Journal of Engineering and Applied Sciences 12 (Special Issue 6): 7849-7858, 2017
ISSN: 1816-949X

A Bibliometric Analysis of the Learning Analytics Application in the Education

Victor Daniel Gil Vera
SISCO "Universidad Catolica Luis Amigo", Tv. 51a # 67B-90, Medellin, Colombia

Abstract: Data analytics allows to identify patterns, trends and to obtain valuable information in large volumes of information. The majority of algorithms used in data analysis usually are designed according to the capabilities of power and flexibility more than by its simplicity and are complex for that educators can use them. This study presents the justification of the need to adapt an algorithm for data analysis that allows to perform predictive learning analysis to university teachers. It was used the methodology of the Systematic Literature Review (SLR) which seeks to answer research questions of scientific research publications. It was concluded that it must implement, adapt or develop algorithms defaults to the educational context which might simplify the configuration and implementation of non-expert users.

Key words: Learning analytics, learning to learn, lifelong learning, multimodal learning analytics, data exchange, publications

INTRODUCTION

Learning analytics refers to the collection, analysis and reporting of data on students and their contexts in order to understand and optimize the learning and the environments in which they occur (Hernandez-Garcia *et al*., 2015; Lonn *et al*., 2015), this arises as an emerging discipline that seeks to improve teaching and learning through a critical evaluation of data and the generation of patterns related to the habits and responses of the students which allow you to provide timely feedback (Hernandez-Garcia *et al*., 2015).

Due to the constant generation of information in schools, colleges and universities related to the academic performance of students, it is necessary to adapt an algorithm for the analysis of data to the needs of teachers who do not have advanced skills in software programming. In the review of the state of the art, we used the methodology of the Systematic Literature Review (SLR), systematic, rigorous, orderly and auditable review whose goal is to answer research questions which takes as a basis research and scientific publications made before (Lonn *et al*., 2015). The research questions were:

- Q1: What are the seminals papers of the area?
- Q2: What are the most cited articles in the area?
- Q3: What are the most important journals in the area?
- Q4: What are the most important conferences in the area?
- Q5: Who are the most cited authors in the area and what are its main research themes?
- Q6: Which is the institution that more published on the topic?
- Q7: Which is the country that more published on the topic?
- Q8: What type of document you publish more about the subject?
- Q9: Which is the area of knowledge that more published on the topic?
- Q10: What year was published more about the subject?
- Q11: What are the main uses of the learning analytics?
- Q12: What is the main problem of the algorithms implementation for data analysis in education?
- Q13: What is the main problem of the algorithms implementation for data analysis in education?
- Q14: What are the possible solutions to the problem?
- Q15: What real cases of learning analytics application have been reported in higher education?

In the rest of the study describes some basic definitions and concepts, the protocol of the review, the search process (document selection, quality assessment and data extraction), results of the process of search, selection of documents and the evaluation of the quality, the analysis of the selected studies and discussion of the answers to the research questions. Finally, the study concludes.

Literature review: SLR are characterized by being focused to answer research questions clearly defined and explicit by using systematic methods to identify, select and critically evaluate relevant research of previously published studies related to a topic to investigate (Cruz-Benito *et al*., 2015; Conde and Hernandez-Garcia,

2015). SLR is a methodical way to identify, evaluate and interpret the empirical studies available on a topic, research question or phenomenon (Lonn *et al*., 2015). The phases or stages to perform a SLR are (Nafria *et al*., 2015; Fulantelli *et al*., 2015; Gasevic *et al*., 2016).

Identification of the research: The objective of a SLR is to find the greatest possible amount of primary studies related to a research question, on the basis of criteria and search parameters before making the final selection. The definition of these criteria is the main difference that distinguishes it from the traditional review of literature (Lonn *et al*., 2015).

Selection of studies: Once the potentially relevant primary studies, one must evaluate their actual quality to filter the studies that can really contribute to answer the question's of research (Lonn *et al*., 2015).

Data extraction: The objective of this stage is to design data extraction forms to accurately record the information that researchers obtain of the primary studies. The data extraction forms must be defined at the time of defining the study protocol (Lonn *et al*., 2015).

Synthesis: Is to collect and summarize the results of the selected primary studies. The synthesis must be descriptive (not quantitative). However, it is sometimes possible to complement a descriptive synthesis with a quantitative summary (Lonn *et al*., 2015).

