

Data mining techniques applied in educational environments:

Literature review

Andrés Villanueva Manjarres

andres13@gmail.com

Institución Universitaria Politécnico Grancolombiano

Luis Gabriel Moreno Sandoval

gabrielmoreno10@gmail.com

Institución Universitaria Politécnico Grancolombiano

Martha Janneth Salinas Suárez

majasalinas@gmail.com

Corporación Universitaria Minuto de Dios

Abstract

Educational Data Mining is an emerging discipline which seeks to develop methods to explore large amounts of data from educational settings, in order to understand students' behavior, interests and results in a better way. In recent years there have been various works related to this specialty and multiple data mining techniques derived from this to address different educational problems have been used. The aim of this paper is to present a review of the works in which data mining techniques were used to solve specific problems of education and to do a classification associated to diverse scenarios in which they have been applied.

Keywords

Educational data mining; Data mining; Clustering; Decision trees; Knowledge discovery.

I. Introduction

Due to the great contributions that technology offers to all disciplines today, it could be explained that all stages of life are changing, especially the way we communicate and the barriers that have collapsed thanks to Internet and the amount of generated and stored data in different information systems. Because of all of this, is not wrong to say that the world evolves adaptively. In this change making decisions is an activity that is not expected, especially when these decisions are determinant for the future of organizations.

In all educational environments, one of the main objectives is to ensure that learning processes allow to understand students and their learning paths. This is where Data Mining Educational or Educational Data Mining (EDM) for its acronym in English, provide a fundamental value to educational institutions and to all entities that support different processes in learning activities.

This work seeks to present in a general way the major components of EDM, as well as to illustrate the context of EDM and its growth since 2008; Moreover it will present the current state of the use of data mining techniques in the resolution or analysis of problems that occur in educational settings. We also present which of them has been used in particular domains of education; Finally, the paper analyzes how was the behavior of the different techniques of data mining in the past two decades and what are the current trends of EDM, looking for providing elements to support decision-making what techniques of data mining can be used in particular situations.

II. Data Mining

Data mining is a set of techniques and procedures that can be developed from various data sources such as data warehouses or relational databases, to flat files without formats that are made from this predictive analysis using statistical study techniques to predict or anticipate statistical measures of certainty based on existing facts. According to Fayyad, data mining "is a non-trivial process of valid identification novel, potentially useful and understandable of comprehensible patterns that are hidden in the data" (R, Sánchez-Guzmán A, & García, 2013). Data mining is also seen as the evolution of information technology (Jiawei & Kamber, 2001), this largely leveraged by the growth that internet has taken great importance in recent years (Society, 2014)(MINTIC, 2015).

Data mining is a multidisciplinary field that allows to obtain relevant information from large amounts of data at the confluence among other areas: artificial intelligence, statistics, databases and information science.

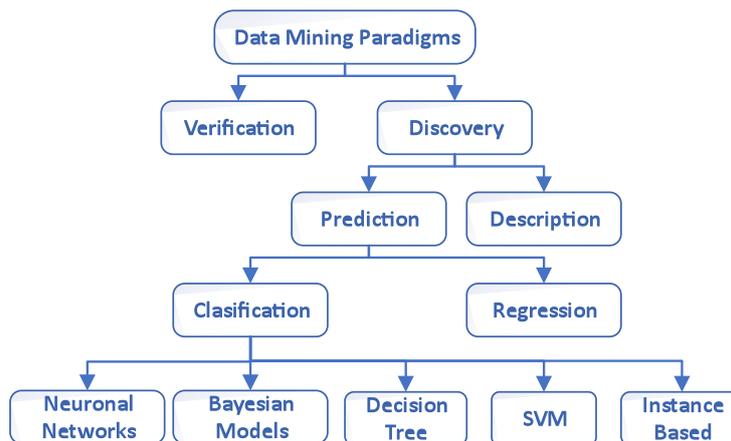


Figure 1. Data Mining Taxonomy.

Source: Maimon, O., & Rokach, (2005)

III. Implementing a Data Mining Project

The data mining projects are implemented with the aim of discovering patterns of relevant and interesting information in large volumes. This is done with the development of four phases (Virsedá Benito & Carrillo, 2008), which are usually:

1. Filtering data.
2. Selection of variables.
3. Extracting knowledge.
4. Interpretation and evaluation.

In general, all techniques have been proved in educational settings (G. Siemens & G. Siemens, 2012), and different case studies have been developed to evaluate the performance of different techniques and to meet the main goals of data mining for education, which seek to identify of behavior patterns of students in their academic environments, classify the types of students according to the recorded performance, classify teachers according to the activities and the use of platforms, identifying successful patterns in the use of virtual learning environments among many others.

IV. Educational Data Mining

The importance that has acquired in recent years the e-learning, has made way to the birth of a discipline that emerges as a tool to solve different paradigms (Almazroui, 2013). This discipline is known as EDM (G. Siemens & G. Siemens, 2012) (R. S. J. Baker & Yacef, 2009) (Cristobal. Romero & Ventura, 2007) (Scheuer & McLaren, 2011); Educational, Data Mining focuses on the development of exploration methods of specific types of data coming from the educational context. Its overall goal is to understand how students learn and identify those aspects that can improve learning and educational aspects. Data Mining for education is so important that in 2007 an international organization dedicated to researching this discipline was founded. Such is the

importance has prompted this research area that it is estimated that by the year 2022 all research related to education will involve analysis and data mining (Baker, Ryan S J, 2014).

Educational information systems now store large amounts of data and its origin can come from different sources, different formats and different granularity levels (Reina, Toral, & Barrero, n.d.). The problems of educational data mining, must be analyzed particularly due to their specific objective determines a singularity when it is solved by data mining techniques.

Data mining in education can analyze the data generated by any system of learning and focus on diverse aspects, both individual and group and take into account underlying, administrative, demographic and motivational data which in turn contain multiple levels of hierarchy, contexts, levels of granularity and historical data. It is called interdisciplinary educational data mining because it can involve the analysis of social networks, educational psychology, cognitive psychology, psychometrics among others.

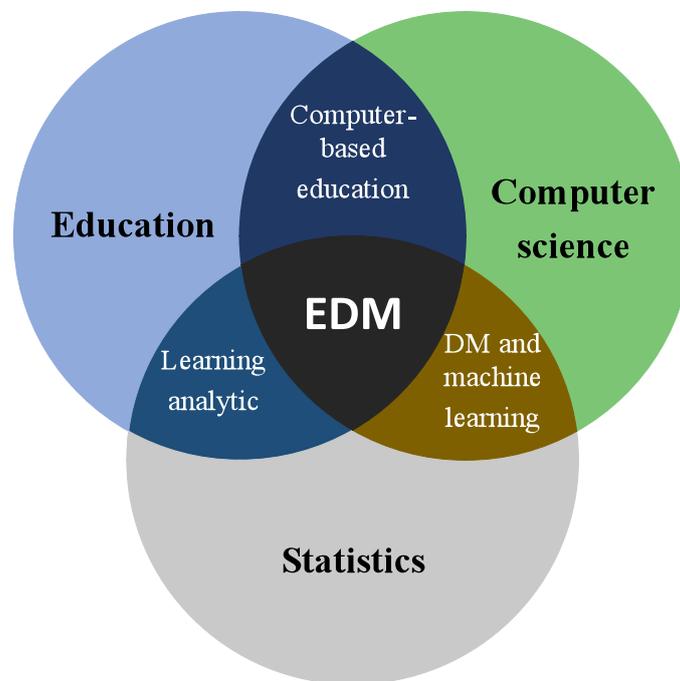


Figure 2. Main areas involved in Educational data mining.

Source: Cristobal Romero & Ventura, (2013)

The most related field areas of the graph is learning analytics (Analysis of Learning) which can be defined as the measurement, collection, analysis and reporting of data about students and their contexts, for understanding and learning purposes and optimizing the environments in which occurs, therefore, Educational data Mining may share many attributes of all and each one of the surrounding areas (Cristobal Romero & Ventura, 2013).

The types of educational environments that exist today, and that can be explored and analyzed by Educational Data Mining are traditional Education and Computer-based Education. Each one of these two environments provide data from different sources that must be pre-processed in particular ways depending on the nature of each of them, the problems and the specific tasks to be solved.

In recent years, the EDM has become very important primarily for the value that this discipline brings to all the actors involved in the educational process, much more, today when the way how to learn and attend to schools and universities has changed so much, this without forgetting that digital learning resources, and social networks are so powerful elements (Mason & Rennie, 2013).

Such importance can be evidenced in the increase of work related to EDM published in the last few years. To illustrate this growth, we show the number of works published in recent years related to EDM in two important sources of scientific information (Harzing & Alakangas, 2016) such as SCOPUS and ScienceDirect.

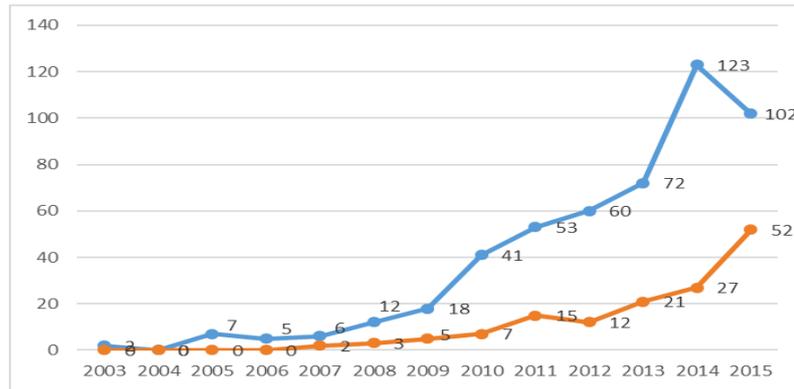


Figure 3. EDM Publications in Scopus.com (blue) and ScienceDirect.com (red).

Source: Own work based on <http://Scopus.com> & <http://ScienceDirect.com>, (2016)

As can be seen in Figure 2, the published articles have been increasing especially since 2008, the year in which the International Educational Data Mining Society (IEDMS) began to advance the international EDM conference. This conference¹ presents full papers (7-8 pages), short papers (4-6 pages) posters (2 pages) and demos related to the use of educational data mining (2 pages). In the figure below the increase in the number of jobs is evident:

		Papers		Posters
		Full	Short	
International Conference on Educational Data Mining	1st. 2008	17		5
	2nd. 2009	20		14
	3th. 2010	24		17
	4th. 2011	20	17	30
	5th. 2012	17	16	17
	6th. 2013	21	25	29
	7th. 2014	24	33	63
	8th. 2015	41	46	39

Table 1. Accepted works in the International Conference on Educational Data Mining

Source: Own work based on <http://educationaldatamining.org/conferences/> (2016)

¹ <http://educationaldatamining.org/EDM2017/call-for-papers-general/>

V. Works in which data mining techniques have used in educational settings

This paper reviewed the data mining techniques used in educational environments and those works analysis. The analysis is presented in the following way:

1. Review of work which used data mining techniques in educational settings.
2. Classified works by educational domains.
3. Classified works by data mining techniques
4. List of relevant authors in EDM.

a. Review of work which used data mining techniques in educational settings

In this paper we have reviewed those papers in which at least one data mining technique has been used to analyze or solve a problem associated with an educational environment, for which the papers found in search results made in Scopus.com were reviewed. and GoogleScholar.com (Harzing & Alakangas, 2016) associated with "Educational Data Mining" that had at least one appointment that was not from the same author and that had a good methodological structure. Those works that by their content did not allow to define which technique they had used were not taken into account.

The papers analyzed are those published between 1993 and 2015, since in 1993 the first paper was published in which, using a data mining technique (neural networks), an attempt was made to identify those students who would approve the courses and which not by (Gedeon & Turner, 1993). As last year analyzed, the year 2015 was left in order to carry out analyzes by educational domains and data mining techniques and facilitate the continuity of this work in subsequent publications.

