDUNINE)*, 2017, pp. 81–85.
- [58] C. Fernandez-Medina, J. R. Pérez-Pérez, V. M. Álvarez-García, and M. del P. Paule-Ruiz, “Assistance in computer programming learning using educational data mining and learning analytics,” in *Proceedings of the 18th ACM conference on Innovation and technology in computer science education - ITiCSE*, 2013, pp. 237–242.
- [59] J. A. Ruipérez-Valiente, P. J. Muñoz-Merino, and C. D. Kloos, “An architecture for extending the learning analytics support in the Khan Academy framework,” in *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality (TEEM)*, 2013, pp. 277–284.
- [60] E. Aguiar, G. A. Ambrose, N. V. Chawla, V. Goodrich, and J. Brockman, “Engagement vs Performance: Using Electronic Portfolios to Predict First Semester Engineering Student Persistence,” in *Proceedings of International Conference on Learning Analytics and Knowledge (LAK)*, 2014, pp. 103–112.
- [61] M. Worsley and P. Blikstein, “Deciphering the Practices and Affordances of Different Reasoning Strategies through Multimodal Learning Analytics,” in *Proceedings of the 2014 ACM workshop on Multimodal Learning Analytics Workshop and Grand Challenge (MLA)*, 2014, pp. 21–27.
- [62] M. Worsley and P. Blikstein, “Leveraging multimodal learning analytics to differentiate student learning strategies,” in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK)*, 2015, pp. 360–367.
- [63] R. Martinez-Maldonado, A. Pardo, N. Mirriahi, K. Yacef, J. Kay, and A. Clayphan, “The LATUX Workflow: Designing and Deploying Awareness Tools in Technology-enabled Learning Settings,” in *Proceedings of the Fifth International Conference on Learning Analytics And Knowledge (LAK)*, 2015, pp. 1–10.
- [64] M. Worsley, Scherer, M. S., L.P., and P. Blikstein, “Exploring Behavior Representation for Learning Analytics,” in *Proceedings of International Conference on Multimodal Interaction (ICMI)*, 2015, pp. 251–258.
- [65] I. Khan and A. Pardo, “Data2U: Scalable real time student feedback in active learning environments,” in *Proceedings of International Conference on Learning Analytics and Knowledge*, 2016, pp. 249–253.
- [66] A. Pardo, F. Han, and R. A. Ellis, “Exploring the relation between self-regulation, online activities, and academic performance,” in *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, 2016, pp. 422–429.
- [67] Y. Altujjar, W. Altamimi, I. Al-Turaiki, and M. Al-Razgan, “Predicting Critical Courses Affecting Students Performance: A Case Study,” *Procedia Comput. Sci.*, vol. 82, no. March, pp. 65–71, 2016.
- [68] M. A. Conde, R. Colomo-Palacios, F. J. García-Peñalvo, and X. Larrucea, “Teamwork assessment in the educational web of data: A learning analytics approach towards ISO 10018,” *Telemat. Informatics*, 2017.
- [69] E. B. Costa, B. Fonseca, M. A. Santana, F. F. de Araújo, and J. Rego, “Evaluating the effectiveness of educational data mining techniques for early prediction of students’ academic failure in introductory programming courses,” *Comput. Human Behav.*, vol. 73, pp. 247–256, 2017.
- [70] R. Asif, A. Merceron, S. A. Ali, and N. G. Haider, “Analyzing undergraduate students’ performance using educational data mining,” *Comput. Educ.*, vol. 113, pp. 177–194, 2017.
- [71] A. Krumm, R. Waddington, S. Lonn, and S. Teasley, “Increasing Academic Success in Undergraduate Engineering Education using Learning Analytics: A Design-Based Research Project,” in *Proceedings of annual meeting of the American Educational Research Association*, 2012, pp. 1–23.
- [72] J. S. Gomes, M. Yassine, M. Worsley, and P. Blikstein, “Analysing Engineering Expertise of High School Students Using Eye Tracking and Multimodal Learning Analytics,” in *Proceedings of the 6th International Conference on Educational Data Mining (EDM)*, 2013, pp. 375–377.
- [73] K. S. Sahedani and P. B. S. Reddy, “A Review : Mining Educational Data to Forecast Failure of Engineering Students,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 3, no. 12, pp. 628–635, 2013.
- [74] M. A. Conde, F. J. Garcia-Penalvo, D. A. Gomez-Aguilar, and R. Theron, “Visual learning analytics techniques applied in software engineering

- subjects,” in *Proceedings of IEEE Frontiers in Education Conference (FIE)*, 2014, pp. 1–9.
- [75] D. B. Knight, V. Tech, U. States, and B. Novoselich, “An Investigation of First-Year Engineering Student and Instructor Perspectives of Learning Analytics Approaches Cory Brozina Youngstown State University , United States Department of Civil and Mechanical Engineering United States Military Academy , United,” *J. Learn. Anal.*, vol. 3, no. 3, pp. 215–238, 2016.
- [76] D. Boulanger, J. Seanosky, V. Kumar, Kinshuk, K. Panneerselvam, and Thamarai Selvi Somasundaram, “Smart Learning Analytics,” in *Emerging Issues in Smart Learning*, Springer, Berlin, Heidelberg, 2014, pp. 289–296.
- [77] R. S. J. D. Baker and K. Yacef, “The State of Educational Data Mining in 2009 : A Review and Future Visions,” *J. Educ. Data Min.*, vol. 1, no. 1, pp. 3–16, 2009.
- [78] T. M. Connolly, E. A. Boyle, E. Macarthur, T. Hainey, and J. M. Boyle, “A systematic literature review of empirical evidence on computer games and serious games,” *Comput. Educ.*, vol. 59, no. 2, pp. 661–686, Sep. 2012.
- [79] R. S. J. D. Baker, “Data mining for education,” *Int. Encycl. Educ.*, vol. 7, no. 1, pp. 112–118, 2010.
- [80] F. Castro, A. Vellido, Á. Nebot, and F. Mugica, “Applying data mining techniques to e-learning problems,” *Stud. Comput. Intell.*, vol. 62, no. 1, pp. 183–221, 2007.
- [81] D. Ben, “Big Data and analytics in higher education: Opportunities and challenges,” *Br. J. Educ. Technol.*, vol. 46, no. 5, pp. 904–920, Sep. 2015.
- [82] L. Zhang, “A Meta-analysis Method to Advance Design of Technology-Based Learning Tool: Combining Qualitative and Quantitative Research to Understand Learning in Relation to Different Technology Features,” *J. Sci. Educ. Technol.*, vol. 23, no. 1, pp. 145–159, Feb. 2014.