

ACCEPTED VERSION (WITHOUT THE JOURNAL LAYOUT).

Learning analytics and collaborative groups of learners in distance education: a systematic mapping study

Lidia M. da SILVA¹, Lucas P. S. DIAS^{1,*}, Jorge L. V. BARBOSA¹,
Sandro J. RIGO¹, Julio C. S. dos ANJOS², Claudio F. R. GEYER²,
Valderi R. Q. LEITHARDT³

¹ Applied Computing Graduate Program (PPGCA), University of Vale do Rio dos Sinos (UNISINOS), Av. Unisinos 950, São Leopoldo, Rio Grande do Sul, Brazil

² Institute of Informatics (INF), Federal University of Rio Grande do Sul (UFRGS), Av. Bento Gonçalves, 9500, Porto Alegre, Rio Grande do Sul, Brazil

³ Research Center for Endogenous Resources Valorization (VALORIZA), Polytechnic Institute of Portalegre, Praça do Município, 11, Portalegre, Portugal

e-mail: lidiasilva@edu.unisinos.br, lucaspfsd@gmail.com, jbarbosa@unisinos.br, rigo@unisinos.br, jcsanjos@inf.ufrgs.br, geyer@inf.ufrgs.br, valderi@ipportalegre.pt

Received: January 2021

Abstract. Advances in information and communication technologies have contributed to the increasing use of virtual learning environments as support tools in teaching and learning processes. Virtual platforms generate a large volume of educational data, and the analysis of this data allows useful information discoveries to improve learning and assist institutions in reducing disqualifications and dropouts in distance education courses. This article presents the results of a systematic mapping study aiming to identify how educational data mining, learning analytics, and collaborative groups have been applied in distance education environments. Articles were searched from 2010 to June 2020, initially resulting in 55,832 works. The selection of 51 articles for complete reading in order to answer the research questions considered a group of inclusion and exclusion criteria. Main results indicated that 53% of articles (27/51) offered intelligent services in the field of distance education, 47% (24/51) applied methods and analysis techniques in distance education environments, 21% (11/51) applied methods and analysis techniques focused on virtual learning environments logs, and 5% (3/51) presented intelligent collaborative services for identification and creation of groups. This article also identified research interest clusters with highlights for the terms recommendation systems, data analysis, e-learning, educational data mining, e-learning platform and learning management system.

Key words: Collaborative Groups, Learning Analytics, Virtual Learning Environments, E-learning.

*Corresponding author.

1. Introduction

Contemporary society is undergoing significant changes, especially the emergence of digital technologies that permeate and affect different social instances, such as the educational sphere. As a result, education has been provoked to rethink its methodologies and give new meaning to its practices by the connective use of multiple technologies in school environments, as subsidiary resources for teaching-learning processes in the digital age.

As indicated by [Brown \(2011\)](#), the advance of distance learning is characterized by the technological contribution available in Virtual Learning Environments (VLE), also known as Learning Management Systems (LMS). The LMS enhances the dynamism, proactivity, and personalization of the teaching-learning process, allowing the interaction between the agents of the process and the interactive use of the available study material.

In general, the number of connected devices showed exponential growth. If the focus is placed on access to VLE, this phenomenon also applies, generating large volumes of data as discussed by [Cambruzzi *et al.* \(2015\)](#). The data generated by the VLE can be used to analyze the behavior and predict students' performance and obtain improvements in the educational process. In addition to enabling interaction and interactivity, virtual environments provide a large amount of information that can be considered valuable to analyze behavior and predict student performance, therefore improving the educational process using appropriate data analysis techniques.

Learning Analytics (LA) is defined by [Chatti *et al.* \(2012\)](#) as collection, analysis, and use of large amounts of data and information from students to improve the understanding of their behaviors and contexts, and improve learning results, increasing the efficiency and effectiveness of the institution. For [Ferguson *et al.* \(2016\)](#), LA consists of collecting and analyzing user data associated with student learning. In turn, [Long and Siemens \(2014\)](#) indicate that LA is defined as the measurement, collection, analysis, and reporting of data on students and their learning contexts in order to understand and optimize learning and the environment in which it occurs.

According to [McAfee and Brynjolfsson \(2012\)](#), computer systems used to manage students in schools and universities worldwide are increasing. The data generated by the use of systems such as Enterprise Resource Planning (ERP) and LMS have rich and useful information that can be used strategically to support student behavior diagnosis.

According to [Siemens \(2013\)](#), LA's use allows dealing with aspects such as visualization of grades and student interaction, generation of behavior patterns and creation of alternative support to learning activities. The goal of LA is to observe and understand learning behaviors in order to enable appropriate interventions, and these interventions can be performed through intelligent systems. According to [Brown \(2011\)](#) the reports generated by LA can be useful for instructors (student activities and progress), students (feedback on their progress), and administrators (for example, possible course aggregations and course progress information).

This article presents the results of a systematic mapping study that aims to understand how LA and intelligent services have been applied in distance education environments. The study has a specific focus on how learning analytics can organize collaborative groups

of learners to enhance learning. The research was carried out in five databases resulting in 55,832 articles. After applications of inclusion and exclusion criteria, 51 articles were selected for complete reading and analysis in order to answer the research questions outlined in the methodological process.

The article is structured in four sections. The first section contextualized the theme and presented the current scenario. In turn, section 2 defines the research methodology. Section 3 comments on the results for each search question. Section 4 presents the final considerations and, finally, the references are listed.

2. Methodological Procedures

Systematic mapping study was used as methodology for developing this work because, according to [Budgen et al. \(2008\)](#) and [Petersen et al. \(2015\)](#), this methodology provides an overview of the studies and their results. The mapping presents the following steps in its execution:

- a. Definition of research questions;
- b. Definition of the search process;
- c. Definition of inclusion and exclusion criteria;
- d. Execution of analysis and classification of results.

2.1. Research Questions

Table 1 presents the research questions associated with each category and divided into three groups: General Questions (GQ), Focal Questions (FQ), and Statistical Questions (SQ).

Table 1
Research Questions

ID	Questions	Categories
General Questions		
GQ1	What methods/techniques of data analysis have been applied in distance education environments?	Methods/techniques of data analysis
GQ2	What benefits have been obtained for students, teachers and managers through data analysis in distance education?	Benefits
Focal Questions		
FQ1	Are there methods/techniques of data analysis that have been using historical log records of students in the field of distance education?	Historical log records
FQ2	Are there intelligent services for collaborative groups of learners in the field of distance education?	Collaborative intelligent services
FQ3	What are the perceived trends?	Trends
Statistical Questions		
SQ1	In which databases were the articles published?	Databases
SQ2	Where were the works developed?	Work countries

The general questions provide insight into the methods and techniques of data analysis and benefits offered by students, teachers and managers by analyzing data on education in

distance education. Focal questions analyze methods and techniques of data analysis that use historical student records, publication trends, and collaborative intelligent services in distance education. FQ2 addresses specific services for forming and managing collaborative groups of learners. The statistical questions present the countries where the works were conducted and in which databases the articles were published.

2.2. Search terms

The study of terms for the definition of the search string was based on the words: Distance, Education, Learning, Educational, Intelligent Services, Environment, Data Analysis, Data Science and Data Mining. The defined terms were joined by the boolean expressions AND and OR and divided into three sets of interests.

Table 2 presents the major and search terms that compose the search string used to retrieve articles from databases. The major terms encompass the most relevant terms present in the research questions.

Table 2
Search string

Major Terms	Search Terms
Distance learning	(Distance learning OR Distance education OR E-Learning OR online education OR Educational technology OR virtual learning environment OR learning management system) AND
Learning Analytics	(Learning Analytics OR learning analytic OR data analysis OR data Science OR educational data mining OR learning data mining OR academic data mining OR school data mining) AND
Intelligent services	(smart services OR smart service OR smart methods OR Intelligent service OR intelligent task OR intelligent systems OR Smart processes OR Intelligent processes)

Table 3 shows the databases in which the search string was executed, and the number of articles initially found.

Table 3
Databases searched

Databases	Initial Search
ACM Digital Library	45.969
IEEE Xplore Digital Library	6.519
ScienceDirect	3.487
Springer Library	3.036
Scopus	308

2.3. Definition of inclusion and exclusion criteria

Table 4 presents the inclusion and exclusion criteria applied in the article selection process. The criteria were used to choose the studies most aligned with the research questions and also, to exclude noise generated by the search.

Table 4
Inclusion and exclusion criteria

Ref.	Criteria
Inclusion criteria (IC)	
IC1	Publications with complete content.
IC2	Publications in conferences, journals and workshops.
IC3	Publications that have methods of data analysis applied in distance education.
IC4	Publications that offer some kind of intelligent service for the area of distance education.
IC5	Publications from 2010 to June 2020.
Exclusion criteria (EC)	
EC1	Publications leading up to 2010.
EC2	Publications with language other than English.
EC3	Theses, dissertations, abstracts, books and systematic reviews.
EC4	Publications not related to the research theme.
EC5	Articles related to short courses.
EC6	Duplicate publications.

Figure 1 shows the complete filtering process. The initial filtering removed the impurities using the exclusion criteria EC1, EC2 and EC3. Then, the texts were filtered by EC4 considering title and keywords. Finally, the studies were filtered according to the abstracts using EC4 and EC5. This process resulted in 58 articles. Seven duplicate texts were excluded according to EC6, resulting in 51 works. The steps of addition by heuristics and filter by the three-pass method did not include articles, resulting the complete filtering on 51 articles that were read in full to ensure their suitability for this systematic mapping study.

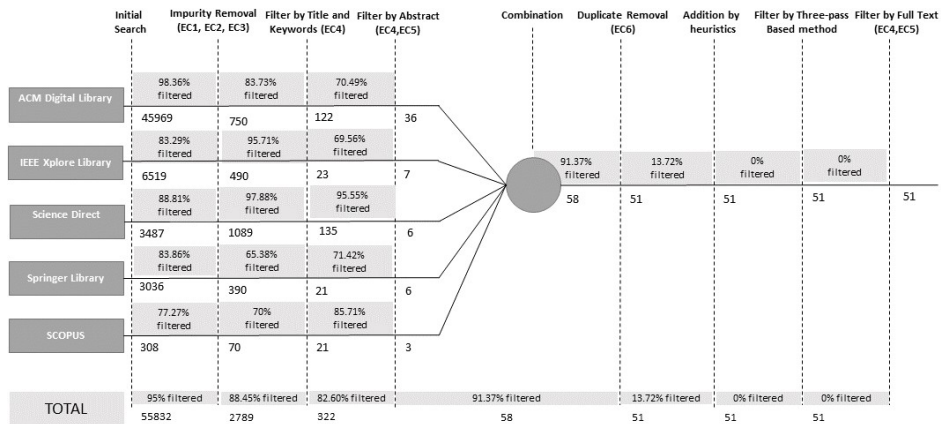


Figure 1. Initial search, application of criteria, join, removal and result.

3. Results and Discussions

Table 5 presents the list of 51 selected articles, containing reference, year of publication, countries of authors, database of the publication and summary on the focus of the article.

The first column also contains a numerical ID for the articles, which will be used throughout the text to simplify the writing and figures. The next section discusses the results of the literature review, describing the articles that answered the research questions according to the analysis methodology.

3.1. *GQ1 - What data analysis methods / techniques (Data Science) have been applied in distance education environments?*

Among the 51 selected articles, 47% (24/51) contain methods/techniques of data analysis applied to distance education. Olivé *et al.* (2018) proposed the creation of software structures for the development of educational data mining forecasting models (EDM) and LA capable of predicting which students are at risk of abandoning a course before its completion. These models allow educators to take appropriate intervention measures before the end of the course.

In the evaluation and use of predicted models of Supervised Learning in LMS, the main elements of the prediction models were abstracted and incorporated in an analytical structure for Moodle.

The software framework manages the complete cycle that predictive models follow until they are used in production. This includes resource calculations and labels of raw LMS database data, normalization, resource engineering, model evaluation, and ready-to-generate insights for forecast users. The software framework presented by the authors provides a foundation for the development of EDM and LA prediction models. The framework simplifies the implementation of new prediction models in online educational contexts.