Works review will be presented chronologically not by used techniques, in that way they will be referenced only once; facilitating reading thereof. In this order, the first referenced work belongs to 1993 and we will conclude with a published work in 2015.

The first known work is the one published in 1993 by Gedeon T.D. and Turner H.S. entitled "Explaining student grades Predicted by a neural network" (Gedeon & Turner, 1993), that one used neural networks to predict the final grades of students. Later, in 1995, Fausett L.V. and Elwasif W. published "Predicting performance from test scores using backpropagation and counter propagation" (Fausett & Elwasif, 1994) work in which neural networks are used to predict the students performance. In 1995 Sanjeev A.P and Zytchow J.M. in "Discovering knowledge in university enrollment databases" (Sanjeev & Zytchow, 1995) presents the usage of association rules to find patterns from enrollment data analysis of a set of on university data; in 2000, Ha S., Bae S. and Park S., published "Web Mining for Distance Education" (Ha, Bae, & Park, 2000) a work that shows how, through association rules, organizations can establish patterns in students academic tastes and effectiveness in the structure of the courses.

In 2001, three EDM papers were published, in which through association rules three different problems are addressed. The first one developed by Yu P., C. and Lin L. who "On learning behavior analysis of web based interactive environment" (Yu, Own, & Lin, 2001) have a job that allows to determine the relationship between learning and behavior patterns so that instructors can promote collaborative learning in WEB environments. Moreover Tsai C. J. Tseng S.S. and Lin C.Y. publish a job in which, using association rules, you can find information that allow instructors to refine or reorganize learning materials and tests adaptive learning environments, this was published in "A Two-phase fuzzy mining and learning algorithm for adaptive learning environment" (Tsai, Tseng, & Lin, 2001), Additionally Yamanishi K. and H. Li in " mining from open answers in questionnaire

data" (Li & Yamanishi, n.d.) show how can you extract information from open responses using the same technique of data mining.

In 2002, three papers were published. In the first, Wang F.H. "On Using Data-Mining Technology for Browsing Log File Analysis in Asynchronous Learning Environment" (F. Wang, 2002) is able to identify learning patterns will allow teachers to modify the structure of teaching in order to make it more efficient. Zaiane O. in his work "Building A Recommender Agent for e-Learning Systems" (Zaiane, 2002) using association rules shows how you can create an agent to recommend activities in e-learning systems; This year, it was also published "Using Neural Networks to Predict Student's Performance" (T. Wang & Mitrović, 2002) where Wang T. and Mitrovic A. display an approach that seeks to predict the performance of students, identifying the number of errors that they may have in a number of problems, using neural networks.

In 2003, Sheard J., Ceddia J., Hurst J. and Tuovinen J. publish "Student Learning Behavior inferring from Website Interactions: A Usage Analysis" (Sheard & Hurst, 2003) where, using decision and clustering trees, analyze the course material and assign homework to students by levels of complexity: Machado L. and Becker K. using association rules evaluate the design of an educational website, that was published in "Distance education: a Case Study web Usage Mining for the Evaluation of Learning Sites " (Machado & Becker, 2003).

In 2004, 6 EDM papers were published, three of them used association rules, one in clustering and the other one in sequential patterns. The ones which use association rules are " Mining interesting contrast rules for a web-based educational system" (Minaei-Bidgoli, Tan, & Punch, 2004) in which Minaei-Bidgoli B., Tan P. and Punch W. discover associations between students' attributes, attributes of identified problems and solution strategies in order to improve education systems online. Moreover, Romero C., Ventura S and De Bra P. in "Knowledge discovery with genetic programming for providing feedback to courseware author. User Modeling and User-Adapted Interaction " (Cristóbal Romero, Ventura, & De Bra, 2004) used this kind of rules to obtain useful information to be used by course authors on how to improve their courses. Likewise, Mor E. and Minguillón J. in " E-learning Personalization based on Itineraries and Long-term navigational behavior " (Mor & Minguillón, 2004) define customizations based on navigational behaviour using the same technique. In the same year, Shen L. and Shen R. used sequential with the purpose of select different objects to students based on learning profiles "Learning content recommendation service based-on a simple sequencing specification" (Shen & Shen, 2004); Lu J. in "A personalized e-learning materials recommender system" (J. Lu, 2004) uses association rules to generate recommendations on learning materials in order to help students finding learning contents and Wang F.H. Shao and H. M. on "Effective personalized recommendation based on time-framed navigation clustering and association mining" (F. H. Wang & Shao, 2004) apply the clustering technique to generate models of future recommendations for students who have similar performances.

In 2005 we found four writings, two of them used sequential patterns and the other two used association rules. Ramli A.A. "Web usage mining using priory algorithm: UUM learning care website case" (Ramli, 2005) shows a case study conducted in UUM University that optimizes the e-learning portal content used at the university from Prior algorithm. Moreover, Kristofic, A. on "Recommender system for adaptive hypermedia applications" (Krištofič, 2005) presents how to generate content recommendations that should see by the student in the future, using adaptive systems; in the same year, Karampiperis P. and Sampson D. also use sequential patterns to produce the sequence of learning resources, adaptively this work was published in "Adaptive learning resources in educational hypermedia sequencing systems." (Karampiperis & Sampson, 2005) and finally, Markellou P., Mousourouli I., Spiros S. and Tsakalidis A. in "using semantic web

mining technologies for personalized e-learning experiences" (Markellou, Mousourouli, Spiros, & Tsakalidis, 2005) show how with the association rules usage can be generated recommendations for learning materials in e-learning systems.

In 2006 Retalis S., Papasalouros A., Psaromilogkos Y., Siscos S. and Kargidis T. published "Towards networked learning analytics - A concept and a tool" (Retalis, Papasalouros, Psaromilogkos, Siscos, & Kargidis, 2006) in which, by using cluster techniques and association rules, achieve to assess the quality of online courses taking students views; In the same way, Spacco J., Winters T., and Payne T. in "inferring use cases from unit testing" (Spacco, Winters, & Payne, 2006) show how using clustering It is possible to get relationships from an evaluation matrix that allows instructors to generate evidence from large amounts of data. On the other hand, Kay J., Maisonneuve N. Yacef K. and Zaiane O.R. published "Mining Patterns of Events in Students' Teamwork Data" (Kay, Maisonneuve, Yacef, & Zaiane, 2006), where, with the use of sequential patterns, they get to identify significant sequences regarding problems or achievements in order to support students to solve problems.. In the same year Chu H.C., Hwang G.J., Tseng J.C.R. and Hwang G.H make use of sequential patterns to produce personalized learning suggestions by analyzing test results and related concepts, this work was published in "A computerized approach to student learning diagnosing problems in health educations" (Engineering, 2006); and finally, Ayers E. and Junker B.W. published "Do skills combine additively to predict task difficulty in eighth grade mathematics" (Ayers & Junker, 2006), a paper which shows how to predict the outcome of the year-end test through student activity with tutors using Bayesian networks.

In 2007, 10 papers were published in which, through the use of data mining techniques, problems in educational environments are solved. In "Using MotSaRT to support online teachers in student motivation" (Weibelzahl, Hurley, & Weibelzahl, 2007) Hurley T. and Weibelzahl S. rely on the use of decision trees to predict in which cases the instructor may recommend certain strategies of students motivation from certain established profiles; on the other hand Vranic M., Painting D., Skocir Z. on "The use of data mining in education environment" (Vranić, Pintar, & Skočir, 2007) and Lu F., Li X., Liu Q., Yang Z., Tan G. and He T. in "Research on Personalized E-Learning System using Fuzzy clustering Algorithm based September" (F. Lu et al., 2007) support the use of clustering and association rules to improve some qualitative aspects of the teaching process and generate recommendations about course based on materials learning habits. Also, Baruque C. B., Amaral M. A., Barcellos, A., Da Silva Freitas J.C. and Longo C. J. employ association rules in "Analyzing users' access to Improve logs in Moodle e-learning" (Baruque, Amaral, Barcellos, da Silva Freitas, & Longo, 2007) to analyze data access Moodle to improve virtual learning; Moreover, Ba-omar Petrounias I. and Anwar F. in "A framework for using web usage mining for personalize e-learning" (Ba-Omar, Petrounias, & Anwar, 2007), using sequential patterns to develop personalized learning scenarios that students can use assisted by systems based on learning styles; using this same technique Liu F. and Shih B. in "Learning activity-based e-learning materials recommendation system" (Liu & Shih, 2007) develop the design of a recommendation system materials based on student learning actions previously stored. Finally, this year, Haddawy P., Thi N. and Hien T.N. in "A decision support system for Evaluating international student applications" (Hien & Haddawy, 2007) show how to predict the students' performance using Bayesian networks and Pardos Z., Heffernan N., Anderson B. and Heffernan C. on "The Effect of Model Granularity on student performance Prediction using Bayesian networks " (ZA Pardos & Heffernan, 2007) from the use of Bayesian networks show how to model the user's knowledge and predict student performance in a mentoring system.

In 2008, Ouyang Y. and Zhu M. in "eLORM: Learning Object Repository based Relationship Mining" (Ouyang & Zhu, 2008) show how to discover patterns to recommend learning objects to

students using sequential patterns; Ranjan J. and Khalil S. in "Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective" (Ranjan & Khalil, 2008) use decision trees and Bayesian networks to support the admission process and to analyze the quality of the education process and student performance in India; Otherwise, Merceron A. and Yacef K. employ association rules to analyze learning data and determine whether students use academic resources and which of them may have greater impact, work published in "Interestingness Measures for Association Rules in Educational Data " (Merceron & Yacef, 2008); using this same technique, Ventura S., Romero C. and Hervás C. in " Analyzing rule evaluation measures with educational datasets: a framework to help the teacher" (Ventura, Romero, & Hervás, 2008) analyze measures assessment rules of educational data in order to identify interesting patterns; Chanchary F.H, Haque I. and Khalid M. S. in "Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment" (Chanchary, 2008a) find relations in access to LMS (Learning Management System) and student behavior to identify patterns Internet usage by the students; Vialardi C., Bravo J. and Ortigosa A. on "Improving AEH courses through log analysis" ("Improving AEH Courses through Log Analysis .," 2015) explain how to improve the design of the course from recommendations generated by *log* analysis of courses. Using this same technique, Zheng, S. Xiong S., Huang Y. and Wu S. in "Using methods of association rules mining optimization in mobile web-based learning system" (Zheng, S., Xiong, Huang, & Wu, 2008) explain how to find relationships between attributes and solution strategies adopted by students in a mobile learning system based on the web. In the same year, Pechenizkiy M., Calders T., Vasilyeva E. and De Bra P. in "the Student Assessment Data Mining: Lessons Drawn from a Small Scale Case Study" (Pechenizkiy, Calders, Vasilyeva, & De Bra, 2008) show a proposal to the use in the extraction data of student assessment, this, using clustering, decision trees and association rules. On the other hand, six papers were published in 2008 in which sequential patterns technique was used, these are: "Personalized recommendation system based on instructing web mining" (L. Zhang, Liu, & Liu, 2008) where Zhang L., and Liu X. show how to customize recommendations based on learning styles and habits of Internet use; in "Sequential pattern analysis software for educational event data" (Nesbit, Xu, Winne, & Zhou, 2008) a paper presented by Nesbit J.C., Xu Y., Winne P. H and Zhou M. who study the behavior of the students eyes, to detect the focal fixations in the courses; "Analyzing rule evaluation measures with educational datasets: a framework to help the teacher" (Ventura et al., 2008) developed by Ventura S., Romero C. and Hervás C. who show how from this technique can generate customized student activities, to help instructors; in "Content recommendation based on Education- contextualized events for browsing web-based personalized learning" (F. H. Wang, 2008) where Wang. F.H. works in generating content recommendations based on events generated by Web browsers to learning customizations; in "A Rule-Based Recommender System for Online Discussion Forums" (Paper, Ibert, & Universidade, 2008) present a framework that allows to display recommendations of interest to students from discussed topics in the discussion forums, paper by Abel F., Bittencourt I.I., Henze N., Krause D. and Vassileva J; "Effective e-learning system based on recommendation self-organizing maps and association mining" (Wen-Shung Tai, Wu, & Li, 2008) where Tai D.W., Wu H. J. and Li P.H. produce a recommendations system based on optional content. In the same year they were also works where Bayesian networks were used to work educational situations, the first of them is "Predicting student's academic performance artificial using neural network: A case study of an engineering course" (Oladokun, Ph, Adebajo, & Sc, 2008) where Oladokun VO, Adebajo A.T and Charles-owaba O.E. explain how to predict the performance a candidate might have if it is accepted in some university courses and "The Composition Effect: conjunctive or Compensatory? An Analysis of Multi-Skill Math Questions in ITS" (Zach Pardos, Beck, Ruiz, & Heffernan, 2008) prepared by Pardos Z., Beck J.E, Ruiz C. and Heffernan N. model two different approaches for determining the probability that a multiple-choice math question should be corrected.

In 2009, 12 papers were published which purpose is to solve problems in educational environments using data mining techniques, among these are correlation analysis, decision trees, Markov chains, classification, clustering, sequential patterns, neural networks and association rules. Published works are "Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data" (Cristóbal Romero, González, Ventura, Jesus, & Herrera, 2009) where Romero C., Gonzalez P., Ventura S., Del Jesus M.J. and Herrera F. find relationships between educational materials and student learning using association rules. These relationships (positive and negative) are primarily given to instructors. In "Mobile formative assessment tool based on data mining techniques for supporting web-based learning" (C. M. Chen & Chen, 2009) developed by Chen C. and Chen M. use correlation analysis, clustering, classification and association rules to support the students' evaluation by the instructors providing information to understand the main factors that influence student performance. "Diagnostic Assessment Data Mining for Concept Similarity" (Madhyastha & Hunt, 2009) where Madhyastha T. and Hunt E. present an analysis on multiple choice data to find concept similarities based on responses ; in "Mining fuzzy association rules from questionnaire data" (Y. L. Chen & Weng, 2009) developed by Chen Y. and Weng C. use association rules to analyze questionnaire data finding patterns in the data of those questionnaires; in "Applying Web Usage Mining for Personalizing Hyperlinks in Web-based Adaptive Educational System" (Cristóbal Romero, Ventura, Zafra, & Bra, 2009) by Romero C., Ventura S., Zafra A. and De Bra P. apply sequential patterns to recommend links to students where they can find interest content based on adaptive educational systems; in "An architecture for making Recommendations to courseware authors using association rule mining and collaborative filtering" (García, Romero, Ventura, & Castro, 2008) written by Garcia E. Romero C., Ventura S. and Castro C. This study use association rules to produce recommendations to course creators on how to improve adaptive courses; in "Implement web learning environment based on data mining" (Guo & Zhang, 2009) paper developed by Guo Q. and Zhang M. where they use neural networks and decision trees to provide support for adaptive and personalized learning; in "Recommendation in higher education using data mining techniques" (Sacín, Agapito, Shafti, & Ortigosa, 2009) prepared by Vialardi C., Brav J. Shafti L. and Ortigosa A. rely on association rules to help students by suggesting subjects should be signed up higher education degrees; in "Data mining for adaptive learning sequence in English language instruction" (Y. H. Wang, Tseng, & Liao, 2009) developed by Wang Y., Tseng M. and Liao H. use decision trees to recommend optimal learning sequences that seek to facilitate the students learning process and to maximize their learning outcomes. Finally in 2009, Stamper J. and Barnes T. published "Unsupervised MDP value selection for Automating ITS capabilities" (Stamper & Barnes, 2009) paper that indicates how to produce automatically adaptation tips using Markov chains.

In 2010 twelve works were published in which, supported by the use of data mining techniques, they deal with specific situations of educational environments. In the same year diverse data mining techniques were used, finding jobs that implement clustering, decision trees, neural networks, classification algorithms, lineal regression, Bayesian networks, association rules and sequential patterns. These works are "Clustering student learning activity data" (Bian, 2010) where Bian H. identifies groups of activities in users with similar performances using clustering; "Classifiers for educational data mining" (Hämäläinen & Vinni, n.d.) prepared by Hämäläinen W. and Vinni M. They make comparisons of some techniques to classify the students situation in learning environment, this study used decision trees, Bayesian networks, neural networks, classification and lineal regression; "Recommender system for predicting student performance" (Thai-nghe, Drumond, Krohn-grimberghe, & Schmidt-thieme, 2010) a paper where Thai-Nghe N., Drumond L., Krohn-Grimberghe A. and Schmidt-Thieme L. propose an approach for using data mining techniques especially those predict the performance of students and using linear regression; " Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks"

(Z. A. Pardos, Heffernan, & Anderson, 2010) prepared by Pardos Z.A., Heffernan N.T., Anderson, B. Heffernan C.L. and Schools W. where they develop a granular model based on Bayesian networks to predict performance of the students; "Mining Rare Association Rules from e-Learning Data" (Cristóbal Romero, Romero, Luna, & Ventura, 2010) writing in which Romero C., Romero J.R, Luna J.M. and Ventura S., analyze and explore association "rare" rules in the usage of LMS Moodle by students; "Use Data Mining to Improve student retention in Higher Education - A Case Study" (Y. Zhang, Oussena, Clark, & Kim, 2010) developed by Zhang Y., Oussena S., Clark T., and Hyensook K. where supported by decision trees and naive Bayes present a case of study where the result of the use of three data mining techniques to improve students retention in higher education is showed; "Early Prediction of Student Success: Mining Students Enrolment Data" (Kovačić, 2010) made by Kovačić, Z. J who performs an analysis of socio-demographic variables and academic environments in students from Open Polytechnic of New Zealand to influence in the permanence or desertion of students based on the decision trees technique. "Mining higher educational students data to analyze students admission in various discipline" (Bhargava, Rajput, & Shrivastava, 2010) presented by Bhargava N., Rajput A. and Shrivastava P. analyze data from college students to decide admission to various disciplines; "Contextual Slip and Prediction of Student Performance After Use of an Intelligent Tutor" (R. Baker et al., 2010) developed by Baker R., Corbett A.T., Gowda, S., Wagner A., MacLaren B. and Kauffman L. who make a comparison of variants used in Bayesian networks to measure student performance, after the use of intelligent tutoring systems; "Data Mining and Student e-Learning Profiles" (Zhou, 2010), prepared by Zhou M. who uses sequential patterns to outline students based on the use of virtual learning environments; "Class Association Rule Mining from Students' Test Data" (Cristóbal Romero, Ventura, Vasilyeva, & Pechenizkiy, 2010) produced by Romero C., Ventura S., Vasilyeva E. and Pechenizkiy, M. proposers of the use of special association rules for discovering relationships in students taking data from one LMS moodle; and "Discovering and Recognizing Student Interaction Patterns in Exploratory Learning Environments" (Bernardini & Conati, 2010) a paper where Bernardini A. and Conati C. use association rules to identify patterns of behavior and classify students of online courses.

In 2011 we have ten works which use diverse data mining techniques such as: association rules, clustering, Bayesian networks, classification, decision trees, neural networks and sequential patterns in order to address educational issues. These publications are "Mining log data for the analysis of learners' Behavior in web-based learning management systems" (Psaromiligkos, Orfanidou, Kyttagias, & Zafiri, 2011) writing in which Psaromiligkos Y., Orfanidou., , Kyttagias C. and Zafiri E. use association rules to improve the continuous feedback process throughout the educational process; "An Empirical Study of the Applications of Data Mining Techniques in Higher Education" (V. Kumar & Chadha, 2011) paper done by Kumar V., Chadha A. where it is shown how the use of data mining techniques such as, association rules and clustering can support administrative and technical activities in higher education; "Data Mining: A prediction of performer or underperformer using classification" (U. K. Pandey & Pal, 2011) work developed by Pandey U. and Pal S. who expose how can predict student performance by using Bayesian networks; "Spectral Clustering in Educational Data Mining" (Trivedi, Pardos, Sárközy, & Heffernan, 2016) developed by Trivedi S., Pardos, Z., Sárközy, G. and Heffernan N.T. who show how, through the use of Spectral Clustering, you can predict student performance ; " *Predicción del Fracaso Escolar mediante Técnicas de Minería de Datos*" (Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, 2012) where Marquez-Vera C., Romero C. and Ventura, S. propose the use of classification techniques and decision trees to predict school failure in programs of the Autonomous University of Zacatecas - Mexico; "Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions" (Trivedi, Pardos, & Heffernan, 2011) published by Trivedi S., Pardos Z., and Heffernan, N. present a proposal on the use of clustering to generate different

test models based on students grouping, "a Data Mining view on Class room Teaching language" (Umesh Kumar Pandey & Pal, 2011) prepared by Pandey U. and Pal S. who use association rules to analyze the impact of language used in the classroom; "Modeling Students' Activity in Online Discussion Forums A Strategy based on Time Series and agglomerative hierarchical clustering" (Cobo et al., 2011) developed by Cobo G. Garcia-Solorzano D., Santamaria E. Moran J., Melenchón J. and Monzo C. who apply the agglomerative hierarchical cluster to group students according to activities; "Anticipating Teachers' Performance" (Barracosa & Antunes, 2011) paper presented by Barracosa J. and Antunes C. who propose a methodology to anticipate the instructors performance from pedagogical surveys using sequential patterns and classification; finally, in 2011 it was published "Improving Student's Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education" (Moucary, Khair, & Zakhem, 2011) where Moucary C., Khair M. and Zakhem W. offer the use of clustering and neural networks to generate different testing models based on students grouping.

In 2012 we found 12 EDM papers, the first of them is written by Rau M.A and Scheines R. "Searching for variables and models to investigate Mediators of learning from multiple representations" (Rau & Scheines, 2012) who use Bayesian networks to compare learnings from multiple graphical representations; Likewise, Eagle M., Johnson M. and Barnes T in "Interaction networks: generating high level hints based on clustering network community" (Eagle, Johnson, & Barnes, 2012) designed a data structure for analyzing interaction information. This data is taken from the of open problems solution where they used clustering; in "Evaluation of model selection strategies for cross-level two-way differential item Functioning analysis" (Patarapichayatham, Kamata, & Kanjanawasee, 2012) developed by Patarapichayatham, C., Kamata, A. and Kanjanawasee, S. who use correlation analysis to assess the impact of model strategic selection; in the same year Baradwaj B. and Pal S. published "Mining Educational Data to Analyze Students' Performance" (Baradwaj & Pal, 2012a) work in the association rules, clustering and decision trees were used to discover student behavior; in the work presented by Bhardwaj B. and Pal S. entitled "Data Mining: A prediction for performance improvement using classification" (Bhardwaj, 2012) a methodological definition was presented to predict student performance by using classification algorithms; In "Application of data mining in academic educational databases for predicting trends and patterns" (Parack, Zahid, & Merchant, 2012) Parack S. Merchant and Zahid F. Z. and used association rules and clustering to define the profile of students; Likewise, Lopez M., J. Moon, Romero C. and S. Ventura present a work that seeks to predict the final outcome of students based on the units of these in the forums using clustering and published in "Classification via clustering for predicting based on student marks end participation in forums" (López, Luna, Romero, & Ventura, 2012); in "Mining Association Rules in Student's Assessment Data" (Varun Kumar & Chadha, 2012) Kumar V. and Chadha A.. present how, from association rules can uncover the factors that are likely to affect academic results in order to obtain successful results in students; in the paper by Bayer J., Bydžovská H., Géryk J., Obsivac T. and Popelinsky L. entitled "Predicting drop-out from social Behavior of students" (Bayer, Bydzovská, & Géryk, 2012) used the technique of classification to determine the possibility of dropping out of college students based on social behaviors; on "Integrating Data Mining in Program Evaluation of K-12 Online Education" (Hung, Hsu, & Rice, 2012) a work published by Hung J., Hsu, Y. and Rice K. apply clustering and decision trees to be integrated into the evaluation program K-12; on "Predicting and analyzing secondary education placement-test scores: A data mining approach" (Şen, Uçar, & Delen, 2012) Şen B., Uçar E. and Delen D. rely on neural networks and decision trees to predict outcomes of high school students in Turkey. Agarwal S., Pandey G., Tiwari M. in "Data Mining in Education: Data Classification and decision tree approach" (Agarwal, Pandey, & Tiwari, 2012) propose a decision tree approach to select students during the courses and includes a comparative analysis of SVM taking student data.