Chaffai *et al.* (2017) created an architecture of the scalable ETL pipeline infrastructure, using Spark in Hadoop Yarn as their main computing mechanism that can provide new data for reactive panels in order to summarize student interactions. The system design illustrates the data pipeline from which to collect data, transform, analyze, store and deploy the model in production to report the analysis results. The authors used an approach based on Big Data to design a modern data pipeline in real time. The authors adopted the Hadoop and Spark association to build a distributed storage and computing mechanism by combining Apache Flume with Kafka to move a large amount of data from the Moodle database.

According to the authors, this solution does not affect moodle platform performance compared to existing solutions that use platform capabilities to produce statistics. The solution proposed by the authors is not limited to data structured as a MySQL database. The flexible architecture of the different layers can be adjusted to different user tracking data structures generated by different e-learning platforms. Experiments were used with the moodle virtual environment data and the results confirmed the efficiency of the proposed solution.

Table 5
List of mapped articles

ID	References/Authors	Countries	Source	Analytics Models/Decision Making Method.
1	Altaf <i>et al.</i> (2019)	Pakistan/ New Zealand/ Malaysia	ACM	Neural Networks.
2	Anaya <i>et al.</i> (2013)	Spain	ACM	Recommendation system based on an ID.
3	Balderas <i>et al.</i> (2013)	Spain	ACM	Domain-specific language.
4	Chaffai <i>et al.</i> (2017)	Morocco	ACM	Pipeline infrastructure architecture.
5	Chanaa and Faddouli (2018)	Morocco	IEEE	Custom template.
6	Chang and Chu (2010)	China	ScienceDirect	Behavioral learning model based on Learning Behavioral Petri Nets (LBPN).
7	Chen <i>et al.</i> (2020)	China / France	IEEE	Improved recommendation method.
8	Clarizia <i>et al.</i> (2018)	Italy	ACM	Mixed term chart (MGTs).
9	Dahdouh <i>et al.</i> (2018)	Morocco	SPRINGER	Learning systems.
10	Dahdouh <i>et al.</i> (2019)	Morocco	SPRINGER	System of recommendation of courses.
11	Dimopoulos <i>et al.</i> (2013)	Greece	ACM	Evaluation tool.
12	El Fouki <i>et al.</i> (2017)	Morocco	ACM	Decision-making system to help instructors.
13	El Moustamid <i>et al.</i> (2017)	Morocco	ACM	System capable of analyzing the profile of students and indexing videos on the Web.
14	Florian <i>et al.</i> (2011)	Spain/ Netherlands/ Colombia	SPRINGER	Prototype.
15	Graf <i>et al.</i> (2011)	Canada	ACM	Academic Analysis Tool (AAT).
16	Guo <i>et al.</i> (2020)	China	ACM	Learning Analytics-based method.
17	Hamada (2012)	Japan	SPRINGER	Model of an e-learning system based on Java2D Technology.
18	Huang <i>et al.</i> (2019)	China/USA	ACM	Deep Reinforcement Learning Structure.
19	Hussain <i>et al.</i> (2018)	China / India / Greece	ACM	Statistical technique of machine learning.
20	Iqbal <i>et al.</i> (2019)	England / Pakistan / Korea	IEEE	Algorithm - Kernel Context Recommendation System.
21	Islam <i>et al.</i> (2019)	Saudi Arabia	ACM	Data mining techniques such as grouping and binding.
22	Joy <i>et al.</i> (2019)	India	ACM	Ontology model.
23	Jovanović <i>et al.</i> (2020)	Serbia/ Australia/ United States	ACM	Explanatory model of student behavior.
24	Kapembe and Quenum (2019)	Namibia	ACM	Hybrid Recommendation (RS).
25	Kim and Kim (2020)	Korea	IEEE	Individualized Artificial Intelligence Tutor.
26	Kolekar <i>et al.</i> (2018)	India	ScienceDirect	Custom interface.
27	Kotsiantis <i>et al.</i> (2010)	Greece	ScienceDirect	System that combines an incremental version of Naive Bayes, the algorithms 1-NN and WINNOR.
28	Kozierkiewicz-Hetmańska and Zysk (2013)	Poland	SPRINGER	Algorithm based on the technique of optimization of ant colonies.
29	Lagman and Mansul (2017)	Philippines	ACM	Capture system.
30	Lara <i>et al.</i> (2014)	Spain	SCOPUS	Educational Data Mining System (EDMS).
31	Lavoie and Proulx (2019)	Canada	ACM	Learning management system (LMS).
32	Manhães <i>et al.</i> (2014)	Brazil	ACM	WAVE architecture.
33	Okubo <i>et al.</i> (2017)	Japan	ACM	Recurrent Neural Network (RNN).
34	Olivé <i>et al.</i> (2018)	Australia	ACM	Software frameworks.
35	Omar and Abdesselam (2017)	Algeria	ACM	Clustering algorithms applied to an e-learning platform.
36	Peñañel <i>et al.</i> (2018)	Ecuador/ Spain	ACM	Computational techniques such as text mining and sentiment analysis.
37	Qu <i>et al.</i> (2018)	China	IEEE	Student performance prediction structure.
38	Olanrewaju <i>et al.</i> (2016)	Malaysia / India	SCOPUS	Distributed Feedback Analysis Engine.
39	Rayón <i>et al.</i> (2014)	Spain	ACM	System (SCALA - Scalable Competency Assessment).
40	Ros <i>et al.</i> (2017)	Spain	ACM	Algorithm (Main Component Analysis).
41	Sharma and Ahuja (2016)	India	ACM	Semantic recommendation.
42	Sorour <i>et al.</i> (2015)	Japan	ACM	Methods based on a statistical class.
43	Spatiotis <i>et al.</i> (2018)	Greece/United Kingdom	ACM	Mining platform.
44	Thai-Nghe <i>et al.</i> (2010)	Germany	ScienceDirect	Recommendation system techniques for educational data mining.
45	Uzir <i>et al.</i> (2020)	United Kingdom/ Australia/ Serbia	ACM	Complementary techniques - hierarchical cluster agglomeration, epistemic analysis of networks and process mining.
46	Venugopalan <i>et al.</i> (2016)	India	ACM	Content-based recommendation system.
47	Wang <i>et al.</i> (2019)	China	IEEE	Intelligent teaching assistant system.
48	Yang <i>et al.</i> (2014)	China	IEEE	Pattern recognition technique.
49	Zakrzewska (2012)	Poland	SPRINGER	Agent-based recommendation system.
50	Zaoudi and Belhadaoui (2020)	Morocco	ACM	Learner Behavior Analytics model (LBA).
51	Zorrilla <i>et al.</i> (2010)	Spain	ACM	System oriented to results analysis.

Qu *et al.* (2018) also created a student performance prediction structure, which includes processing using data in the data warehouse. They proposed a layer-supervised multilayer perceptron (MLP) method to predict student performance and the oversights provided to each corresponding hidden layer of the MLP to improve student performance. Student behavior patterns, including previous course grades and web records, were considered to find the relationship between behaviors and student performance.

The projected framework includes data processing and student performance, explores behaviors, and builds multi-level ratings to improve student performance prediction. The experiment involved 455 students from a given school. The work compared the proposed method with four algorithms, being Naive Bayes (NB), Logistic Regression (LR), Support Vector Machine (SVM), and Perceptron Multi-Layer (MLP). The result showed that the structure designed by the authors obtained the best performance.

Zorrilla *et al.* (2010) presented a system oriented to the analysis of results for different e-learning platforms. The system was designed based on a modular architecture to perform the pre-processing tasks related to the application of data mining algorithms and to store this data in the data warehouse database. The authors made use of three open-source data mining software packages, being RapidMiner, Weka and Keel.

In this work the authors proposed a decision-making system which didn't require prior knowledge of data mining. The system prepared the entry file according to a model or standard, interpreted the results, and adopted the necessary educational measures, such as new activities, new content, new information, debates, among others, thus promoting improvements in the teaching and learning process.

The article presents two main challenges, first, determining the input variables, the technique, and the parameters with which to perform the algorithms to answer the teachers' questions; and, secondly, the work defines a graphical interface that allows instructors to interpret the results easily.

According to Ros *et al.* (2017), by analyzing the different parameters recorded by the logging system and the structure of the content in Moodle for an online course, it is possible to classify them into different learning traits. These characteristics allow users to analyze the students' learning behavior and classify them to understand how they are learning, how to improve the structure of the content and the students' grades. Therefore, helping students to minimize inappropriate practices, detect abandonment, and the loneliness of the student who takes a distance course.

Ros *et al.* (2017) implemented Principal Component Analysis (PCA) algorithm to reduce the data parameters and study the most representative ones in the classification of the students' behavior, thus reducing the dimensionality of the data sets. The authors performed the classification of the learning behavior of the students by proposing five different traits: general course information, learning resources, evaluations, interaction tools and implications of the students.

Sorour *et al.* (2015) proposed new methods based on a latent statistical class for the task of forecasting the student's grade. The methods convert student comments using latent semantic analysis (LSA) and probabilistic latent semantic analysis (PLSA). They generate prediction models using support vector machine and artificial neural network to predict

students' final grades. The authors applied text mining techniques on comments in a course with 15 classes to predict students' grades.

The work used LSA and PLSA models to understand the students' attitudes and learning situations. LSA constructs a conceptual vector space in which each comment is represented as a vector in space. This not only greatly reduces dimensionality, but also reveals the associative relationship between comments. The PLSA finds aspects of the words known as topics that can deeply distinguish comments with different meanings.

The authors presented the PLSA to improve the results of the LSA. The data from the comments analyzed by the PLSA have benefits for the evaluation of students in fields of education, such as understanding the behavior and situations of students, focusing on topics with different meanings, topics that may reflect students' learning attitudes, understanding of subjects, learning difficulties and learning activities of students in each class.

The model was constructed from the feedback data in one class and then as a test, the model was applied to the feedback data of another class and finally, the predicted score was compared with the notes in the corresponding original data, repeating the procedure in each lesson and calculating the results. This approach significantly improved the overall results of forecast accuracy compared to the LSA and PLSA models, then ANN (neural and artificial networks) and SVM (support vector machine) techniques were used to build models for predicting students' final grades. After testing both methods, the results were compared and showed that the methods used by the authors can accurately predict the students' grades based on feedback data.

Peñafiel *et al.* (2018) used an application of data mining based on computational techniques such as text mining and sentiment analysis, with the aim of evaluating the open questions of online surveys conducted by university professors. The results provided relevant information concerning the time that teachers use when incorporating online platforms in the process of teaching-learning, as well as the teacher's acceptance or rejection of the use of these tools.

The data used for this study correspond to the research carried out in the learning platform of the National Polytechnic School, Higher Education Institution of South America from September 2014 to January 2015, applied to teachers to make a diagnosis of the perception of the use of virtual classrooms as a teaching tool in the classroom. The methodology proposed by the authors can be used to evaluate open or optional questions in the educational context, such as support for traditional evaluation methods. Consequently, they can be used to make decisions that should help improve the teaching and learning process.

Clarizia *et al.* (2018) used the mixed term graph (MGTs), obtained by the latent directory allocation approach, as a tool for the classification of feelings. The method is based on the construction of reference MGTs from documents labeled according to students' feelings. The proposed method was applied in e-learning to measure the classroom's mood in relation to topics, allowing teachers the option to better adjust their teaching approach. The proposed method was tested in real cases with effective results.

Spatiotis *et al.* (2018) presented an opinion mining platform capable of classifying the classes participants' opinions according to their polarity and analyzing them to contribute

to improving the teaching procedure. The data used in this work are a set of opinions collected from questionnaires of more than 2,600 teachers who participated in e-learning courses through the HEP system.

Islam *et al.* (2019) explored the various aspects of student interaction data using data mining techniques such as clustering and association to identify relevant behavior patterns. The research used a set of data generated by the Blackboard learning management system. The data covered the online activities of students of EAD courses from January 9 to november 31, 2016.