In 2013, we find thirteen works of interest, the first one is "Examining students' online interaction in a live video streaming environment using data mining and text mining" (He, 2013) published by He W. who uses the clustering technique to examine the students' interaction with live video tools; Also Priya K. and Kumar, A. published "Improving the Student's Performance Using Educational Data Mining" (Priya, 2013) and use decision trees to improve student performance in their courses; in "Data mining for providing a personalized learning path in creativity: An application of decision trees" (Lin, Yeh, Hung, & Chang, 2013) published by Lin C., Yeh Y., Hung Y. and Chang R. present the usage of decision trees to establish and facilitate learning routes in order to optimize creativity of learning; on the other hand Martínez-Maldonado R., Yacef K. and Kay J. published "Data Mining in the Classroom: Discovering Groups' Strategies at a Multi-tabletop Environment" (Martínez-maldonado, Yacef, & Kay, 2013) where they use data mining techniques to identify strategies followed by small groups of students in classrooms; in the same year, Romero C., Romero P., Zafra A., Romero R. and Ventura S. published "Web Usage Mining for Predicting Final Marks of Students That Use Moodle Courses" (Cristóbal Romero, Espejo, Zafra, Romero, & Ventura, 2013) and use decision trees to predict the final grade of courses for university students; additionally, Priyama A., Abhijeeta R., Ratheeb A. and Srivastava S. published the paper "Comparative Analysis of Decision Tree Classification Algorithms" (Priyam, Gupta, Rathee, & Srivastava, 2013) where comparing the results of applying decision tree algorithms to predict students' performance; Likewise, Ramesh V., Parkavi P. and Ramar K. publish "Predicting Student Performance: A Statistical and Data Mining Approach" (Ramesh, 2013) where they support on the classification technique to identify factors influencing the result of the students' final exam and establish how to predict the students who may have risk in the courses' approval; otherwise, Blagojević M. and Micić Ž. published "A web-based intelligent report e-learning system using data mining techniques" (Blagojevic, 2013); in the same year, Ali Yahya A., Osman A. and Abdu Alattab A. presented "Educational Data Mining: A Case Study of Teacher's Classroom Questions" (Ali Yahya, Osman, & Abdu Alattab, 2013) a paper where by means of the classification analyze questions asked teachers in the classroom; in "Association rule mining using genetic programming to Provide feedback to instructors from multiple-choice quiz data" (Cristóbal Romero, Zafra, Luna, & Ventura, 2013), Romero C., Zafra A., Luna J.M and Ventura S. used association rules and genetic programming to improve testing and courses at the university level; additionally, Sundar P. published "A Comparative Study for Predicting Students Academic Performance using Bayesian Network Classifiers" (Sundar, 2013) where they make a comparison of obtained results using Bayesian networks in predicting student performance; Likewise, Bhise, R., Thorat, S. and Supekar A. publish "Importance of Data Mining in Higher Education System" (Bhise, Thorat, & Supekar, 2013) writing in which based on clustering to help instructors improving student performance and finally, Jha J. and Ragha L. published "Educational data Mining using Improved Prior algorithm" (Jha & Ragha, 2013) work where present some problems that occur when using the Prior algorithm on data from educational settings, and they make an improvement proposal to the algorithm.

In 2014 we found papers in which data mining techniques were used, such as decision trees, clustering, association rules, neural networks and classification to address educational situations. The first one is "Data Mining: A prediction for Student's Performance Using Classification Method" (Badr, Din, & Elaraby, 2014) work developed by Ahmed A. and Elaraby I. where decision trees are used to predict the final students' grade; on "Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques" (Chalaris, Gritzalis, Maragoudakis, & Sgouropoulou, 2014) Chalaris M., Gritzalis S. and Maragoudakis M. use association rules to provide knowledge related to educational institutions' processes; additionally, Belsis P., Chalaris I., Chalaris M., Skourlas C. and Tsolakidis A. published in "The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules" (Belsis, Chalaris, Chalaris, & Skourlas, 2014) how, from clustering and association rules extraction can analyze the study length

in higher education; in the same year, Guruler H. and Istanbulu A. published "Modeling Student Performance in Higher Education Using Data Mining" (Mugla, 2014) where employ decision trees to identify factors that impact the success of students in higher education; also found a job Rabbany R., Elatia S., Takaffoli M. and Zaiane O. entitled "Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective" (Rabbany, Elatia, Takaffoli, & Zaiane, 2014) where the clustering technique is used to analyze the social networks usage (ARS) and student interaction in forums; additionally, Yukselturk E., Ozekes S. and Türel Y. publish "Predicting dropout student an application of data mining methods in an online education program" (Yukselturk & Education, 2014) a writing where through the use of decision trees, neural networks and classification, examine the students dropout in online programs; in the same year, Chen X., Vorvoreanu M. and Madhavan K. published "Mining Social Media Data for Understanding Students' Learning Experiences" (X. Chen, Member, Vorvoreanu, & Madhavan, 2014a) in this paper with the use of classification technique analyze the students learning experience based on the information discussed in social networks; and finally, Hu Y., Lo C. and Shih S. in the paper entitled "Developing early warning systems to predict students' online learning performance" (Hu, Lo, & Shih, 2014) used classification and regression trees to create an early warning system of students performance in the LMS.

The last year of which have been revised work and where they have used data mining techniques to analyze or treat problems that occur in educational settings is 2015. In this year we found nine papers where different techniques converge, including clustering, decision trees, classification, neural networks and sequential patterns. The analyzed papers are "Educational Data Mining: Performance Evaluation of Decision Tree and Clustering Techniques using WEKA Platform" (Saxena, 2015) developed by Saxena R., this paper used the WEKA tool for comparing the performance of decision trees and clustering techniques in data from educational sector; in "Decision Tree C4.5 algorithm and Its enhanced approach for Educational Data Mining" (Patidar, Dangra, & Rawar, 2015) Patidar P, Dangra J. and Rawar M. describe the use of data mining techniques to improve the efficiency of academic performance in educational institutions; Furthermore, Pradeep A. and Thomas J. publish "Predicting College Students Dropout using EDM Techniques" (Thomas, 2015) use classification and decision trees to identify students likely to have low academic performance; on the other hand, Kaur P., Singh M. and Josan G. on "Classification and prediction based data mining algorithms to predict slow learners in education sector," (Kaur, Singh, & Singh, 2015) based on classification technique to identify students with slow progress and present a predictive model based on classification algorithms; in the same year, Shahiri A. and Husain W. in "A Review on Predicting Student's Performance Using Data Mining Techniques" (Shahiri & Husain, 2015) checked classification, neural networks and decision trees techniques that can be used to predict student performance; In addition, "An Examination of Online Learning Effectiveness Using Data Mining" (Shukor, Tasir, & Meijden, 2015) Shukor N., Taseer Z. and Van der Meijden H. show a predictive model of student performance in online courses using the decision trees technique to the logs of courses; in "Cognitive Skill Analysis for Students through Problem Solving Based on Data Mining Techniques" (Mayilvaganan & Kalpanadevi, 2015) presented by Mayilvaganan M. and Kalpanadevi D. use data mining techniques to assess student skills based on problems solved; on the other hand, Dutt A., Aghabozrgi S., Ismail M. and Mahroeian H. published "Clustering Algorithms Applied in Educational Data Mining" (Dutt, Aghabozrgi, Akmal, Ismail, & Mahroeian, 2015) where they revise some clustering algorithms applied to educational settings data; and finally, Campagni R., Merlini D., Sprugnoli R. and Verri M. published "Data mining models for student careers" (Campagni, Merlini, Sprugnoli, & Verri, 2015) work where supported by clustering and sequential patterns analyze the most suitable careers in graduated students based on the results of course examinations.

In total 127 papers have been revised, where 305 authors have participated and 12 data mining techniques were identified.

The following sections classify the analyzed studies into three groups.

1. Classification of works by educational domains.
2. Papers classified by data mining techniques
3. Data mining techniques used by educational domains.

b. Classification of works by educational domains

There are multiple case studies in which data mining techniques are applied for education, each one of them seeking an answer to a particular situation that contributes to the objective sought by the EDM. In this review they were taken into account published papers between 1993 and 2015 in data mining techniques have been used in any of the following domains widely linked to education:

- *Dropping out or Retention Analysis*: papers in which factors related to dropout are analyze or studies seeking to promote student retention.
- *VLO or VLE Analysis*: those papers in which VLO virtual learning objects or Virtual Learning Environment (VLE) have been analyzed.
- *Performance and students evaluation Analysis*: studies in which the use of data mining techniques was guide to analyze the performance of students or their assessment made during face-to-face or virtual courses.
- *Generation of Educational Recommendations*: domain in which papers that generate recommendations during the educational process are associated.
- *Learning pattern Identification*: in this group is the work that analyzes the way in which virtual students develop in the learning environment and tries to establish the way in which they learn. Generally, these employers seek to group students according to similar characteristics.
- *Students patterns Identification*: data analysis of educational environments which identified patterns among students.
- *Students related Prediction*: in this domain we find studies that allowed to do predictions relating to students, predictions in final grades, performance, behavior in specific courses etc.

Domains	References
Dropping out or Retention Analysis	(Y. Zhang et al., 2010), (Kovačić, 2010), (Cano, Romero, & Ventura, 2013), (Bayer et al., 2012), (Thomas, 2015), (Yukselturk & Education, 2014)
VLO or VLE Analysis	(Li & Yamanishi, n.d.), (Sheard & Hurst, 2003), (Machado & Becker, 2003), (Cristóbal Romero et al., 2004), (Shen & Shen, 2004), (Ramli, 2005), (Karampiperis & Sampson, 2005), (Retalis et al., 2006), (Ba-Omar et al., 2007), (Merceron & Yacef, 2008), (Zheng,

	S. et al., 2008), (Cristóbal Romero, González, et al., 2009), (Madhyastha & Hunt, 2009), (Nesbit et al., 2008), (Ba-Omar et al., 2007), (Guo & Zhang, 2009), (Zach Pardos et al., 2008), (Rau & Scheines, 2012), (Eagle et al., 2012), (Patarapichayatham et al., 2012), (He, 2013), (Blagojevic, 2013), (Ali Yahya et al., 2013), (Rabbany et al., 2014), (X. Chen, Member, Vorvoreanu, & Madhavan, 2014b), (Dutt et al., 2015)
Performance and students evaluation Analysis	(Liu & Shih, 2007), (Saxena, 2015), (Patidar et al., 2015), (Mou Cary et al., 2011), (Bhardwaj, 2012), (Priya, 2013), (Cristóbal Romero, Espejo, et al., 2013), (Priyam et al., 2013), (Ramesh, 2013), (Cristóbal Romero, Zafra, et al., 2013), (Sundar, 2013), (Bhise et al., 2013), (Badr et al., 2014), (Badr et al., 2014), (Hu et al., 2014), (Shahiri & Husain, 2015), (Shukor et al., 2015)
Generation of Educational Recommendations	(Siemens & Baker, 2012), (Vranić et al., 2007), (Psaromiligkos et al., 2011), (Baruque et al., 2007), ("Improving AEH Courses through Log Analysis .," 2015), (Krištofič, 2005), (F. H. Wang & Shao, 2004), (F. Lu et al., 2007), (Sacín et al., 2009), (Y. H. Wang et al., 2009), (Stamper & Barnes, 2009), (V. Kumar & Chadha, 2011), (Hung et al., 2012), (Chalaris et al., 2014)
Learning pattern Identification	(Yu et al., 2001), (Tsai et al., 2001), (Minaei-Bidgoli et al., 2004), (F. Wang, 2002), (Ha et al., 2000), (Spacco et al., 2006), (Y. L. Chen & Weng, 2009), (L. Zhang et al., 2008), (Cristóbal Romero, Ventura, et al., 2009), (Zaiane, 2002), (Markellou et al., 2005), (F. H. Wang, 2008), (Paper et al., 2008), (Engineering, 2006), (García et al., 2008), (J. Lu, 2004), (Wen-Shung Tai et al., 2008), (Hämäläinen & Vinni, n.d.), (Umesh Kumar Pandey

	& Pal, 2011), (Lin et al., 2013), (Bhise et al., 2013), (Jha & Ragha, 2013), (Chalaris et al., 2014), (Belsis et al., 2014), (Mayilvaganan & Kalpanadevi, 2015)
Students patterns Identification	(Sanjeev & Zytchow, 1995), (Mor & Minguillón, 2004), (Kay et al., 2006), (Ranjan & Khalil, 2008), (Chanchary, 2008b), (Pechenizkiy et al., 2008), (C. M. Chen & Chen, 2009), (Bian, 2010), (Cristóbal Romero, Romero, et al., 2010), (Y. Zhang et al., 2010), (Bhargava et al., 2010), (Zhou, 2010), (Cristóbal Romero, Ventura, et al., 2010), (Bernardini & Conati, 2010), (Cobo et al., 2011), (Trivedi et al., 2011), (Trivedi et al., 2011), (Parack et al., 2012), (Varun Kumar & Chadha, 2012), (Agarwal et al., 2012), (Martinez-maldonado et al., 2013), (Mugla, 2014), (Campagni et al., 2015)
Students related Prediction	(Weibelzahl et al., 2007), (C. M. Chen & Chen, 2009), (Ventura et al., 2008), (Fausett & Elwasif, 1994), (Gedeon & Turner, 1993), (T. Wang & Mitrović, 2002), (Oladokun et al., 2008), (Hien & Haddawy, 2007), (ZA Pardos & Heffernan, 2007), (Ayers & Junker, 2006), (Thai-nghe et al., 2010), (Z. A. Pardos et al., 2010), (R. Baker et al., 2010), (U. K. Pandey & Pal, 2011), (Trivedi et al., 2016), (Barracosa & Antunes, 2011), (López et al., 2012), (Oladokun et al., 2008), (Şen et al., 2012), (Kaur et al., 2015)

Table 2. Classification of papers per present domains in education.

Source: Own work

Figure 4 shows the graph of the number of EDM papers by domains

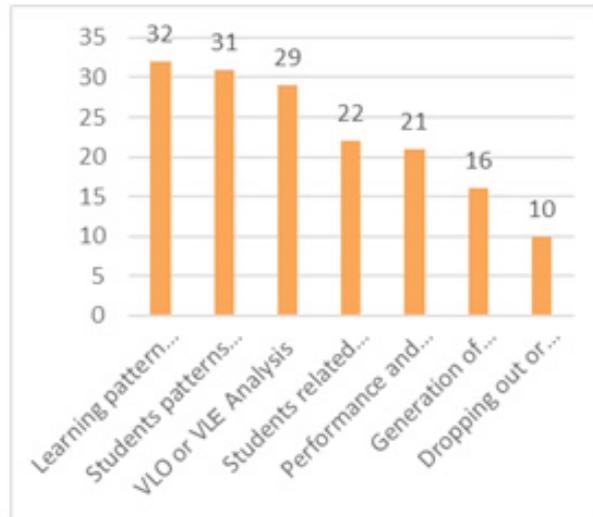


Figure 4. Classification of papers per present domains in education.

Source: Own work

c. Papers classified by data mining techniques

In the analyzed studies was found that 13 data mining techniques are used in total, these are:

- *Correlation Analysis: Technique that allows us to understand how the variables of our interest are related, usually used to convert frequencies into correlation coefficients (Maimon, O., & Rokach, 2005).*
- *Decision Trees: Set of hierarchically organized conditions in the form of a tree from the root to the leaves (Quinlan, 1987).*
- *Regression Trees: Technique that seeks to predict a variable from a set of predictor variables using a set of simple rules (Yohannes & Hoddinott, 1999).*
- *Markov Chains: is a series of events, in which the probability of an event occurring depends on the previous event (Jiawei & Kamber, 2001).*
- *Classification: Classification techniques can be predictive as descriptive, they allow classifying elements within previously defined groups (LÓPEZ, 2017).*
- *Clustering: Technique that automatically identifies groupings of elements according to some elements that make them similar (LÓPEZ, 2017).*
- *Differential Sequence Mining: Technique that identifies and classifies repetition patterns in groupings of elements (LÓPEZ, 2017).*
- *Sequential Patterns: They are an extension of the algorithms of associations by incorporating a temporal component (Shen & Shen, 2004).*
- *Bayesian Networks: Graphic model that allows to represent conditional independencies in a set of variables (Z. A. Pardos et al., 2010) .*
- *Neural Networks: Is an interconnected assembly of simple processing elements, units or nodes, whose functionality is loosely based on the animal neuron (Gurney, 2014).*
- *Association rules: A technique that allows identifying situations that commonly occur in a data set (Jiawei & Kamber, 2001).*
- *Linear regression: It is a statistical technique used to determine the relationship between a dependent variable and independent variables determining their relationship, to create a prediction (Jiawei & Kamber, 2001).*

The analyzed studies in this systematic review are associated with different data mining techniques and that classification is presented below.

Applied Technique	Reference
Correlation Analysis	(Patarapichayatham et al., 2012), (C. M. Chen & Chen, 2009)
Decision Trees	(Sheard & Hurst, 2003), (Weibelzahl et al., 2007), (Ranjan & Khalil, 2008), (Pechenizkiy et al., 2008), (Guo & Zhang, 2009), (Y. H. Wang et al., 2009), (Hämäläinen & Vinni, n.d.), (Y. Zhang et al., 2010), (Kovačić, 2010), (Bhargava et al., 2010), (Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, 2012), (Baradwaj & Pal, 2012b), (Hung et al., 2012), (Şen et al., 2012), (Agarwal et al., 2012), (Priya, 2013), (Lin et al., 2013), (Cristóbal Romero, Espejo, et al., 2013), (Priyam et al., 2013), (Blagojevic, 2013), (Badr et al., 2014), (Mugla, 2014), (Yukselturk & Education, 2014), (Saxena, 2015), (Patidar et al., 2015), (Thomas, 2015), (Shahiri & Husain, 2015), (Shukor et al., 2015)
Regression Trees	(Hu et al., 2014)
Markov Chains	(Stamper & Barnes, 2009)
Classification	(C. M. Chen & Chen, 2009), (Hämäläinen & Vinni, n.d.), (Y. Zhang et al., 2010), (Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, 2012), (Barracosa & Antunes, 2011), (Bhardwaj, 2012), (Bayer et al., 2012), (Ramesh, 2013), (Ali Yahya et al., 2013), (Yukselturk & Education, 2014), (X. Chen et al., 2014b), (Hu et al., 2014), (Thomas, 2015), (Kaur et al., 2015), (Shukor et al., 2015), (Mayilvaganan & Kalpanadevi, 2015)

Clustering	(Sheard & Hurst, 2003), (Mor & Minguillón, 2004), (F. H. Wang & Shao, 2004), (Retalis et al., 2006), (Spacco et al., 2006), (Vranić et al., 2007), (F. Lu et al., 2007), (Pechenizkiy et al., 2008), (C. M. Chen & Chen, 2009), (Madhyastha & Hunt, 2009), (Bian, 2010), (V. Kumar & Chadha, 2011), (Cobo et al., 2011), (Moucary et al., 2011), (Trivedi et al., 2011), (Eagle et al., 2012), (Baradwaj & Pal, 2012b), (Parack et al., 2012), (López et al., 2012), (Hung et al., 2012), (He, 2013), (Bhise et al., 2013), (Chalaris et al., 2014), (Belsis et al., 2014), (Rabbany et al., 2014), (Saxena, 2015), (Dutt et al., 2015), (Campagni et al., 2015)
Differential Sequence Mining	(Martinez-maldonado et al., 2013)
Sequential Patterns	(Shen & Shen, 2004), (Krištofič, 2005), (Karampiperis & Sampson, 2005), (Kay et al., 2006), (Engineering, 2006), (Ba-Omar et al., 2007), (Ouyang & Zhu, 2008), (Liu & Shih, 2007), (L. Zhang et al., 2008), (Nesbit et al., 2008), (Ventura et al., 2008), (F. H. Wang, 2008), (Paper et al., 2008), (Wen-Shung Tai et al., 2008), (Cristóbal Romero, Ventura, et al., 2009), (Zhou, 2010), (Barracosa & Antunes, 2011), (Campagni et al., 2015)
Bayesian Networks	(Ayers & Junker, 2006), (Hien & Haddawy, 2007), (ZA Pardos & Heffernan, 2007), (Ranjan & Khalil, 2008), (Zach Pardos et al., 2008), (Hämäläinen & Vinni, n.d.), (Z. A. Pardos et al., 2010), (R. Baker et al., 2010), (U. K. Pandey & Pal, 2011), (Rau & Scheines, 2012), (Sundar, 2013)
Neural Networks	(Gedeon & Turner, 1993), (Fausett & Elwasif, 1994), (T. Wang & Mitrović, 2002), (Oladokun et al., 2008), (Guo & Zhang, 2009), (Hämäläinen & Vinni, n.d.), (Moucary et al., 2011), (Şen et al., 2012), (Blagojevic, 2013), (Yukselturk & Education, 2014), (Shahiri & Husain,

	2015)
Association rules	(Sanjeev & Zytkow, 1995), (Ha et al., 2000), (Yu et al., 2001), (Tsai et al., 2001), (Li & Yamanishi, n.d.), (F. Wang, 2002), (Zaiane, 2002), (Machado & Becker, 2003), (Minaei-Bidgoli et al., 2004), (Cristóbal Romero et al., 2004), (J. Lu, 2004), (Ramli, 2005), (Markellou et al., 2005), (Retalis et al., 2006), (Vranić et al., 2007), (Baruque et al., 2007), (Merceron & Yacef, 2008), (Chanchary, 2008b), ("Improving AEH Courses through Log Analysis .," 2015), (Zheng, S. et al., 2008), (Cristóbal Romero, González, et al., 2009), (C. M. Chen & Chen, 2009), (Y. L. Chen & Weng, 2009), (García et al., 2008), (Sacín et al., 2009), (Cristóbal Romero, Romero, et al., 2010), (Cristóbal Romero, Ventura, et al., 2010), (Bernardini & Conati, 2010) , (Psaromiligkos et al., 2011), (V. Kumar & Chadha, 2011), (Baradwaj & Pal, 2012b), (Parack et al., 2012), (Varun Kumar & Chadha, 2012), (Cristóbal Romero, Zafra, et al., 2013), (Jha & Raha, 2013), (Chalaris et al., 2014), (Belsis et al., 2014)
Linear regression	(Hämäläinen & Vinni, n.d.), (Thai-nghe et al., 2010)

Table 3. Classification of papers by data mining techniques

Source: Own work

Figure 5 presents the graph of the number of EDM studies classified by techniques used. It is important to underline that in some referenced papers several techniques are used, in those cases these works are related in all techniques they have used, this in order to do not skew the analysis of some techniques in particular.