The study used k-averaging grouping, association rule mining, Map-Reduce, IQR, and other algorithms to prepare data and identify multiple online profiles of students who exhibited similar patterns in learning activities and strategies. A structure was also proposed to analyze data based on R and Hadoop platforms to correlate online profiles with student performance.

Based on the identified online profiles, students' involvement can be identified when they change from one profile to another during the academic period. The proposed analytical framework can help to further outline online profiles at more granular levels of activity that are associated with student grades. Consequently, providing a basis for predicting a student's expected final academic performance.

Hussain *et al.* (2018) applied statistical machine learning techniques for predictive analysis. They proposed a model which predicts the outcome or behavior based on data from LMS moodle records, identifying under performing and inactive students, allowing the instructor to make decisions before the final exam. In this study, the dataset used for the experiments was obtained from the anonymous database "Learn Moodle".

The data were collected in the sessions of the free online course designed for educational purposes in any educational institution from August 7 to September 4, 2016, taught by Moodle PTY LTD. In this total of online courses, 6119 were enrolled and 2 instructors facilitated the students in the online course. The data refer to the participation of the students in a discussion forum, filling out questionnaires, taking the exam, participating in the workshop, and holding events during the course.

The authors concluded that the classification technique of the diffuse disordered rule induction algorithm (FURIA) achieves high accuracy in the detection of inactive students, it also predicts the different categories of the student during the Moodle course and the K-means cluster is also able to group inactive and active users and users with unsatisfactory performance. The result of the experiment demonstrates that the proposed system can be easily integrated into the Moodle system to send alerts to inactive and low-performance students during the completion of the course and build an efficient educational environment for students.

The aim of Altaf *et al.* (2019) research was to assess whether neural networks can be used to predict student performance based on data from a Campus Management System (CMS) log file. The applicability of neural networks was assessed through two case studies, comparing the predictive performance in the data set obtained from the log records of 900 students divided into 10 courses, between the years 2016 and 2017, that used the virtual Moodle environment.

The resources used for the training originated from the LMS data acquired in the development of each course and from data such as the grades obtained in the course assignments and questionnaires. The study aimed to assess the extent to which Neural Networks can predict student performance and assess the need for academic assistance. The authors concluded that when considering performance indicators, the neural network is an excellent classifier and can be used to predict whether the student needs academic help.

Olanrewaju *et al.* (2016) introduced a new approach called Feedback Analysis Mechanism distributed (DFAM) for Big Data based on knowledge that can be applicable to the improvement of the concepts associated with the e-learning mechanisms on the cloud platform. The proposed system design perspective also included the analysis of feedback of different types of feelings provided by the respective users of the proposed system. The proposed DFAM design principle included collaborative and distributed characteristics where collaborative characteristics signify the evolution of the transmission and processing of a huge flow of heterogeneous data, such as structured, unstructured, and semi-structured data.

DFAM's planned performance evaluation adopted experiential characters. In this study, a complete collaborative interaction application was designed. The user profile is made based on the collected user data and then a semantic analysis of the educational data from the necessary discussion forum was evaluated. Experimental prototyping uses an algorithmic approach to mining and motto modules that were designed in DFAM. According to the authors, the proposed study will be useful for the continuation of research in Data Mining and Big Data analytics.

Yang *et al.* (2014) proposed a learning style prediction method based on a pattern recognition technique, which works as a middleware. That technique can be applied to other intelligent tutoring systems, while it can process topic dependent data for tracking and update the learning style results recursively.

The forecasting process was divided into two steps. The first step identified the main dimension of the learning style, which is used to turn a multilabel classification problem into a single label classification problem. The second step used continuously updated learning information to classify new students into three groups based on the dimension of the selected learning style. The concepts and methodology used in the development of the mathematical model can also be applied to other approaches to the learning style and other intelligent learning systems after some modifications. The method predicts learning styles by observing critical learning behaviors. The experimental evaluations demonstrated the efficacy of this prediction method.

Kotsiantis *et al.* (2010) proposed a system that combines an incremental version of Naive Bayes, 1-NN and WINNOW algorithms, using the voting methodology to predict student performance in a distance learning system. With the help of the proposed technique, tutors are able to know which of their students will complete a module or course accurately.

The application of the technique proposed by the authors in predicting the performance of students proved useful in identifying poor performances and it may allow tutors to take preventive measures at an earlier stage, even at the beginning of the school year, in order

to provide additional assistance to risk groups. According to the authors, the probability of a more accurate diagnosis of students' performance increases as new curriculum data is entered during the school year, offering tutors results that are more effective.

Chang and Chu (2010) presented a behavioral learning model based on Learning Behavioral Petri Nets (LBPN) to simulate a situation in which students participate in an e-learning course and then generate their behavioral patterns. The results can be used to recommend appropriate learning content for students automatically and efficiently. Behavioral patterns generated based on the LBPN-based behavioral learning model were compared with actual data collected from 117 elementary school students, and the compared results confirmed that LBPN facilitates research for web-based learning environments.

Omar and Abdesselam (2017) carried out studies with clustering algorithms applied to an e-learning platform. The algorithms were applied to the interaction traces in log files extracted from the Bechar University e-learning platform. Experiments were carried out with some classification algorithms, to test and compare their performance. Three algorithms that form the two large clustering classification families (partitioning, hierarchical) were used: K-means, CLARA and BIRCH. The author's work consisted of developing a comparative survey between automatic classification algorithms, where several concepts related to data mining and the KDD process were defined and the execution of some data mining algorithms. Each algorithm is adapted to a specific context so that it can succeed in one context and fail in another. According to the authors, the choice of the right algorithm is based on data and needs and, among the classification algorithms used, k-means proved very effective since properly configured.

Okubo *et al.* (2017) used a method to predict final grades of students through a Recurrent Neural Network (RNN) from the data stored in educational systems. The data represented the learning activities of students who used the learning management system, the electronic portfolio system and the e-book system. The authors applied this method to the registration data of 108 students of the "Information Science" course, which began in April 2016. Students were required each week to submit a report, answer a questionnaire, write a logbook for a class, and read slides and review using the three systems.

The tests performed by the authors showed that the accuracy of the prediction by the RNN is higher than 90% using the registration data until the 6th week and that, compared with the multiple regression analysis, the RNN is effective for the early prediction of the final notes.

Graf *et al.* (2011) presented an Academic Analysis Tool (AAT) developed in the Moodle Analytics project that incorporates functionality to analyze data related to students' behavior in the learning systems. AAT is a software application that allows users to access and analyze student behavior data in learning systems, also it allows users to extract detailed information about how students interact and learn in online courses using a learning system. The app is primarily intended for learning designers who want to get feedback on how students use and learn in courses, but can also be used by teachers.

This tool can provide valuable information about students' learning processes, enabling the identification of difficult or inappropriate learning materials and therefore can contribute significantly to the design of student support activities and resources.

The tool allows users to run predefined and customized queries on any learning system that stores their data in a SQL-accessible database, it also allows users to progressively enhance the tool's analytical capabilities with a simple-to-use graphical user interface. It is possible to generate automated interventions to increase student retention, motivation, and/or learning, and generate customized dashboards to share progress information with tutors and students, thus meeting the institutional objectives of quality and access.

Rayón *et al.* (2014) designed and developed a system Scalable Competence Assessment (SCALA) through a Learning Analysis approach, which integrates how the user interacts with resources and how students and teachers interact with each other. The system tracks the data to support competency assessment and the results are obtained by applying clustering and association rules mining algorithms. SCALA is an integrated, extensible web platform to support competencies assessment that visualizes learning metrics in a dashboard and extracts data using analytics techniques for discovering student patterns and metric relationships in web-based educational systems.

The experiment carried out with SCALA presented to the professors a panel with enriched headings of combined data sets obtained from six learning evaluation activities, carried out with a group of 28 students of the 2013/14 school course of the Engineering Degree Course of the University of Deusto working on the competence of teamwork. The authors also showed how the discovery of patterns through different data mining algorithms and visualization techniques suggests a simple pedagogical policy.

Uzir *et al.* (2020) presented a new learning analysis methodology that combines three complementary techniques - agglomerative hierarchical cluster, epistemic network analysis and process mining. The methodology allows the identification and interpretation of self-regulated learning in terms of the use of learning strategies.

This new methodology allows new insights into learning strategies, studying the frequency, strength of the connections between ordering and the execution time of time management, and learning tactics. The authors conducted an investigation on a first-year basic studies course at an Australian university. Tracking data were collected from students enrolled in 2017 and 2018. The number of students enrolled in 2017 and 2018 was 250 (124 women, 107 men, 19 others) and 232 (131 women, 79 men, 22 others), respectively. The course lasted 13 weeks and included 12 course topics.

This investigation provided empirical evidence and contributed to the understanding of the diversity of strategies adopted by students during the study in a virtual environment. The research showed the importance of various time management and learning tactics to promote effective learning strategies to improve self-regulation and academic performance. From an educator's perspective, this study can inform productive educational practices to help students succeed in combined and online learning. It can support educators in encouraging the effective use of learning strategies by making the necessary modifications to their teaching approach and/or planning actionable feedback interventions.

Guo *et al.* (2020) presented an LA-based method, which has a structure designed and implemented to meet the problem of predicting students' attitude towards mixed learning classes. The authors introduced sentiment analysis to transform textual data into numerical sentiment scores. They also made comparisons between several typical classification and

regression algorithms that showed better forecasting performance, through experiments in a combined C++ programming class, which featured an online blended learning platform, the prediction for the students' attitude was verified as feasible.

From the comparison between several typical classifications and regression algorithms, the Support Vector Machine classifier and the Support Vector Regression regressor showed better prediction performance. Besides, the importance of sentiment analysis for textual data was also significant for the forecast objective.

Table 6
List of articles with methods/techniques of data analysis

References/Authors	Data analysis methods / techniques
Altaf <i>et al.</i> (2019)	Neural Networks to predict student performance based on data from the log file.
Chaffai <i>et al.</i> (2017)	Scalable ETL pipeline infrastructure architecture.
Chang and Chu (2010)	Learning behavioral model based on LBPN (learning behavioral Petri Nets).
Clarizia <i>et al.</i> (2018)	Mixed graph of terms, obtained by using the latent directory allocation approach, as a tool for the classification of feelings.
Graf <i>et al.</i> (2011)	Academic analysis tool developed in the Moodle Analytics project.
Guo <i>et al.</i> (2020)	LA-based method, which has a structure designed and implemented to meet the problem of predicting the attitude of student.
Hussain <i>et al.</i> (2018)	Machine Learning statistical technique using predictive analysis.
Islam <i>et al.</i> (2019)	Data mining techniques such as clustering and association.
Jovanović <i>et al.</i> (2020)	Explanatory model of student behavior, identifying patterns in online activity.
Kotsiantis <i>et al.</i> (2010)	System that combines an incremental version of Naive Bayes, 1-NN and WINNOW algorithms, using the voting methodology.
Lara <i>et al.</i> (2014)	Educational Data Mining System.
Okubo <i>et al.</i> (2017)	Method for predicting final grades of students by an RNN from the registration data stored in educational systems.
Olivé <i>et al.</i> (2018)	Software structures for developing educational data mining and LA forecasting models.
Omar and Abdesselam (2017)	Clustering algorithms applied to an e-learning platform.
Peñafiel <i>et al.</i> (2018)	Application of data mining using computational techniques, such as text mining and sentiment analysis.
Qu <i>et al.</i> (2018)	Student performance prediction framework, which includes data processing using data stored in the warehouse.
Olanrewaju <i>et al.</i> (2016)	Approach called Feedback Analysis Mechanism distributed for Big Data.
Rayón <i>et al.</i> (2014)	SCALA System - Scalable Competence Assessment through a Learning Analysis approach.
Ros <i>et al.</i> (2017)	PCA Algorithm - Principal Component Analysis.
Sorour <i>et al.</i> (2015)	Methods based on a latent statistical class for the task of forecasting the student's grade.
Spatiotis <i>et al.</i> (2018)	Opinion mining platform capable of classifying the opinions of participants in classes.
Uzir <i>et al.</i> (2020)	Learning analysis methodology that combines three complementary techniques - agglomerative hierarchical cluster, epistemic network analysis and process mining.
Yang <i>et al.</i> (2014)	Learning style prediction method based on a pattern recognition technique.
Zorrilla <i>et al.</i> (2010)	System oriented to results analysis, generic for different e-learning platforms.