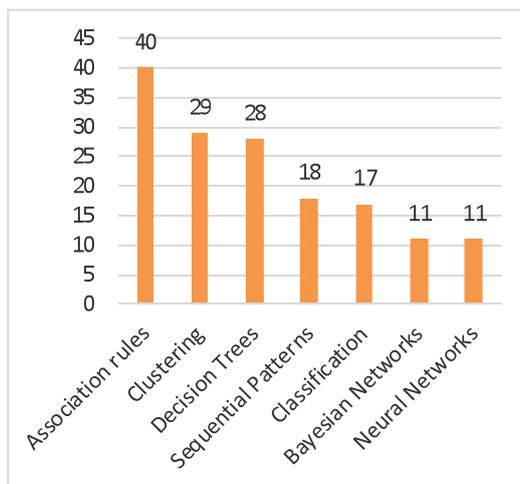


Figure 5. Classification of works by data mining techniques

Source: Own work

d. Data mining techniques used by educational domains

As part of the analysis done in this paper, it is presented below a table that allows to see what data mining techniques are used in educational domains mentioned in section 5.2. This classification seeks to offer the reader a general overview of what techniques are used in particular domains.

The above table identify in a quickly way what techniques are most commonly used in each domain.

Technique \ Domain	Dropping out or Retention Analysis	VLO or VLE Analysis	Performance and students evaluation Analysis	Generation of Educational Recommendations	Learning pattern Identification	Students patterns Identification	Students related Prediction
Correlation Analysis		1				1	
Decision Trees	5	3	8	2	2	6	2
Regression Trees			1				
Markov Chains				1			
Classification	4	2	4		3	1	3
Clustering		7	3	5	3	9	2
Differential Sequence Mining						1	
Sequential Patterns		4	1	1	7	3	2
Bayesian Networks		2	1		1	1	6
Neural Networks	1	2	2		1		5
Association rules		8	1	7	14	9	1
Linear regression					1		1

Table 4. Data mining techniques used in educational domains.

Source: Own work

VI. Conclusions

It is evident the importance that currently EDM has. More than a research discipline, this area has become a tool used at all educational levels, especially in higher education. Over the last 20 years over twelve data mining techniques have been used to analyze contexts or particular domains of education, where the association rules, clustering, decision trees and sequential patterns are the most commonly used. Moreover, it can be identified that most of the situations presented in educational environments can be analyzed using data mining techniques. The domains most commonly analyzed in education by using data mining techniques are learning pattern identification, VLO or VLE Analysis and Students patterns Identification. However, since 2010, many studies have focused on the Dropping out analysis itself.

In general, we could identify that studies produced around EDM have been mainly focused on case studies, and data analysis especially stored in LMS.

The result of this work in which more than 100 documents were reviewed where data mining techniques were used to analyze, understand or solve a particular situation in an educational environment is presented in table number 4. This table is the summary of the work done and is the association of the 7 domains analyzed and the 12 data mining techniques used. This table can be

used by all those who wish to work on EDM projects to guide the work based on the advanced review presented throughout the work.

Future work should describe the growing forms of analytics that are presented in different domains (academic, business, other) and the fundamental role that machine learning and artificial intelligence must face in these processes and especially in the educational field, the systematic review of the literature presented is expected to be a starting point for new research that shows these types of challenges. It is also expected to continue characterizing in a more granular way in which part of the educational process these techniques of text mining and analytics should impact in the educational field, and know if concepts such as learning analytics and other concepts of these will focus on the teaching process, student process knowledge discovery or educational management.

References

- Agarwal, S., Pandey, G. N., & Tiwari, M. D. (2012). Data Mining in Education□: Data Classification and Decision Tree Approach. *International Journal of E-Education, e-Business, e-Management and e-Learning*, 2(2), 140–144.
- Ali Yahya, A., Osman, A., & Abdu Alattab, A. (2013). Educational Data Mining□: A Case Study of Teacher ' s Classroom Questions. *IEEE 13th International Conference On*, (February), 92–97.
- Almazroui, Y. a. (2013). A survey of Data mining in the context of E-learning. *International Journal of Information Technology & Computer Science*, 7, 8–18. Retrieved from http://www.ijitcs.com/volume_7_No_3/Yousef+Almazroui.pdf
- Ayers, E., & Junker, B. W. (2006). Do skills combine additively to predict task difficulty in eighth grade mathematics? *Educational Data Mining Workshop at {AAAI}*, 14–20.
- Ba-Omar, H., Petrounias, I., & Anwar, F. (2007). A Framework for Using Web Usage Mining to Personalise E-learning. *Seventh IEEE International Conference on Advanced Learning Technologies (ICALT 2007)*, 1(2006), 937–938. <https://doi.org/10.1109/ICALT.2007.13>
- Badr, A., Din, E., & Elaraby, I. S. (2014). Data Mining□: A prediction for Student ' s Performance Using Classification Method. *World Journal of Computer Application and Technology*, 2(2), 43–47. <https://doi.org/10.13189/wjcat.2014.020203>
- Baker, Ryan S J, P. S. (2014). Educational data mining and learning analytics.
- Baker, R., Corbett, A. T., Gowda, S. M., Wagner, A. Z., Maclaren, B. A., Kauffman, L. R., ... Giguere, S. (2010). Contextual Slip and Prediction of Student Performance After Use of an Intelligent Tutor. *User Modeling, Adaptation, and Personalization*, 52–63.
- Baker, R. S. J., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *Journal of Educational Data Mining*, 1(1), 3–17. Retrieved from http://www.educationaldatamining.org/JEDM/images/articles/vol1/issue1/JEDMVol1Issue1_BakerYacef.pdf
- Baradwaj, B. K., & Pal, S. (2012a). Mining Educational Data to Analyze Students' Performance, 2(6), 63–69.
- Baradwaj, B. K., & Pal, S. (2012b). Mining Educational Data to Analyze Students " Performance, 2(6), 63–69.

- Barracosa, J., & Antunes, C. (2011). Anticipating Teachers' Performance. *KDD 2011 Workshop: Knowledge Discovery in Educational Data*, 77–82.
- Baruque, C. B., Amaral, M. A., Barcellos, A., da Silva Freitas, J. C., & Longo, C. J. (2007). Analysing users' access logs in Moodle to improve e learning. *Euro American Conference on Telematics and Information Systems*, 72.
- Bayer, J., Bydzovská, H., & Géryk, G. (2012). Predicting drop-out from social behaviour of students. *International Educational Data Mining Societ, (Dm)*, 103–109.
- Belsis, P., Chalaris, I., Chalaris, M., & Skourlas, C. (2014). The Analysis of the Length of Studies in Higher Education based on Clustering and the Extraction of Association Rules. *Procedia - Social and Behavioral Sciences*, 147, 567–575. <https://doi.org/10.1016/j.sbspro.2014.07.159>
- Bernardini, A., & Conati, C. (2010). Discovering and Recognizing Student Interaction Patterns in Exploratory Learning Environments. *Intelligent Tutoring Systems*, 125–134.
- Bhardwaj, B. K. (2012). Data Mining□: A prediction for performance improvement using classification, 9(4).
- Bhargava, N., Rajput, A., & Shrivastava, P. (2010). Mining higher educational students data to analyze student's admission in various discipline. *Binary Journal of Data Mining & Networking*, 1, 1–5.
- Bhise, R. B., Thorat, S. S., & Supekar, A. K. (2013). Importance of Data Mining in Higher Education System. *IOSR Journal Of Humanities And Social Science*, 6(6), 18–21.
- Bian, H. (2010). Clustering Student Learning Activity Data. *Proceedings of the 3rd International Conference on Educational Data Mining*, 277–278.
- Blagojevic, M. (2013). A web-based intelligent report e-learning system using data, 39, 465–474.
- Campagni, R., Merlini, D., Sprugnoli, R., & Verri, M. C. (2015). Data mining models for student careers. *Expert Systems with Applications*, 42, 5508–5521.
- Cano, C. M. A., Romero, C., & Ventura, S. (2013). Predicting student failure at school using genetic programming and different data mining approaches with high dimensional and imbalanced data, 315–330. <https://doi.org/10.1007/s10489-012-0374-8>
- Chalaris, M., Gritzalis, S., Maragoudakis, M., & Sgouropoulou, C. (2014). Improving Quality of Educational Processes Providing New Knowledge using Data Mining Techniques. *Procedia - Social and Behavioral Sciences*, 147, 390–397. <https://doi.org/10.1016/j.sbspro.2014.07.117>
- Chanchary, F. H. (2008a). Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment. *ResearchGate*. <https://doi.org/10.1109/WKDD.2008.139>
- Chanchary, F. H. (2008b). Web Usage Mining to Evaluate the Transfer of Learning in a Web-Based Learning Environment . Web Usage Mining to Evaluate the Transfer of Learning in a Web-based Learning Environment, (October 2015). <https://doi.org/10.1109/WKDD.2008.139>
- Chen, C. M., & Chen, M. C. (2009). Mobile formative assessment tool based on data mining techniques for supporting web-based learning. *Computers & Education*, 52, 256–273.
- Chen, X., Member, S., Vorvoreanu, M., & Madhavan, K. (2014a). Mining Social Media Data for Understanding Students' Learning Experiences, 7(3), 246–259.
- Chen, X., Member, S., Vorvoreanu, M., & Madhavan, K. (2014b). Mining Social Media Data for Understanding Students ' Learning Experiences, 7(3), 246–259.
- Chen, Y. L., & Weng, C. H. (2009). Mining fuzzy association rules from questionnaire data. *Knowledge-Based Systems*, 22, 46–56.

- Cobo, G., García-Solórzano, D., Santamaria, E., Morán, J. A., Melenchón, J., & Monzo, C. (2011). Modeling Students' Activity in Online Discussion Forums: A Strategy based on Time Series and Agglomerative Hierarchical Clustering. *EDM 2011*, (February 2016), 253–258.
- Dutt, A., Aghabozrgi, S., Akmal, M., Ismail, B., & Mahroeian, H. (2015). Clustering Algorithms Applied in Educational Data Mining, *5*(2), 112–116. <https://doi.org/10.7763/IJIEE.2015.V5.513>
- Eagle, M., Johnson, M., & Barnes, T. (2012). Interaction Networks: Generating High Level Hints Based on Network Community Clustering. *International Educational Data Mining Society*, 164–167. Retrieved from http://eric.ed.gov/?q=intelligent+tutoring+systems&ff1=dtSince_2011&pg=14&id=ED537223
- Engineering, I. (2006). A Computerized Approach to Diagnosing Student Learning Problems in Health Education. *Health (San Francisco)*, *1*(1), 43–60.
- Fausett, L. V., & Elwasif, W. (1994). Predicting performance from test scores using back propagation and counter propagation. *IEEE World Congress on Computational Intelligence*, 3398–3402.
- G. Siemens, R. S. j. d. B., & G. Siemens, R. S. j. d. B. (2012). Learning analytics and educational data mining: towards communication and collaboration. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge*.
- García, E., Romero, C., Ventura, S., & Castro, C. De. (2008). An architecture for making recommendations to courseware authors using association rule mining and collaborative filtering. *User Modelling and User-Adapted Interaction*, *19*(1–2 SPEC. ISS.), 99–132. <https://doi.org/10.1007/s11257-008-9047-z>
- Gedeon, T. D., & Turner, H. S. (1993). Explaining student grades predicted by a neural network. *International Joint Conference IEEE*, *1*, 609–612.
- Guo, Q., & Zhang, M. (2009). Implement web learning environment based on data mining. *Knowledge-Based Systems*, *22*, 439–442.
- Gurney, K. (2014). *An introduction to neural networks*. *Neural Network World* (Vol. 6). [https://doi.org/10.1016/S0140-6736\(95\)91746-2](https://doi.org/10.1016/S0140-6736(95)91746-2)
- Ha, S. H., Bae, S. M., & Park, S. C. (2000). Web mining for distance education. *Proceedings of the 2000 IEEE International Conference on Management of Innovation and Technology*, *2*, 715–719. <https://doi.org/10.1109/ICMIT.2000.916789>
- Hämäläinen, W., & Vinni, M. (n.d.). Classifiers for educational data mining. *Handbook of Educational Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery Series*, 57–71.
- Harzing, A. W., & Alakangas, S. (2016). Google Scholar, Scopus and the Web of Science: a longitudinal and cross-disciplinary comparison. *Scientometrics*. <https://doi.org/10.1007/s11192-015-1798-9>
- He, W. (2013). Examining Students' Online Interaction in a Live Video Streaming Environment Using Data Mining and Text Mining Computers in Human Behavior. *Computers in Human Behavior*, (February), 90–102. <https://doi.org/10.1016/j.chb.2012.07.020>
- Hien, N. T. N., & Haddawy, P. (2007). A decision support system for evaluating international student applications. *Proceedings - Frontiers in Education Conference, FIE*, (NOVEMBER 2007), 1–6. <https://doi.org/10.1109/FIE.2007.4417958>
- Hu, Y. H., Lo, C. L., & Shih, S. P. (2014). Developing early warning systems to predict students' online learning performance. *Computers in Human Behavior*, 469–478.

- Hung, J., Hsu, Y., & Rice, K. (2012). Integrating Data Mining in Program Evaluation of K-12 Online Education, *15*, 27–41.
- Improving AEH Courses through Log Analysis . (2015), (October).
- Jha, J., & Ragha, L. (2013). Educational Data Mining using Improved Apriori Algorithm, *3*(5), 411–418.
- Jiawei, H., & Kamber, M. (2001). Data mining: concepts and techniques. *San Francisco, CA, Itd: Morgan Kaufmann*, 377–385. [https://doi.org/10.1002/1521-3773\(20010316\)40:6<9823::AID-ANIE9823>3.3.CO;2-C](https://doi.org/10.1002/1521-3773(20010316)40:6<9823::AID-ANIE9823>3.3.CO;2-C)
- Karampiperis, P., & Sampson, D. (2005). Adaptive Learning Resources Sequencing in Educational Hypermedia Systems. *Educational Technology & Society*, *8*, 128–147.
- Kaur, P., Singh, M., & Singh, G. (2015). Classification and prediction based data mining algorithms to predict slow learners in education sector. *Procedia - Procedia Computer Science*, *57*, 500–508. <https://doi.org/10.1016/j.procs.2015.07.372>
- Kay, J., Maisonneuve, N., Yacef, K., & Zaïane, O. R. (2006). Mining patterns of events in students ' teamwork data. *Proceedings of the Workshop on Educational Data Mining at the 8th International Conference on Intelligent Tutoring Systems*, 1–8.
- Kovačić, Z. J. (2010). Early Prediction of Student Success□: Mining Students Enrolment Data. *In Proceedings of Informing Science & IT Education Conference (InSITE)*, 647–665.
- Krištofič, A. (2005). Recommender System for Adaptive Hypermedia Applications. *Student Research Conference*, 229–234.
- Kumar, V., & Chadha, A. (2011). An Empirical Study of the Applications of Data Mining Techniques in Higher Education. *International Journal of Advanced Computer Science and Application*, *2*(3), 80–84.
- Kumar, V., & Chadha, A. (2012). Mining Association Rules in Student ' s Assessment Data. *International Journal of Computer Science Issues*, *9*(5), 211–216.
- Li, H., & Yamanishi, K. (n.d.). Mining from Open Answers in Questionnaire Data.
- Lin, C. F., Yeh, Y. C., Hung, Y. H., & Chang, R. I. (2013). Computers & Education Data mining for providing a personalized learning path in creativity□: An application of decision trees. *Computers & Education*, *68*, 199–210. <https://doi.org/10.1016/j.compedu.2013.05.009>
- Liu, F. J., & Shih, B. J. (2007). Learning Activity-Based E-Learning Material Recommendation System. *ISMW'07. Ninth IEEE International Symposium*, 343–348.
- LÓPEZ, C. P. (2017). Minería de datos: técnicas y herramientas. *International Conference on Universidad Nacional de Colombia*. Retrieved from <http://disi.unal.edu.co/profesores/eleonguz/cursos/md/>
- López, M. I., Luna, J. M., Romero, C., & Ventura, S. (2012). Classification via clustering for predicting final marks based on student participation in forums. *International Educational Data Mining Society.*, 148–151.
- Lu, F., Li, X., Liu, Q., Yang, Z., Tan, G., & He, T. (2007). Research on Personalized E-Learning System Using Fuzzy Set Based Clustering Algorithm. *Computational Science–ICCS*, 587–590.
- Lu, J. (2004). A Personalized e-Learning Material Recommender System. *Proceedings of the 2nd International Conference on Information Technology for Application (ICITA 2004)*, (Icita), 374–380.
- Machado, L., & Becker, K. (2003). Distance education: a Web usage mining case study for the

- evaluation of learning sites. *Proceedings 3rd IEEE International Conference on Advanced Technologies*, 360–361. <https://doi.org/10.1109/ICALT.2003.1215123>
- Madhyastha, T., & Hunt, E. (2009). Mining Diagnostic Assessment Data for Concept Similarity. *JEDM - Journal of Educational Data Mining*, 1(1), 72–91. <https://doi.org/Similarity>
- Maimon, O., & Rokach, L. (2005). *Data Mining and Knowledge Discovery Handbook. Data Mining and Knowledge Discovery Handbook*. https://doi.org/10.1007/0-387-25465-x_2
- Markellou, P., Mousourouli, I., Spiros, S., & Tsakalidis, a. (2005). Using semantic web mining technologies for personalized e-learning experiences. *Proceedings of the Web-Based Education*, 461–826. Retrieved from http://www.hci.gr/files/461-826_Markellou.pdf
- Martinez-maldonado, R., Yacef, K., & Kay, J. (2013). Data Mining in the Classroom: Discovering Groups' Strategies at a Multi-tabletop Environment. *International Conference on Educational Data Mining*, 121–128.
- Mason, R., & Rennie, F. (2013). *E-learning and social networking handbook: Resources for higher education*.
- Mayilvaganan, M., & Kalpanadevi, D. (2015). Cognitive Skill Analysis for Students through Problem Solving Based on Data Mining Techniques. *Procedia - Procedia Computer Science*, 47, 62–75. <https://doi.org/10.1016/j.procs.2015.03.184>
- Merceron, A., & Yacef, K. (2008). Interestingness Measures for Association Rules in Educational Data. *Proceedings of the 1st International Conference on Educational Data Mining*, 57–66. Retrieved from <papers3://publication/uuid/E0EAD44B-49CB-44D3-BA64-E29F555E8470>
- Minaei-Bidgoli, B., Tan, P.-N. T. P.-N., & Punch, W. F. (2004). Mining interesting contrast rules for a web-based educational system. *2004 International Conference on Machine Learning and Applications, 2004. Proceedings*. <https://doi.org/10.1109/ICMLA.2004.1383530>
- MINTIC. (2015). Panorama TIC | Publicado en marzo de 2015. *PanoramaTIC*.
- Mor, E., & Minguillón, J. (2004). E-learning personalization based on itineraries and long-term navigational behavior. *Proceedings of the 13th International World Wide Web Conference on Alternate Track Papers Posters WWW Alt 04*, 264. <https://doi.org/10.1145/1013367.1013427>
- Moucary, C. E., Khair, M., & Zakhem, W. (2011). Improving Student's Performance Using Data Clustering and Neural Networks in Foreign-Language Based Higher Education. *The Research Bulletin of Jordan ACM*, 2(3), 27–34.
- Mugla, H. G. (2014). Modeling Student Performance in Higher Education Using Data Mining Modeling Student Performance in Higher Education Using Data Mining, (February 2016). <https://doi.org/10.1007/978-3-319-02738-8>
- Nesbit, J. C., Xu, Y., Winne, P. H., & Zhou, M. (2008). Sequential pattern analysis software for educational event data Data Mining Software for Educational Log Analysis. *Measuring Behavior 2008*, 160.
- Oladokun, V. O., Ph, D., Adebajo, a T., & Sc, B. (2008). Predicting Students' Academic Performance using Artificial Neural Network: A Case Study of an Engineering Course. *The Pacific Journal of Science and Technology*, 9(1), 72–79.
- Ouyang, Y., & Zhu, M. (2008). eLORM: Learning Object Relationship Mining based Repository. *Online Information Review*, 32, 254–265.
- Pandey, U. K., & Pal, S. (2011). A Data Mining view on Class Room Teaching Language, 8(2), 277–282.
- Pandey, U. K., & Pal, S. (2011). Data Mining: A prediction of performer or underperformer using

- classification, 2(2), 686–690.
- Paper, C., Ibert, I., & Universidade, B. (2008). Adaptive Hypermedia and Adaptive Web-Based Systems, 5149(October). <https://doi.org/10.1007/978-3-540-70987-9>
- Parack, S., Zahid, Z., & Merchant, F. (2012). Application of Data Mining in Educational Databases for Predicting Academic Trends and Patterns. *Technology Enhanced Education (ICTEE)*, 1–4.
- Pardos, Z. A., Heffernan, N. T., & Anderson, B. (2010). Using Fine-Grained Skill Models to Fit Student Performance with Bayesian Networks. *Handbook of Educational Data Mining*, 417.
- Pardos, Z., Beck, J. E., Ruiz, C., & Heffernan, N. (2008). The Composition Effect: Conjunctive or Compensatory? An Analysis of Multi-Skill Math Questions in ITS. *Proceedings of the 1st International Conference on Educational Data Mining*, 147–156.
- Pardos, Z., & Heffernan, N. (2007). The effect of model granularity on student performance prediction using Bayesian networks. *User Modeling 2007. Springer Berlin Heidelberg*, 435–439. Retrieved from http://link.springer.com/chapter/10.1007/978-3-540-73078-1_60
- Patarapichayatham, C., Kamata, a., & Kanjanawasee, S. (2012). Evaluation of Model Selection Strategies for Cross-Level Two-Way Differential Item Functioning Analysis. *Educational and Psychological Measurement*, 72, 44–51. <https://doi.org/10.1177/0013164411409743>
- Patidar, P., Dangra, J., & Rawar, M. K. (2015). Decision Tree C4.5 algorithm and its enhanced approach for Educational Data Mining, 7(2), 1–14.
- Pechenizkiy, M., Calders, T., Vasilyeva, E., & De Bra, P. (2008). Mining the Student Assessment Data: Lessons Drawn from a Small Scale Case Study. *Educational Data Mining*, 187–191.
- Priya, K. S. (2013). Improving the Student 's Performance Using Educational Data Mining, 1685, 1680–1685.
- Priyam, A., Gupta, R., Rathee, A., & Srivastava, S. (2013). Comparative Analysis of Decision Tree Classification Algorithms, 334–337.
- Psaromiligkos, Y., Orfanidou, M., Kytagiias, C., & Zafiri, E. (2011). Mining log data for the analysis of learners' behaviour in web-based learning management systems. *Operational Research*, 11, 187–200.
- Quinlan, J. R. (1987). Generating production rules from decision trees. *Proceedings of the Tenth International Joint Conference on Artificial Intelligence*, 30107, 304–307. Retrieved from <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.98.9054&rep=rep1&type=pdf>
- R, B., Sánchez-Guzmán A, & García, R. (2013). Minería de datos educativa: Una herramienta para la investigación de patrones de aprendizaje sobre un contexto educativo. (Spanish). *Latin-American Journal of Physics Education*, 7(4), p662-668.
- Rabbany, R., Elatia, S., Takaffoli, M., & Zaiane, O. R. (2014). Collaborative Learning of Students in Online Discussion Forums: A Social Network Analysis Perspective. *EDM 2014*, 1–25.
- Ramesh, V. (2013). Predicting Student Performance: A Statistical and Data Mining Approach, 63(8), 35–39.
- Ramli, A. A. (2005). Web usage mining using apriori algorithm: uum learning care portal case. *In International Conference on Knowledge Management, Malaysia*, 1–19.
- Ranjan, J., & Khalil, S. (2008). Conceptual Framework of Data Mining Process in Management Education in India: An Institutional Perspective. *Information Technology Journal*, 7(1), 16–23.
- Rau, M. A., & Scheines, R. (2012). Searching for Variables and Models to Investigate Mediators of