[Jovanović *et al.* \(2020\)](#) aimed to establish an explanatory model of student behavior, identifying patterns in online activity, offering new opportunities to identify patterns that can be easily interpreted by instructors, resulting in opportunities for interventions that involve human judgment. The study developed used registration data stored in the Learning Management System (LMS) database of three consecutive offers from a first-year course completely online at an Australian university in 2018. The 10-week course was mandatory

for all online students, regardless of their specialization. The number of students enrolled in the 2018 course editions was 290, 389, and 390, respectively.

The proposed method and the results offered important implications for research and practice. First, it improves student-level personalized process-level feedback to inform and motivate a student’s move towards strategies that have proven successful in the context of the course. Second, teacher-directed feedback provides support for the design of instructional interventions.

Table 6 shows the results of mapped articles that presented data analysis methods and techniques applied in distance learning environments. The methods and techniques of the mapped articles are data mining, classification algorithms, association, clustering, analysis-oriented systems, software structures / architectures, computational techniques, methods based on statistical classification, explanatory models, analysis tools, artificial neural networks, big data and LA.

Figure 2 shows the distribution of the methods / techniques found in the mapped articles. The data analysis technique was the most used, being found in 15 articles, followed by data mining systems with 4 articles, data mining platform and software structure for data mining with 2 articles each and analysis tool with 1 article. Among the main technologies, the most used were software packages for data mining and neural networks respectively.

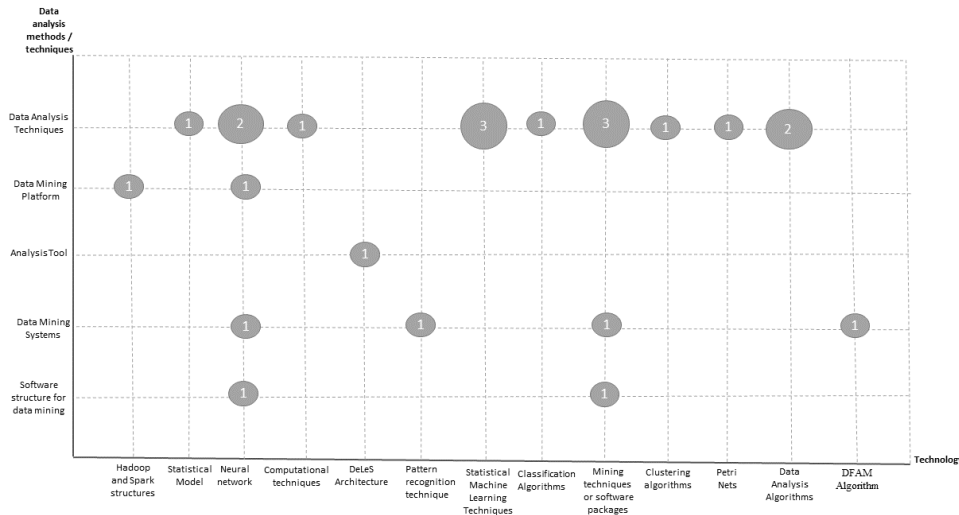


Figure 2. Methods or techniques of data analysis versus technologies used.

3.2. GQ2 - What benefits have been obtained for students, teachers and managers through data analysis in the field of distance education?

The articles indicate benefits obtained through analysis of student data stored in virtual environments or in academic management systems. The benefits can be categorized into

two aspects: 1) methods and techniques and 2) intelligent services. The benefits categorized as methods and techniques are: software frameworks for developing prediction models (Olivé et al., 2018), educational data mining system (Lara et al., 2014), pipeline infrastructure (Chaffai et al., 2017), student performance prediction structure (Qu et al., 2018), system oriented to analysis of results (Zorrilla et al., 2010), algorithms (Ros et al., 2017), methods based on a latent statistical class (Sorour et al., 2015), application of data mining using computational techniques (Peñafiel et al., 2018), mixed chart of terms (Clarizia et al., 2018), opinion mining platform (Spatiotis et al., 2018), data mining techniques such as grouping and association (Islam et al., 2019), statistical machine learning technique (Hussain et al., 2018), distributed feedback analysis engine for Big Data (Altaf et al., 2019), neural networks to predict performance (Olanrewaju et al., 2016), learning style prediction method, system that uses (Yang et al., 2014), voting methodology (Kotsiantis et al., 2010), based learning behavioral model (Chang and Chu, 2010), clustering algorithms (Omar and Abdesselam, 2017), method for predicting students' final grades by an RNR (Okubo et al., 2017), academic analysis tool (Graf et al., 2011), complementary techniques - agglomerate hierarchical cluster, epistemic network analysis and process mining (Rayón et al., 2014), learning analysis methodology (Uzir et al., 2020), method based on Learning Analytics (Guo et al., 2020) and explanatory model of student behavior (Jovanović et al., 2020).

The benefits categorized as intelligent services are: learning management system (Lavoie and Proulx, 2019), semantic recommendation using ontology (Sharma and Ahuja, 2016); hybrid recommendation based on student profile (Kapembe and Quenum, 2019); deep reinforcement learning structure (Huang et al., 2019); decision-making system (El Fouki et al., 2017), content-based recommendation system (Venugopalan et al., 2016), domain-specific language (Balderas et al., 2013), WAVE architecture (Manhães et al., 2014), intelligent teaching assistant system (Wang et al., 2019), profile analysis system (El Moustamid et al., 2017), algorithm based on the technique of optimizing ant colonies (Kozierkiewicz-Hetmańska and Zyśk, 2013), prototype indicators (Florian et al., 2011), online learning systems based on big data technologies (Dahdouh et al., 2018), agent-based recommendation system, Java2D technology-based e-learning system (Hamada, 2012), ID based recommendation system (Zakrzewska, 2012; Anaya et al., 2013), capture system (Lagman and Mansul, 2017), custom model (Chanaa and Faddouli, 2018), ontology model (Joy et al., 2019), evaluation tool (Dimopoulos et al., 2013), adaptive recommendation method (Chen et al., 2020), Kernel Context Recommender System algorithm (Iqbal et al., 2019), distributed course recommendation systems (Dahdouh et al., 2019), custom user interface (Kolekar et al., 2018), recommendation system techniques for educational data mining (Thai-Nghe et al., 2010), individualized artificial intelligence tutor and LBA model (Kim and Kim, 2020) based on a system called SBAN (Zaoudi and Belhadaoui, 2020).

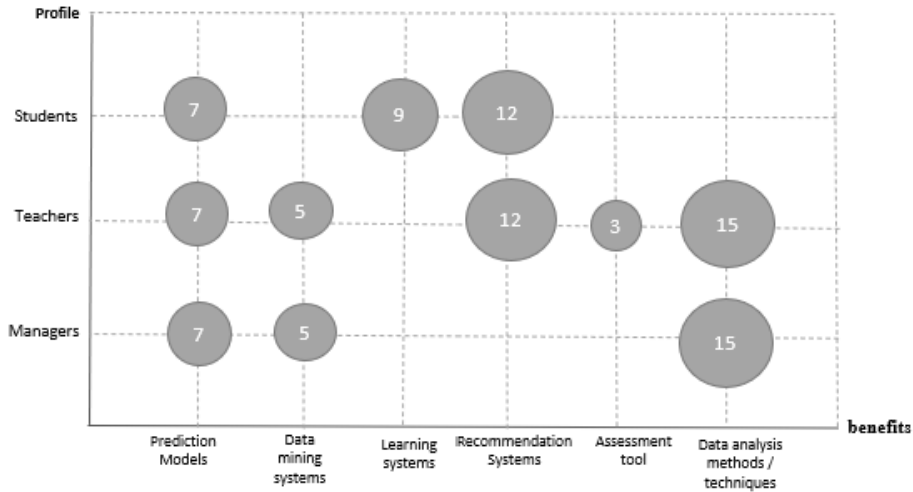


Figure 3. Data analysis benefits.

The application of methods and techniques of data analysis provide student grade prediction, behavior pattern detection, academic progress forecasting, modeling, course dropout risk prediction, also providing student performance feedback to teachers. In addition, managers and teachers can offer recommendation systems to solve problems in the teaching and learning process by providing appropriate content that meets the individual needs and preferences of students.

Figure 3 shows that the use of data analysis benefits not only students and teachers but also managers who can use analytical models to conduct more targeted campaigns, offer differentiated services according to the student's profile and the information obtained through the analysis of the data assist managers in decision making in the short, medium and long term.

3.3. FQ1 - Are there methods/techniques of analysis that have been using historical log records of students in the field of distance education?

Of the 51 mapped articles, 21% (11/51) present the application of analysis methods in the student interaction details left in the logs of the learning management systems. The authors Lara et al. (2014) used data mining techniques in moodle's internal database logs in order to discover the representative patterns of each group of students. Many data mining techniques are applied to educational data to solve problems of grouping, classification, association, time analysis, among others. The proposal aims to solve the problems mentioned globally through an integrated system, which not only solves the problems separately but is also able to identify students who are likely to drop out of the course based on the analysis of the time of the actions they perform in the classroom. The proposed system was evaluated based on real data from students of the Open University of Madrid enrolled in different courses that are part of undergraduate programs generated satisfactory results.

Islam *et al.* (2019) used the data mining techniques applied in the student interaction data to identify behavior patterns and possible key attributes to predict academic performance. The research was conducted at a Saudi educational institute, in a data set generated by the Blackboard learning management system. The data involved the activities of the students from January 9 to November 30, 2016. Grouping of k-means, mining of association rules, Map-Reduce, IQR, and other algorithms were used to prepare the data and identify various profiles of students who exhibited similar patterns in learning activities and strategies. A framework was proposed to analyze this interaction data based on R and Hadoop platforms to correlate online profiles with student performance, and key findings identified various patterns of online user profiles based on a set of learning strategies adopted.

Hussain *et al.* (2018) applied statistical techniques to the log data in the Moodle to identify underperforming or inactive students and allow the instructor to make intelligent decisions before the final exam. The applied techniques aimed at detecting the inactive student by analyzing the interaction data with Moodle and, second, detecting the student with low performance during the course based on the student's grades and events so that the instructor can intervene in real-time. In the third moment, find the activities and events that are strongly related to excellent students and finally, find the appropriate ML algorithms to predict inactive and underperforming students.

Ataf *et al.* (2019) used artificial neural networks to predict student performance based on log data from a campus management system, containing information on 900 students in 10 courses, thus allowing to compare predictive performance between courses and assess whether predictors that identify individual courses affect performance.

Omar and Abdesselam (2017) used classification algorithms in the interaction traits with an e-learning platform and, specifically, in a log file extracted from the e-learning platform of the University of Bechar in order to test and compare the performance of the algorithms.

Graf *et al.* (2011) presented an academic analysis tool developed in Moodle analytics that incorporates functionality to analyze data related to student behavior and provide teachers and course designers with more detailed and meaningful information about student behavior and the use of learning resources in EAD courses. The tool developed incorporates functionality to access and analyze data related to student behavior in online learning systems.

Florian *et al.* (2011) developed a prototype that implements indicators as examples of analytical learning applications. The research focuses on the concept of social plans and analyzes social perspectives for accessing Moodle tracking data, analyzing the reuse of Moodle tracking data for student and group modeling, and Moodle activity logging to build advanced student models based on activities in a social context. The prototyping process indicated that Moodle activity tracking includes data on more complex social structures in virtual learning environments.

Lagman and Mansul (2017) created a system that captures each student's e-learning paths and it can determine difficult topics and subjects, in which it is essential to provide academic intervention to students, serving as a complementary educational tool to help students improve their academic performance. The developed system can capture the

learning trajectories of students. The importance of identifying students' learning paths can lead to the identification of difficult topics and subjects that are vital to students' academic counseling. With this, the necessary academic interventions can be given to students to increase the percentage of approval in the course.

[Chanaa and Faddouli \(2018\)](#) proposed a customized model to improve the course completion rate and provide appropriate content that meets the individual needs and preferences of the student. The personalized model aimed to analyze the cognitive acquisition of students throughout the learning period, combining cognition analysis with sentiment analysis and the learning style of each student. Data were collected and processed based on the tracks of the students' activities from different records of the learning management system.

[Dimopoulos et al. \(2013\)](#) presented an evaluation tool, called Enriched Learning Analytics Rubric (LAe-R), which was developed as a Moodle plug-in containing related criteria and classification levels that are associated with data extracted from student interaction analysis and learning behavior in a moodle course. The article presented by [Kolekar et al. \(2018\)](#) aimed to understand the characteristics of the students and generate the custom user interface according to their learning styles, based on analysis of the web log. The architecture consists of two important modules. The first module identifies learning styles using web log analysis, and the second module provides adaptation to the Moodle-based portal according to the learning styles identified by students, also provides content as well as a user interface tailored to each student according to their learning style.

Table 7 presents the methods/techniques found in articles that used data analysis using log history in order to identify learning behavior patterns, identify underperforming or inactive students, and intervene by offering personalized tools or content as educational support.

Table 7
List of articles that contain analysis methods using records and log history

References/Authors	Analysis Methods/Techniques
Altaf et al. (2019)	Neural networks to predict student performance based on log data.
Chanaa and Faddouli (2018)	Customized model with three main components, sentimental analysis, cognitive analysis and learning style.
Dimopoulos et al. (2013)	Evaluation tool, called Enriched Heading of Learning Analytics (LAe-R).
Florian et al. (2011)	Prototype that implements indicators as examples of learning analytical.
Graf et al. (2011)	Academic analysis tool developed in moodle analytics.
Hussain et al. (2018)	Statistical techniques in log data in Moodle.
Islam et al. (2019)	Data mining techniques applied to student interaction data.
Lagman and Mansul (2017)	System that captures each student's e-learning paths.
Lara et al. (2014)	Data mining techniques in the log data in moodle's internal database.
Kolekar et al. (2018)	User interface customized according to your learning styles, based on web log analysis.
Omar and Abdesselam (2017)	Classification algorithms in the interaction traits with an e-learning platform.

Figure 4 shows the published articles that used data analysis methods based on records and log histories. Various methods or techniques were used. The methods or techniques most used by the authors were: data analysis with 37% ([Islam et al., 2019](#); [Hussain et](#)

al., 2018; Altaf *et al.*, 2019; Omar and Abdesselam, 2017), followed by learning systems with 27% (Lagman and Mansul, 2017; Chanaa and Faddouli, 2018; Kolekar *et al.*, 2018), Assessment tool with 18% (Graf *et al.*, 2011; Dimopoulos *et al.*, 2013), Recommendation System (Florian *et al.*, 2011) and Data mining Systems with 9% (Lara *et al.*, 2014).

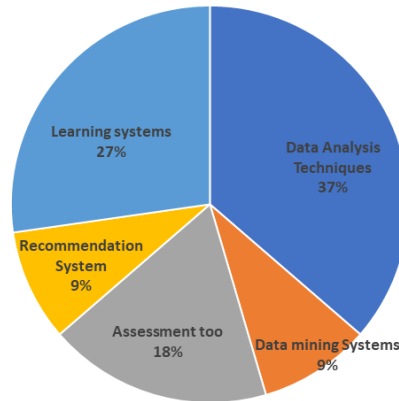


Figure 4. Articles with analysis methods using records and log histories.

3.4. FQ2. Are There Intelligent Services for Collaborative Groups of Learners in the Field of Distance Education?

Group work is an important resource for teachers to promote collaborative learning, but there are many difficulties in identifying students with similar profiles in the formation of solid groups, in the case of distance learning courses, these difficulties increase and research on systems for the analysis of student profiles to recommend collaborative study groups are still scarce. Among the 51 articles mapped, three articles address the theme approached in the focal question 2.

[Kolekar *et al.* \(2018\)](#) developed an e-learning application based on the moodle framework that captures student usage data and analyzes that data to identify group students according to the learning style, generating a personalized interface for the user. The FSLSM learning style model is a combination of three other models that combine the best of all these models (Kolb, Pask, Dunn-Dunn, Myers-Briggs, Felder-Silverman, among others). This model (FSLSM) categorizes students into a predefined set of learning style classes and has eight categories or classes of students, namely Sensing, Intuitive, Global, Sequential, Verbal, Visual, Reflective, and Active.

The group recommendation system proposed by [Zakrzewska \(2012\)](#) aims to suggest that new students choose colleagues, with whom they can learn together, using the same resources as the course. It is assumed that student groups have already been created and that they consist of learners with similar characteristics, such as cognitive styles, usability preferences or whose historical behaviors were very similar. The recommendations are

based on three collections of data: the attributes of the group members, the teaching materials equipped with flags indicating the target students and the third of a new student resource. The system performs two tasks: recommending teaching materials for groups and recommending groups for new students. In the system presented by the authors, agents are implemented to provide each new student with recommendations from classes of students of similar profiles and, consequently, indicate appropriate learning resources, or refer the student to the tutor if any group of similar peers does not exist.

Anaya *et al.* (2013) proposed an influence diagram, which includes the observable variables relevant to assess collaboration and the variable that represents whether the student collaborates or not. The main objective was to create a system that analyzes the tracking and collaboration assessments of students to identify the student's circumstances and propose a personal recommendation to the target student. The analysis provides teachers and students with a friendly explanation that can help them correct deficiencies in the collaboration process, thereby increasing their confidence and improving their learning.

The main recommendation systems have similar characteristics, namely, they are based on the analysis of student profiles and mostly offer learning content recommendation (Sharma and Ahuja, 2016; Venugopalan *et al.*, 2016; Chanaa and Faddouli, 2018; Joy *et al.*, 2019; Kim and Kim, 2020), recommendation of learning objects (Kapembe and Quenum, 2019), exercise recommendation (Huang *et al.*, 2019), course recommendation (El Moustamid *et al.*, 2017; Dahdouh *et al.*, 2019) and learning resources (Chen *et al.*, 2020).

3.5. FQ3 - What are the perceived trends?

The VOSViewer bibliometric mapping tool (van Eck and Waltman, 2018) was used to study research interest in the 51 articles mapped and published from 2010 to 2020 in the ACM Digital Library, IEEE Xplore Digital Library, Springer Library, ScienceDirect and Scopus databases. The tool was used to identify clusters that indicated shared areas of interest based on the content of the publications.

After the normalization of the terms, that is, the identification and classification of synonyms, the VOSViewer tool identified 48 items and 4 clusters among the 51 selected publications. Table 8 shows the terms and the number of occurrences in the articles.

Table 8
Top clusters, terms and number of occurrences.

Cluster	Terms	Number of occurrences
Blue	Student	80
Red	System	77
Green	Data	76
Yellow	Technique	29

Figure 5 shows the density of terms and the formation of clusters of interest, grouped by colors according to the proximity of the terms. Clusters are characterized as follows:

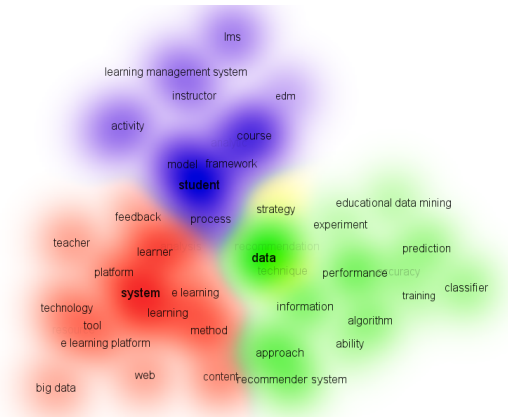


Figure 5. Cluster density.

- Blue Cluster:** This cluster consists of 13 items, and the term “Student” stood out from the others with 80 occurrences. This cluster also relates to course, model, framework, activity, instructor, process, LMS indicating relationships in the context of students and courses. This cluster also connects with other clusters as highlighted in Figure 6.
- Green Cluster:** This cluster houses 15 items. In this cluster, the highlight refers to the term “data” that was observed in 76 occurrences and relating to various algorithms. Figure 7 presents the connection of the main term with other clusters and other terms.

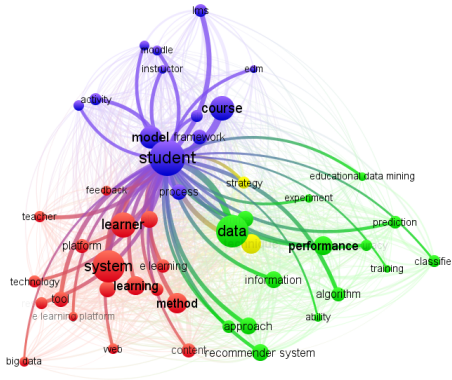


Figure 6. Blue cluster connections.

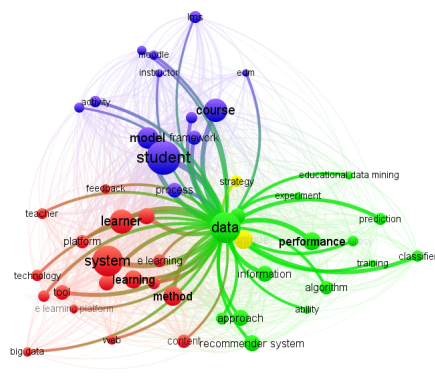


Figure 7. Green cluster connections.

- Red Cluster:** This is the largest cluster, consisting of 18 items, with a focus on “systems” with 77 occurrences, followed by other terms with smaller occurrences, such as “learner”, with 44, and “learning”, with 26. Figure 8 shows that the term “system” connects with the other clusters and with the other terms.

- **Yellow Cluster:** Smaller cluster, composed of only 2 items, and the term “technique” has 29 occurrences and the term “strategy” 11. Like others, the yellow cluster has connections to other clusters and other terms, as shown in Figure 9.

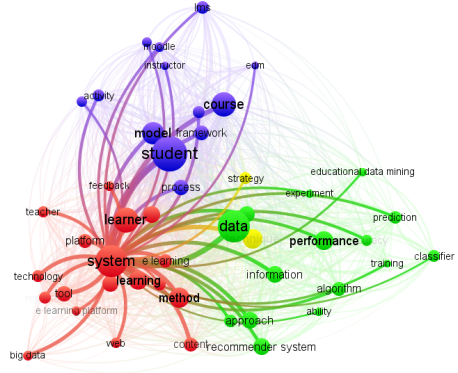


Figure 8. Red cluster connections.

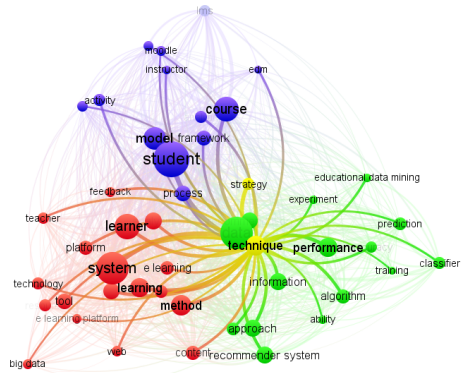


Figure 9. Yellow cluster connections.

Figure 10 provides an overview of the connections between terms from the same cluster or different clusters. Connections are determined by factors such as the occurrence of terms in articles. This model represents the overlap of chronological incidence of terms on the cluster map.

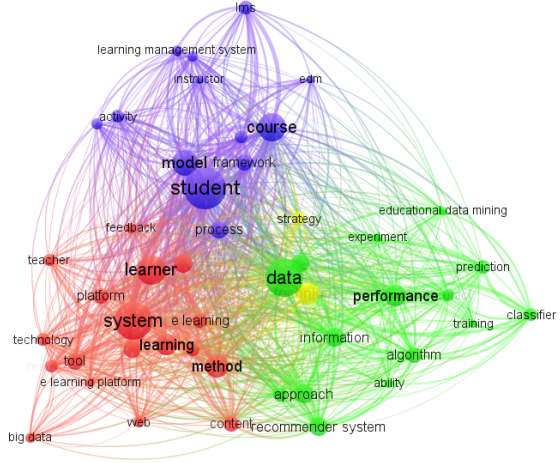


Figure 10. Overview of the relationship between terms.