- Learning from Multiple Representations. *International Educational Data Mining Society*, 110–117. Retrieved from http://eric.ed.gov/?q=intelligent+tutoring+systems&ff1=dtYSince_2011&pg=14&id=ED537203
- Reina, D. G., Toral, S. L., & Barrero, F. (n.d.). Metodologías de Análisis de los Big Data en las Plataformas Educativas.
- Retalis, S., Papasalouros, a, Psaromiligkos, Y., Siscos, S., & Kargidis, T. (2006). Towards Networked Learning Analytics – A concept and a tool. *Networked Learning*, 1–8.
- Romero, C., Espejo, P. G., Zafra, A., Romero, J. R., & Ventura, S. (2013). Web Usage Mining for Predicting Final Marks of Students That Use Moodle Courses. *Computer Applications in Engineering Education*, 135–146. <https://doi.org/10.1002/cae.20456>
- Romero, C., González, P., Ventura, S., Jesus, M. J., & Herrera, F. (2009). Evolutionary algorithms for subgroup discovery in e-learning: A practical application using Moodle data. *Expert Systems with Applications*, 36(2), 1632–1644.
- Romero, C., Romero, J. R., Luna, J. M., & Ventura, S. (2010). Mining Rare Association Rules from e-Learning Data. *EDM 2010*, 171–180.
- Romero, C., & Ventura, S. (2007). Educational data mining: A survey from 1995 to 2005. *Expert Systems with Applications*, 33(1), 135–146. <https://doi.org/10.1016/j.eswa.2006.04.005>
- Romero, C., & Ventura, S. (2013). Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1), 12–27. <https://doi.org/10.1002/widm.1075>
- Romero, C., Ventura, S., & De Bra, P. (2004). Knowledge Discovery with Genetic Programming for Providing Feedback to Courseware Authors, 1–48.
- Romero, C., Ventura, S., Vasilyeva, E., & Pechenizkiy, M. (2010). Class Association Rule Mining from Students' Test Data. *EDM 2010*, 317–318.
- Romero, C., Ventura, S., Zafra, A., & Bra, P. de. (2009). Applying Web usage mining for personalizing hyperlinks in Web-based adaptive educational systems. *Computers & Education*, 53(3), 828–840. <https://doi.org/10.1016/j.compedu.2009.05.003>
- Romero, C., Zafra, A., Luna, J. M., & Ventura, S. (2013). Association rule mining using genetic programming to provide feedback to instructors from multiple-choice quiz data, 30(2), 162–172. <https://doi.org/10.1111/j.1468-0394.2012.00627.x>
- Romero Morales, Cristóbal; Márquez Vera, Carlos; Ventura Soto, S. (2012). Predicción del Fracaso Escolar Mediante Técnicas de Minería de Datos. *Iee-Rita*, 7(3), 109–117.
- Sacín, C. V., Agapito, J. B., Shafti, L., & Ortigosa, A. (2009). Recommendation in Higher Education Using Data Mining Techniques. *Proceedings of the 2nd International Conference on Educational Data Mining*, 191–199. Retrieved from <http://www.educationaldatamining.org/EDM2009/uploads/proceedings/vialardi.pdf>
- Sanjeev, A. P., & Zytchow, J. M. (1995). Discovering Enrollment Knowledge in University Databases., 246–251.
- Saxena, R. (2015). Educational Data Mining: Performance Evaluation of Decision Tree and, 14(April), 1–10.
- Scheuer, O., & McLaren, B. M. (2011). Educational Data Mining.
- Şen, B., Uçar, E., & Delen, D. (2012). Predicting and analyzing secondary education placement-test scores: A data mining approach. *Expert Systems with Applications*, 39, 9468–9476. <https://doi.org/10.1016/j.eswa.2012.02.112>

- Shahiri, A. M., & Husain, W. (2015). A Review on Predicting Student's Performance using Data Mining Techniques. *Procedia Computer Science*, 72, 414–422. <https://doi.org/10.1016/j.procs.2015.12.157>
- Sheard, J., & Hurst, A. J. (2003). Inferring Student Learning Behaviour from Website Interactions: A Usage Analysis Inferring Student Learning Behaviour from, (February 2016). <https://doi.org/10.1023/A>
- Shen, L. P., & Shen, R. M. (2004). Learning Content Recommendation Service Based-on Simple Sequencing Specification. *Advances in Web-Based Learning-ICWL*, 363–370.
- Shukor, N. A., Tasir, Z., & Meijden, H. Van Der. (2015). An examination of online learning effectiveness using data mining. *Procedia - Social and Behavioral Sciences*, 172, 555–562. <https://doi.org/10.1016/j.sbspro.2015.01.402>
- Siemens, G., & Baker, R. S. J. D. (2012). Learning analytics and educational data mining. *Proceedings of the 2nd International Conference on Learning Analytics and Knowledge - LAK '12*, 252. <https://doi.org/10.1145/2330601.2330661>
- Society, I. (2014). Informe Global de Internet 2014, 12.
- Spacco, J., Winters, T., & Payne, T. (2006). Inferring use cases from unit testing. *AAAI Workshop on Educational Data Mining*, 1–7. Retrieved from <http://www.aaai.org/Papers/Workshops/2006/WS-06-05/WS06-05-010.pdf>
- Stamper, J., & Barnes, T. (2009). An unsupervised, frequency-based metric for selecting hints in an MDP-based tutor. *Proceedings of the 2nd International Conference on Educational Data Mining*, 181–190. Retrieved from <http://www.educationaldatamining.org/EDM2009/uploads/proceedings/stamper.pdf>
- Sundar, P. P. (2013). A Comparative Study For Predicting Students Academic Performance using Bayesian Network Classifiers. *IOSR Journal of Engineering*, 3(2), 37–42.
- 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. <https://doi.org/10.1016/j.procs.2010.08.006>
- Thomas, J. (2015). Predicting College Students Dropout using EDM Techniques, 123(5), 26–34.
- Trivedi, S., Pardos, Z. A., & Heffernan, N. T. (2011). Clustering Students to Generate an Ensemble to Improve Standard Test Score Predictions, G. Biswas et al. *Artificial Intelligence in Education*, 377–384. <https://doi.org/10.1007/978-3-642-21869-9>
- Trivedi, S., Pardos, Z. A., Sárközy, G. N., & Heffernan, N. T. (2016). Spectral Clustering in Educational Data Mining. *EDM 2011*, (February), 129–138.
- Tsai, C. J., Tseng, S. S., & Lin, C. Y. (2001). A Two-phase fuzzy mining and learning algorithm for adaptive learning environment. *Computational Science-ICCS 2001*, 429–438.
- Ventura, S., Romero, C., & Hervás, C. (2008). Analyzing rule evaluation measures with educational datasets: A framework to help the teacher C3 - Educational Data Mining 2008 - 1st International Conference on Educational Data Mining, Proceedings. *1st International Conference on Educational Data Mining, EDM 2008*, 177–181. Retrieved from <http://www.scopus.com/inward/record.url?eid=2-s2.0-77955226888&partnerID=40&md5=272421f1ccafa9a684f3896ac40ca696>
- Virseda Benito, F., & Carrillo, J. (2008). Minería de datos y aplicaciones (p. 8).
- Vranić, M., Pintar, D., & Skočir, Z. (2007). The use of data mining in education environment. *9th International Conference on IEEE*, 243–250.

- Wang, F. (2002). On Using Data-Mining Technology for Browsing Log File Analysis in Asynchronous Learning Environment. *Conference on Educational Multimedia, Hypermedia and Telecommunications, 2005–2006*.
- Wang, F. H. (2008). Content recommendation based on education-contextualized browsing events for web-based personalized learning. *Educational Technology and Society, 11(4)*, 94–112.
- Wang, F. H., & Shao, H. M. (2004). Effective personalized recommendation based on time-framed navigation clustering and association mining. *Expert Systems with Applications, 27(3)*, 365–377. <https://doi.org/10.1016/j.eswa.2004.05.005>
- Wang, T., & Mitrović, A. (2002). Using neural networks to predict student's behaviour. *Proc. Int. Conf. Computers in Education, (October)*, 969–973. <https://doi.org/10.1109/CIE.2002.1186127>
- Wang, Y. H., Tseng, M. H., & Liao, H. C. (2009). Data mining for adaptive learning sequence in English language instruction. *Expert Systems with Applications, 36*, 7681–7686.
- Weibelzahl, S., Hurley, T., & Weibelzahl, S. (2007). Creating New Learning Experiences on a Global Scale. *Lecture Notes in Computer Science 4753, (October)*, 518. https://doi.org/10.1007/3-540-68339-9_34
- Wen-Shung Tai, D., Wu, H. J., & Li, P. H. (2008). Effective e-learning recommendation system based on self-organizing maps and association mining. *The Electronic Library, 26*, 329–344.
- Yohannes, Y., & Hoddinott, J. (1999). Classification and regression trees: an introduction, 1–29.
- Yu, P., Own, C., & Lin, L. (2001). On learning behavior analysis of web based interactive. *Conf. on Implementing Curricular Change in Engineering Education*, 1–10.
- Yukselturk, E., & Education, C. (2014). PREDICTING DROPOUT STUDENT: AN APPLICATION OF DATA MINING METHODS IN AN ONLINE EDUCATION PROGRAM, *17(1)*. <https://doi.org/10.2478/eurodl-2014-0008>
- Zaiane, O. R. (2002). Building a recommender agent for e-learning systems. *International Conference on Computers in Education, 2002. Proceedings., (October)*, 55–59. <https://doi.org/10.1109/CIE.2002.1185862>
- Zhang, L., Liu, X., & Liu, X. (2008). Personalized Instructing Recommendation System Based on Web Mining. *Young Computer Scientists, 2517–2521*.
- Zhang, Y., Oussena, S., Clark, T., & Kim, H. (2010). Using data mining to improve student retention in HE: A case study.
- Zheng, S., Xiong, S., Huang, Y., & Wu, S. (2008). Using methods of association rules mining optimization in web-based mobile-learning system. *In Electronic Commerce and Security, 967–970*.
- Zhou, M. (2010). Data Mining and Student e-Learning Profiles. *IEEE 2010 International Conference, (February)*, 5405–5408. <https://doi.org/10.1109/ICEE.2010.1352>