The identification and presentation of clusters point to the main terms explored in the final corpus comprising 51 articles. After analyzing the 48 grouping terms, 6 terms

were identified that stand out as a search trend, being: recommendation system (Thai-Nghe *et al.*, 2010; Anaya *et al.*, 2013; Venugopalan *et al.*, 2016; Kapembe and Quenum, 2019; Iqbal *et al.*, 2019; Dahdouh *et al.*, 2019), educational data mining (Thai-Nghe *et al.*, 2010; Lara *et al.*, 2014; Olivé *et al.*, 2018), analysis (Zorrilla *et al.*, 2010; Graf *et al.*, 2011; Olanrewaju *et al.*, 2016; Ros *et al.*, 2017; Peñafiel *et al.*, 2018; Uzir *et al.*, 2020), learning management system (Okubo *et al.*, 2017; Olivé *et al.*, 2018; Chanaa and Faddouli, 2018; Lavoie and Proulx, 2019), e-learning and platform e-learning (Chang and Chu, 2010; Zorrilla *et al.*, 2010; Hamada, 2012; Anaya *et al.*, 2013; Olanrewaju *et al.*, 2016; Omar and Abdesselam, 2017; El Fouki *et al.*, 2017; Lagman and Mansul, 2017; Clarizia *et al.*, 2018; Joy *et al.*, 2019; Dahdouh *et al.*, 2019). These terms are within the red, blue and green clusters, and they have connections with each other and connections to the other clusters.

Figure 11 focuses on research questions versus the objectives of the mapped articles and the result corroborates the trends identified through the VOSViewer tool. The selected articles mainly refer to the intelligent systems of recommendations, analysis systems and the terms “learning analytics” and “e-learning”.

Figure 11 shows that intelligent services as recommendation systems are highlighted. A less explored scenario is composed of collaborative intelligent services for the identification and creation of groups. Only three articles addressed this theme, Anaya *et al.* (2013), Zakrzewska (2012) and Dahdouh *et al.* (2019).

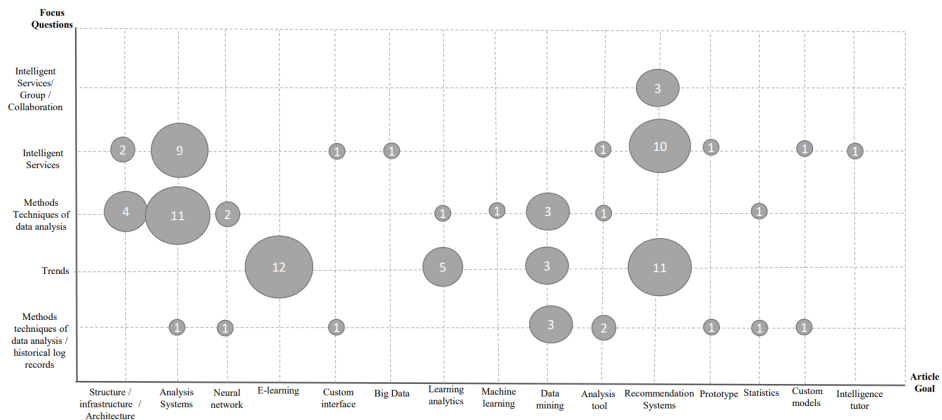


Figure 11. Research questions versus the objectives of the mapped articles.

3.6. *SQL* -In which databases were the works published?

The research carried out between 2010 and June 2020 resulted in 51 publications. Most of the publications found were in the ACM Digital Library database with 63% of the articles, followed by the IEEE Xplore Digital Library databases with 13%, Springer Library with 12%, ScienceDirect with 8%, and Scopus with only 4%.

3.7. SQ2 - Where were the works developed?

The articles were mapped according to the country where the authors' institutions are located. In this sense, the study shows that China participated in the largest number of articles with 8 followed by Morocco with 7, Spain with 6, India with 5, Greece with 4 and Japan with 3 and other countries with fewer publications, as highlighted in Figure 12.

Figure 13 presents the number of studies per year from 2010 to June 2020, emphasizing the identification of the article and the publication base. In 2010, 5 articles were published. From 2010 to 2016, there were increases and decreases in publications with no trend. But as of 2016, the increase in the number of publications is significant. The search for articles ended in June 2020 with the count of six articles which indicates a relevant number compared to previous years.

Figure 13 presents the number of studies per year from 2010 to June 2020, emphasizing the identification of the article and the publication basis.

In 2010, 5 articles were published, with a decrease in 2011 and 2012, an increase in the publication in 2013 and 2014, a significant decrease in 2015, an increase in 2016 and significant growth between 2017 and 2020, with 7 publications in 2017, 9 publications in 2018 and 2019 and 6 publications from January to June 2020.

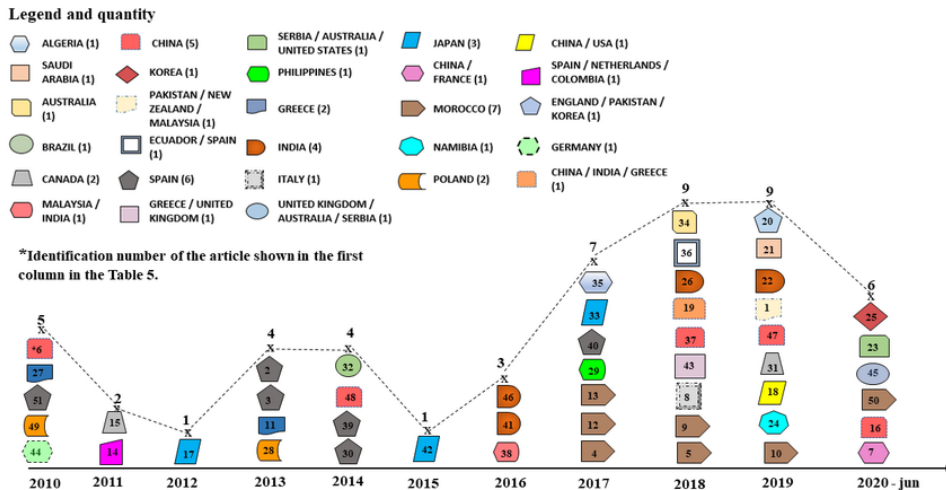


Figure 12. Publishing Countries.

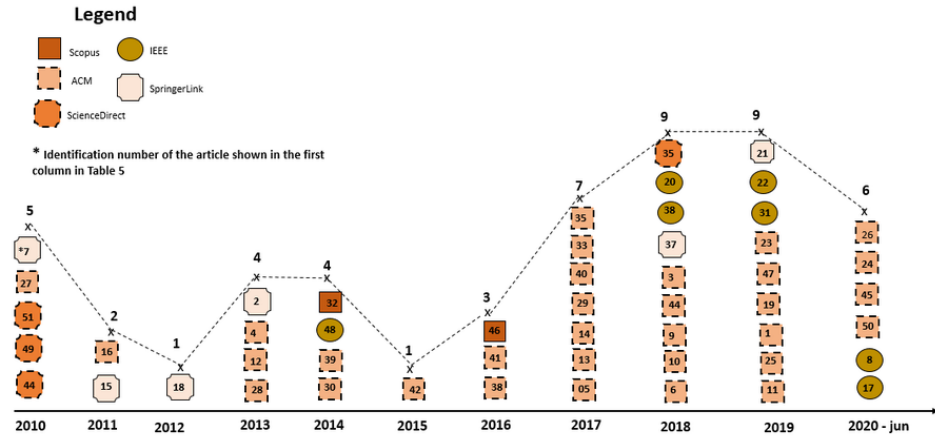


Figure 13. Publication by year and by search base.

4. Final Considerations and Future Research

The distance education has been increasing in the educational area and its use contributes to the generation of data from the interaction and learning behavior of students. The application of methods and techniques of data analysis can identify patterns of behaviors of students, allowing the prediction of academic progress and possible abandonment, and enabling teachers and managers to make decisions in the face of the problems encountered.

This work identified the current scenario in publications regarding data analysis, data mining educations, collaborative groups, LA and intelligent systems applied to distance education. A systematic mapping study was adopted to identify and analyze the articles selected in this research.

The research allowed the identification of articles that used methods and techniques of data analysis in distance education environments. Of the 51 articles selected, 47% (24/51) presented LA methods and/or techniques, algorithms, clustering techniques, association and classification, analysis-oriented systems, software structures/architectures, computational techniques, methods based on statistical classes, analysis tools, neural networks and Big Data. These techniques and models can be used in the discovery of patterns of student behaviors, allowing teachers and managers to use the information to reduce the problems faced in conducting distance education courses, such as failure and dropout.

The offer of intelligent services within the scope of the distance education was observed in 53% (27/51) of the selected articles. The main intelligent services offered are learning system, recommendation systems, systems for profile analysis, data analysis algorithms, machine learning algorithms, intelligent tutoring and recommendation systems. In this way these intelligent services promote improvements in learning and make it possible to minimize the dropout rate in distance education courses.

In 21% (11/51) of the articles, methods and techniques of data analysis applied in the log records of students in the distance education courses were observed, in order to iden-

tify learning behavior patterns, underperforming or inactive students, and intervene by offering personalized tools or content as educational support. The main techniques found are data mining techniques, statistical techniques, artificial neural network, classification algorithms, prototype that implements indicators as examples of learning analytical, custom model and sentiment analysis, cognitive analysis and learning style.

Although the research identified a variety of intelligent services, there was a gap when it came to intelligent collaborative services, among the 51 articles mapped only 5% (3/51) presented intelligent collaborative services to identify and create groups of students. The result found points to the need for a deepening in the research and development of this type of intelligent service.

The study also identified a growing interest increase in research related to the use of intelligent service and LA in distance education. This increase can be observed in the years 2017 to June 2020 with 31 articles published. The results obtained validate the perception and trends of research in recommendation systems, analysis, e-learning, educational data mining and learning management system.

As a future work, it is intended to conduct research with students and teachers in order to obtain data and opinions on the use, challenges, characteristics and potentialities of distance education. In addition, research work will be conducted to develop an intelligent recommendation system that allows the provision of individual and contingent feedback to the academic performance of the student based on the results of the analysis of this systematic mapping study and conduct a state-of-the-art study on the main causes of dropout in distance courses.

References

- Altaf, S., Soomro, W., Rawi, M.I.M. (2019). Student Performance Prediction Using Multi-Layers Artificial Neural Networks: A Case Study on Educational Data Mining. In: *Proceedings of the 2019 3rd International Conference on Information System and Data Mining*. ICISDM 2019. Association for Computing Machinery, New York, NY, USA, pp. 59–64. 9781450366359. <https://doi.org/10.1145/3325917.3325919>.
- Anaya, A.R., Luque, M., García-Saiz, T. (2013). Recommender system in collaborative learning environment using an influence diagram. *Expert Systems with Applications*, 40(18), 7193–7202. <https://doi.org/10.1016/j.eswa.2013.07.030>.
- Balderas, A., Ruiz-Rube, I., Palomo-Duarte, M., Dodero, J.M. (2013). A Generative Computer Language to Customize Online Learning Assessments. In: *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*. TEEM '13. Association for Computing Machinery, New York, NY, USA, pp. 141–147. 9781450323451. <https://doi.org/10.1145/2536536.2536558>.
- Brown, M. (2011). Learning analytics: The coming third wave. In: *Washington, DC: Learning Initiative*, Washington, DC, pp. 1–4. Available in: <https://library.educause.edu/-/media/files/library/2011/4/elib1101-pdf.pdf>, access: December 13 2020..

- Budgen, D., Turner, M., Brereton, P., Kitchenham, B. (2008). Using mapping studies. *Software engineering in Proceedings of PPIG. Lancaster University.*, 8, 195–204.
- Cambrozzi, W., Rigo, S.J., Barbosa, J.L.V. (2015). Dropout Prediction and Reduction in Distance Education Courses with the Learning Analytics Multitrail Approach. *Journal of Universal Computer Science*, 21(1), 23–47. <https://doi.org/10.3217/jucs-021-01-0023>.
- Chaffai, A., Hassouni, L., Anoun, H. (2017). E-Learning Real Time Analysis Using Large Scale Infrastructure. In: *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications. BDCA'17*. Association for Computing Machinery, New York, NY, USA, pp. 1–6. 9781450348522. <https://doi.org/10.1145/3090354.3090378>.
- Chanaa, A., Faddouli, N.E. (2018). Deep learning for a smart e-learning system. In: *2018 4th International Conference on Cloud Computing Technologies and Applications (Cloudtech), Brussels, Belgium, 2018, pp. 1-8, DOI:*, pp. 1–8. <https://doi.org/10.1109/CloudTech.2018.8713335>.
- Chang, Y.-C., Chu, C.-P. (2010). Applying learning behavioral Petri nets to the analysis of learning behavior in web-based learning environments. *Information Sciences*, 180(6), 995–1009. Special Issue on Modelling Uncertainty. <https://doi.org/doi.org/10.1016/j.ins.2009.11.022>.
- Chatti, M.A., Dyckhoff, A.L., Schroeder, U., Thüs, H. (2012). A Reference Model for Learning Analytics. *Int. J. Technol. Enhanc. Learn.*, 4(5/6), 318–331. <https://doi.org/10.1504/IJTEL.2012.051815>.
- Chen, H., Yin, C., Li, R., Rong, W., Z, X., David, B. (2020). Enhanced learning resource recommendation based on online learning style model. in *Tsinghua Science and Technology*, 25(3), 348–359. <https://doi.org/10.26599/TST.2019.9010014>.
- Clarizia, F., Colace, F., De Santo, M., Lombardi, M., Pascale, F., Pietrosanto, A. (2018). E-Learning and Sentiment Analysis: A Case Study. In: *Proceedings of the 6th International Conference on Information and Education Technology. ICIET '18*. Association for Computing Machinery, New York, NY, USA, pp. 111–118. 9781450353595. <https://doi.org/10.1145/3178158.3178181>.
- Dahdouh, K., Dakkak, A., Oughdir, L., Messaoudi, F. (2018). Big data for online learning systems. *Education and Information Technologies*, 23(6), 2783–2800. <https://doi.org/10.1007/s10639-018-9741-3>.
- Dahdouh, K., Dakkak, A., Oughdir, L., Ibriz, A. (2019). Large-scale e-learning recommender system based on Spark and Hadoop. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0169-4>.
- Dimopoulos, I., Petropoulou, O., Retalis, S. (2013). Assessing Students' Performance Using the Learning Analytics Enriched Rubrics. In: *Proceedings of the Third International Conference on Learning Analytics and Knowledge. LAK '13*. Association for Computing Machinery, New York, NY, USA, pp. 195–199. 9781450317856. <https://doi.org/10.1145/2460296.2460335>.
- El Fouki, M., Aknin, N., El. Kadiri, K.E. (2017). Intelligent Adapted E-Learning System Based on Deep Reinforcement Learning. In: *Proceedings of the 2nd International*

- Conference on Computing and Wireless Communication Systems. ICCWCS'17. Association for Computing Machinery, New York, NY, USA, pp. 1–6. 9781450353069. <https://doi.org/10.1145/3167486.3167574>.
- El Moustamid, A., En-Naimi, E., El Bouhdidi, J. (2017). Integration of Data Mining Techniques in E-Learning Systems: Clustering Profil of Lerner and Recommender Course System. In: *Proceedings of the 2nd International Conference on Big Data, Cloud and Applications*. BDCA'17. Association for Computing Machinery, New York, NY, USA, pp. 1–4. 9781450348522. <https://doi.org/10.1145/3090354.3090453>.
- Ferguson, R., Brasher, A., Clow, D., Cooper, A., Hillaire, G., Mittelmeier, J., Rienties, B., Ullmann, T., Vuorikari, R. (2016). Research Evidence on the Use of Learning Analytics-Implications for Education Policy. *Research Evidence on the Use of Learning Analytics - Implications for Education Policy*. <https://doi.org/10.2791/955210>.
- Florian, B., Glahn, C., Drachslar, H., Specht, M., Fabregat Gesa, R. (2011). Activity-Based Learner-Models for Learner Monitoring and Recommendations in Moodle. In: Kloos, C.D., Gillet, D., Crespo García, R.M., Wild, F., Wolpers, M. (Eds.), *Towards Ubiquitous Learning*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 111–124. 978-3-642-23985-4. <https://doi.org/10.1007/978-3-642-23985-4-10>.
- Graf, S., Ives, C., Rahman, N., Ferri, A. (2011). AAT: A Tool for Accessing and Analysing Students' Behaviour Data in Learning Systems. In: *Proceedings of the 1st International Conference on Learning Analytics and Knowledge*. LAK '11. Association for Computing Machinery, New York, NY, USA, pp. 174–179. 9781450309448. <https://doi.org/10.1145/2090116.2090145>.
- Guo, C., Yan, X., Li, Y. (2020). Prediction of Student Attitude towards Blended Learning Based on Sentiment Analysis. In: *Proceedings of the 2020 9th International Conference on Educational and Information Technology*. ICEIT 2020. Association for Computing Machinery, New York, NY, USA, pp. 228–233. 9781450375085. <https://doi.org/10.1145/3383923.3383930>.
- Hamada, M. (2012). Learning Style Model for e-Learning Systems. In: Huang, R., Ghorbani, A.A., Pasi, G., Yamaguchi, T., Yen, N.Y., Jin, B. (Eds.), *Active Media Technology*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 186–195. 978-3-642-35236-2.
- Huang, Z., Liu, Q., Zhai, C., Yin, Y., Chen, E., Gao, W., Hu, G. (2019). Exploring Multi-Objective Exercise Recommendations in Online Education Systems. In: *Proceedings of the 28th ACM International Conference on Information and Knowledge Management*. CIKM '19. Association for Computing Machinery, New York, NY, USA, pp. 1261–1270. 9781450369763. <https://doi.org/10.1145/3357384.3357995>.
- Hussain, M., Hussain, S., Zhang, W., Zhu, W., Theodorou, P., Abidi, S.M.R. (2018). Mining Moodle Data to Detect the Inactive and Low-Performance Students during the Moodle Course. In: *Proceedings of the 2nd International Conference on Big Data Research*. ICBDR 2018. Association for Computing Machinery, New York, NY, USA, pp. 133–140. 9781450364768. <https://doi.org/10.1145/3291801.3291828>.
- Iqbal, M., Ghazanfar, M.A., Sattar, A., Maqsood, M., Khan, S., Mehmood, I., Baik, S.W. (2019). Kernel Context Recommender System (KCR): A Scalable Context-Aware Rec-

- ommender System Algorithm. in *IEEE Access*, 7, 24719–24737. <https://doi.org/10.1109/ACCESS.2019.2897003>.
- Islam, O., Siddiqui, M., Aljohani, N.R. (2019). Identifying Online Profiles of Distance Learning Students Using Data Mining Techniques. In: *Proceedings of the 2019 The 3rd International Conference on Digital Technology in Education*. ICDTE 2019. Association for Computing Machinery, New York, NY, USA, pp. 115–120. 9781450372206. <https://doi.org/10.1145/3369199.3369249>.
- Jovanović, J., Dawson, S., Joksimović, S., Siemens, G. (2020). Supporting Actionable Intelligence: Reframing the Analysis of Observed Study Strategies. In: *Proceedings of the Tenth International Conference on Learning Analytics and Knowledge*. LAK '20. Association for Computing Machinery, New York, NY, USA, pp. 161–170. 9781450377126. <https://doi.org/10.1145/3375462.3375474>.
- Joy, J., Raj, N.S., G, R.V. (2019). An Ontology Model for Content Recommendation in Personalized Learning Environment. In: *Proceedings of the Second International Conference on Data Science, E-Learning and Information Systems*. DATA '19. Association for Computing Machinery, New York, NY, USA, pp. 1–6. 9781450372848. <https://doi.org/10.1145/3368691.3368700>.
- Kapembe, S.S., Quenum, J.G. (2019). A Personalised Hybrid Learning Object Recommender System. In: *Proceedings of the 11th International Conference on Management of Digital EcoSystems*. MEDES '19. Association for Computing Machinery, New York, NY, USA, pp. 242–249. 9781450362382. <https://doi.org/10.1145/3297662.3365810>.
- Kim, W., Kim, J. (2020). Individualized AI Tutor Based on Developmental Learning Networks. in *IEEE Access*, 8, 27927–27937. <https://doi.org/10.1109/ACCESS.2020.2972167>.
- Kolekar, S.V., Pai, R.M., Pai M.M., M. (2018). Adaptive User Interface for Moodle based E-learning System using Learning Styles. *Procedia Computer Science*, 135, 606–615. The 3rd International Conference on Computer Science and Computational Intelligence (ICCCSCI 2018) : Empowering Smart Technology in Digital Era for a Better Life. <https://doi.org/10.1016/j.procs.2018.08.226>.
- Kotsiantis, S., Patriarcheas, K., Xenos, M. (2010). A combinational incremental ensemble of classifiers as a technique for predicting students' performance in distance education. *Knowledge-Based Systems*, 23(6), 529–535. <https://doi.org/10.1016/j.knosys.2010.03.010>.
- Kozierkiewicz-Hetmańska, A., Zyśk, D. (2013). A Method for Determination of an Opening Learning Scenario in Intelligent Tutoring Systems. In: Selamat, A., Nguyen, N.T., Haron, H. (Eds.), *Intelligent Information and Database Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 129–138. 978-3-642-36543-0.
- Lagman, A.C., Mansul, D.M. (2017). Extracting Personalized Learning Path in Adaptive E-Learning Environment Using Rule Based Assessment. In: *Proceedings of the 2017 International Conference on Information Technology*. ICIT 2017. Association for Computing Machinery, New York, NY, USA, pp. 335–340. 9781450363518. <https://doi.org/10.1145/3176653.3176679>.

- Lara, J.A., Lizcano, D., Martínez, M.A., Pazos, J., Riera, T. (2014). A System for Knowledge Discovery in E-Learning Environments within the European Higher Education Area - Application to Student Data from Open University of Madrid, UDIMA. *Comput. Educ.*, 72(C), 23–36. <https://doi.org/10.1016/j.compedu.2013.10.009>.
- Lavoie, F.B., Proulx, P. (2019). A Learning Management System for Flipped Courses. In: *Proceedings of the 2019 The 3rd International Conference on Digital Technology in Education*. ICDTE 2019. Association for Computing Machinery, New York, NY, USA, pp. 73–76. 9781450372206. <https://doi.org/10.1145/3369199.3369216>.
- Long, P., Siemens, G. (2014). Penetrating the fog: Analytics in learning and education. In: *TD63 - Dossier - Learning Analytics* (Vol. 22), pp. 31–40. <https://doi.org/10.17471/2499-4324/195>.
- Manhães, L.M.B., da Cruz, S.M.S., Zimbrão, G. (2014). WAVE: An Architecture for Predicting Dropout in Undergraduate Courses Using EDM. In: *Proceedings of the 29th Annual ACM Symposium on Applied Computing*. SAC '14. Association for Computing Machinery, New York, NY, USA, pp. 243–247. 9781450324694. <https://doi.org/10.1145/2554850.2555135>.
- McAfee, A., Brynjolfsson, E. (2012). Big Data: The Management Revolution. *Harvard Business Review*, 60–68.
- Okubo, F., Yamashita, T., Shimada, A., Ogata, H. (2017). A Neural Network Approach for Students' Performance Prediction. In: *Proceedings of the Seventh International Learning Analytics and Knowledge Conference*. LAK '17. Association for Computing Machinery, New York, NY, USA, pp. 598–599. 9781450348706. <https://doi.org/10.1145/3027385.3029479>.
- Olanrewaju, R., Khan, B., Mir, R., Baba, A., Anwar, F. (2016). Dfam: A distributed feedback analysis mechanism for knowledge based educational big data. *Jurnal Teknologi*, 78, 31–38. <https://doi.org/10.11113/jt.v78.10020>.
- Olivé, D.M., Huynh, D.Q., Reynolds, M., Dougiamas, M., Wiese, D. (2018). A Supervised Learning Framework for Learning Management Systems. In: *Proceedings of the First International Conference on Data Science, E-Learning and Information Systems*. DATA '18. Association for Computing Machinery, New York, NY, USA, pp. 1–8. 9781450365369. <https://doi.org/10.1145/3279996.3280014>.
- Omar, L., Abdesselam, B. (2017). Applying Clustering Algorithms to Solve E-Learning Problems. In: *Proceedings of the International Conference on Computing for Engineering and Sciences*. ICCES '17. Association for Computing Machinery, New York, NY, USA, pp. 45–49. 9781450353090. <https://doi.org/10.1145/3129186.3129195>.
- Peñafiel, M., Vásquez, S., Vásquez, D., Zaldumbide, J., Luján-Mora, S. (2018). Data Mining and Opinion Mining: A Tool in Educational Context. In: *Proceedings of the 2018 International Conference on Mathematics and Statistics*. ICoMS 2018. Association for Computing Machinery, New York, NY, USA, pp. 74–78. 9781450365383. <https://doi.org/10.1145/3274250.3274263>.

- Petersen, K., Vakkalanka, S., Kuzniarz, L. (2015). Guidelines for conducting systematic mapping studies in software engineering: An update. *Information and Software Technology*, 64, 1–18. <https://doi.org/10.1016/j.infsof.2015.03.007>.
- Qu, S., Li, K., Zhang, S., Wang, Y. (2018). Predicting Achievement of Students in Smart Campus. *IEEE Access*, 6(18495379), 60264–60273. <https://doi.org/10.1109/ACCESS.2018.2875742..>
- Rayón, A., Guenaga, M., Núñez, A. (2014). Supporting Competency-Assessment through a Learning Analytics Approach Using Enriched Rubrics. In: *Proceedings of the Second International Conference on Technological Ecosystems for Enhancing Multiculturality*. TEEM '14. Association for Computing Machinery, New York, NY, USA, pp. 291–298. 9781450328968. <https://doi.org/10.1145/2669711.2669913>.
- Ros, S., Lázaro, J.C., Robles-Gómez, A., Caminero, A.C., Tobarra, L., Pastor, R. (2017). Analyzing Content Structure and Moodle Milestone to Classify Student Learning Behavior in a Basic Desktop Tools Course. In: *Proceedings of the 5th International Conference on Technological Ecosystems for Enhancing Multiculturality*. TEEM 2017. Association for Computing Machinery, New York, NY, USA, pp. 1–6. 9781450353861. <https://doi.org/10.1145/3144826.3145392>.
- Sharma, M., Ahuja, L. (2016). A Novel and Integrated Semantic Recommendation System for E-Learning Using Ontology. In: *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. ICTCS '16. Association for Computing Machinery, New York, NY, USA, pp. 1–5. 9781450339629. <https://doi.org/10.1145/2905055.2905110>.
- Siemens, G. (2013). Learning Analytics: The Emergence of a Discipline. *American Behavioral Scientist*, 57(10), 1380–1400. <https://doi.org/10.1177/0002764213498851>.
- Sorour, S., Goda, K., Mine, T. (2015). Correlation of Topic Model and Student Grades Using Comment Data Mining. In: *Proceedings of the 46th ACM Technical Symposium on Computer Science Education*. SIGCSE '15. Association for Computing Machinery, New York, NY, USA, pp. 441–446. 9781450329668. <https://doi.org/10.1145/2676723.2677259>.
- Spatiotis, N., Perikos, I., Mporas, I., Paraskevas, M. (2018). Evaluation of an Educational Training Platform Using Text Mining. In: *Proceedings of the 10th Hellenic Conference on Artificial Intelligence*. SETN '18. Association for Computing Machinery, New York, NY, USA, pp. 1–5. 9781450364331. <https://doi.org/10.1145/3200947.3201049>.
- Thai-Nghe, N., Drumond, L., Krohn-Grimberghe, A., Schmidt-Thieme, L. (2010). Recommender system for predicting student performance. *Procedia Computer Science*, 1(2), 2811–2819. Proceedings of the 1st Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2010). <https://doi.org/10.1016/j.procs.2010.08.006>.
- Uzir, N.A., Gašević, D., Jovanović, J., Matcha, W., Lim, L.-A., Fudge, A. (2020). Analytics of Time Management and Learning Strategies for Effective Online Learning in Blended Environments. In: *Proceedings of the Tenth International Conference on*

- Learning Analytics and Knowledge*. LAK '20. Association for Computing Machinery, New York, NY, USA, pp. 392–401. 9781450377126. <https://doi.org/10.1145/3375462.3375493>.
- van Eck, N.J., Waltman, L. (2018). VOSviewer Manual. Available in: http://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.8.pdf, access: December 13 2020..
- Venugopalan, S., Srinath, M.V., Rodrigues, P. (2016). Recommender System for E-Learning through Content and Profile Based Approach. In: *Proceedings of the Second International Conference on Information and Communication Technology for Competitive Strategies*. ICTCS '16. Association for Computing Machinery, New York, NY, USA, pp. 1–5. 9781450339629. <https://doi.org/10.1145/2905055.2905103>.
- Wang, Y., Sun, Y., Chen, Y. (2019). Design and Research of Intelligent Tutor System Based on Natural Language Processing. In: *2019 IEEE International Conference on Computer Science and Educational Informatization (CSEI)*, pp. 33–36. <https://doi.org/10.1109/CSEI47661.2019.8939031>.
- Yang, J., Huang, Z., Gao, Y., Liu, H. (2014). Dynamic Learning Style Prediction Method Based on a Pattern Recognition Technique. *Learning Technologies, IEEE Transactions on*, 7(2), 165–177. <https://doi.org/10.1109/TLT.2014.2307858>.
- Zakrzewska, D. (2012). *Building Group Recommendations in E-Learning Systems*. Springer Berlin Heidelberg, Berlin, Heidelberg, pp. 144–163. 978-3-642-32066-8. https://doi.org/10.1007/978-3-642-32066-8_7.
- Zaoudi, M., Belhadaoui, H. (2020). Adaptive E-Learning: Adaptation of Content According to the Continuous Evolution of the Learner During His Training. In: *Proceedings of the 3rd International Conference on Networking, Information Systems and Security*. NISS2020. Association for Computing Machinery, New York, NY, USA, pp. 1–6. 9781450376341. <https://doi.org/10.1145/3386723.3387890>.
- Zorrilla, M., García, D., Álvarez, E. (2010). A Decision Support System to Improve E-Learning Environments. In: *Proceedings of the 2010 EDBT/ICDT Workshops*. EDBT '10. Association for Computing Machinery, New York, NY, USA, pp. 1–8. 9781605589909. <https://doi.org/10.1145/1754239.1754252>.

Lidia M. Silva is a Ph.D. student in Applied Computing at the University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, RS, Brazil. She received a master's degree in Computer Science from the Methodist University of Piracicaba (UNIMEP), Piracicaba, SP. Specialized in Instructional Design, in 2018, and Specialized in Technological and Inclusive Training, in 2009, by the Federal Institute of Education, Science, and Technology of Mato Grosso (IFMT), Specialization in Informatics in Education, in 2010, by the Centro Universitário Cândido Rondon (UNIRONDON) and graduated in Data Processing Technology, in 1997, by the Faculdade Cândido Rondon, Cuiabá, MT. Currently, she is the Coordinator of the Execution Nucleus for the distance learning courses - Project of the State Court of Auditors (TCE) and the Federal University of Mato Grosso, and also acts as Coordinator of the Technology Courses in Analysis and Development of Systems, Technology in Computer Networks and Bachelor in Computer Science at the INVEST Faculty of Science and Technology in Cuiabá - MT. She participates in the UFMTON-LINE Project at the Federal University of Mato Grosso.

Lucas P. S. Dias received the M.Sc. degree in applied computing from the University of Vale do Rio dos Sinos (UNISINOS), São Leopoldo, Brazil, in 2018 and starts the Ph.D. degree at UNISINOS in 2020. He integrates a research team at Mobile Computing Laboratory (Mobilab/Unisinos) focused on Ubiquitous Computing and Mental Health. He also works in the industry as a Tech Lead.

Jorge L. V. Barbosa received M.Sc. and Ph.D. in computer science from the Federal University of Rio Grande do Sul, Porto Alegre, Brazil. He conducted post-doctoral studies at Sungkyunkwan University (SKKU, Suwon, South Korea) and University of California Irvine (UCI, Irvine, USA). Jorge is a full professor at the Applied Computing Graduate Program (PPGCA) of the University of Vale do Rio dos Sinos (UNISINOS), head of the university's Mobile Computing Lab (MOBILAB), and a researcher at the Brazilian Council for Scientific and Technological Development (CNPq). His research interests are Ubiquitous Computing, Ambient Intelligence, Big Data, Internet of Things (IoT), Machine Learning, Data Analysis, Games Development, Context Histories, Context Prediction mainly through similarity and pattern analysis, Applied Computing in Health, Accessibility, Learning, Industry and Agriculture.

Sandro J. Rigo is Professor/Researcher in the Applied Computing Graduate Program (UNISINOS); Dean of UNISINOS Polytechnic School. Head of the UNISINOS Graduate Degree Program in Applied Computer Science (2016–2017). Head of the UNISINOS Undergraduate Degree Program in Computer Science (2011–2015). Bsc in Computer Science at Pontifícia Universidade Católica do Rio Grande do Sul PUCRS (1990); MSc in Computer Science at Universidade Federal do Rio Grande do Sul UFRGS (1993); PhD in Computer Science at Universidade Federal do Rio Grande do Sul UFRGS (2008). Post-doctorate at Friedrich-Alexander Universität Erlangen-Nürnberg/Germany (2018). Professor and Researcher at UNISINOS University, since 1995. Research areas: Artificial Intelligence, Semantic Web, Natural Language Processing, Distance Education.

Julio C. S. ANJOS received the Ph.D. in Computer Science from the Federal University of Rio Grande do Sul (UFRGS/RS) in 2017 and M.Sc. in Computer Science in UFRGS/RS in 2012. Electrical Engineer by the Pontifical Catholic University of Rio

Grande do Sul (PUC/RS), in 1991. It is a Postdoc tenure in Computer Science at UFRGS since 2017. He is a Researcher in Big Data, MapReduce and Streaming Processing, Distributed and Heterogeneous Systems, Hybrid Infrastructures, and Big Data analytics with Machine Learning and Deep Learning.

Claudio F. R. GEYER received the Ph.D. degree in informatics from the Université de Grenoble I, in 1991, and the M.Sc. degree in computer science from UFRGS/RS, in 1986. He was a Mechanics Engineer at UFRGS/RS in 1978, where he is also a Tenured Professor in computer science. His areas of interests include pervasive and ubiquitous computing, grid and volunteer computing, scheduling and data-intensive computing, and data mining with deep learning on big data.

Valderi R. Q. LEITHARDT received the Ph.D. degree in computer science from INF-UFRGS, Brazil, in 2015. He is currently an Adjunct Professor with the Polytechnic Institute of Portalegre and a Researcher integrated with the VALORIZA Research Group, School of Technology and Management (ESTG). He is also a collaborating Researcher with the following research groups: COPELABS, Universidade Lusófona de Lisboa, Portugal, the IT Branch Covilhã, Telecommunications Institute of Portugal, Portugal, the Department of Informatics, University Beira Interior, Covilhã, Portugal, the Laboratory of Embedded and Distributed Systems, University of Vale do Itajaí (UNIVALI), Brazil, and the Expert Systems and Applications Laboratory, University of Salamanca, Spain. His research interests include distributed systems with a focus on data privacy, communication, and programming protocols, involving scenarios and applications for the Internet of Things, smart cities, big data, cloud computing, and Blockchain.