MATERIALS AND METHODS

Planning of the review: This study presents the parameters used in the design of the research which were established by reference to the publications of (Freitas *et al*., 2015; Romero and Ventura, 2010). A search was made in the scientific databases: Scopus, IEEE Xplore, ISI web of knowledge assuming that there were previous publications on the subject of study. The keywords were obtained from the titles of articles and books specialized on learning analytics, subsequently obtained synonyms and similar words. After the first search were incorporated into the words that were not considered at the start. In the end, they formed the search strings with the combinations of words: "Learning analytics", "21st century skills", "multimodal learning analytics", "data sharing", "learning to learn", "lifelong learning", "transferable skills" and "complex systems" in the title, abstract and key words. The search strings used were:

- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (21st century skills))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (multimodal learning analytics))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (data sharing))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (learning to learn))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (lifelong learning))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (transferable skills))
- (TITLE-ABS-KEY (learning analytics) AND TITLE-ABS-KEY (complex systems))

Included articles, conference articles, books and book chapters published in Scopus, IEEE Xplore and ISI web of knowledge. Were discarded internet publications, conference abstracts and editorials. The publications found in the jerarquizaron review and the themes of each one in order to answer the research questions Q11-Q15.

RESULTS AND DISCUSSION

This study presents the analysis of the results found in the review which are focused on answering the research questions:

Q1; What are the seminals papers of the area? For the identification of these papers was used Tree of Science (ToS) intelligent web tool for the selection of scientific papers. The majority of the studies of Table 1, focus on analysing the main uses of the learning analytics in virtual environments.

Q2; What are the most cited articles in the area? The majority of papers of Table 2, focus on presenting the challenges and opportunities of the learning analytics which are related to the identification of different types of technologies, uses and applications in the pedagogy field.

Q3; What are the most important journals in the area? The journal which has the highest number of publications on the subject was "lecture notes in Computer Science including subseries lecture notes in Artificial Intelligence and Lecture notes in Bioinformatics" in this published the latest advances in research in all areas of computing.

Q4; What are the most important conferences in the area? The conference which has the highest number of publications on the subject was "Proceedings Frontiers in Education Conference Fie" international conference on educational innovations and research in Engineering and Computer Science (Table 3 and 4).

Table 1: What are the seminals papers of the area?

Seminal study	Years	Researchers	H index
Educational data mining: a survey from 1995-2005	2007	Romero and Ventura	29-33
Course V is: a graphical student monitoring tool for supporting instructors in web-based distance courses	2007	Mazza and Dimitrova	16-22
The state of educational data mining in 2009: a review and future visions	2009	Baker and Yacef	41-26
Educational data mining: a review of the state of the art	2010	Romero and Ventura	29-33
Mining LMS data to develop an "early warning system" for educators: A proof of concepts	2010	Macfadyen and Dawson	15-23
Learning analytics: the definitions, the processes and the potential	2011	Lias and Elias	0-0
Penetrating the fog: analytics in learning and education	2011	Siemens and Long	37-0
Learning analytics: drivers, developments and challenges	2012	Ferguson	17
A reference model for learning analytics	2012	Chatti, Dyckhoff, Schroeder and H. Thus	22 0 0 0
Can we predict success from log data in VLEs? Classification of interactions for learning analytics and their relation with performance in VLE-supported F2F and online learning	2014	Agudo-Peregrina, Iglesias-Pradas, Conde-Gonzalez and Hernandez-Garcia	6 9 32 13

Table 2: What are the most cited articles in the area?

Title	Researchers	Years	Cites
Social learning analytics	Shum and Ferguson	2012	164
Learning analytics: drivers, developments and challenges	Ferguson	2012	158
Cognitive style, gender and learning from multi-media materials in 11 years old children	Riding and Grimley	1999	83
A reference model for learning analytics	Chatti, Dyckhoff, Schroeder and thus	2012	77
Translating learning into numbers: a generic framework for learning analytics	Reserchers of Document Greller and Drachsler	2012	77
Mayday integrative analytics for expression data	Reserchers of Document Battke, Symons and Nieselt	2010	73
Design and implementation of a learning analytics toolkit for teachers	Dyckhoff, Zielke, Bultmann, Chatti and Schroeder	2012	70
Learning analytics dashboard applications	Verbert, Duval, Klerkx, Govaerts and Santos	2013	69
Learning analytics: the emergence of a discipline	Reserchers of Document Siemens	2013	58
Dataset-driven research to support learning and knowledge analytics	Verbert, Manouselis, Drachsler and Duval	2012	52
Learning analytics: ethical issues and dilemmas	Slade and Prinsloo	2013	46

Table 3: What are the most important journals in the area?

Journal	Amounts
Lecture notes in computer science including subseries lecture notes in artificial intelligence and lecture notes in bioinformatics	11
Computers and education	7
British journal of educational technology	5
Educational technology and society	4
Computers in human behavior	3
Education and information technologies	3
International journal of engineering education	3
International review of research in open and distance learning	3
American statistician	2
IEEE transactions on learning technologies	2

Table 4: What are the most important conferences in the area?

Conference	Amounts
Proceedings frontiers in education conference fie	26
Communications in computer and information science	22
Proceedings of the ACM SIGKDD international conference on knowledge discovery and data mining	20
Proceeding computer science	18
Proceedings of the ACM SIGMOD international conference on management of data	18
Proceedings of international conference of the learning sciences icls	17
Advances in intelligent systems and computing	16
Proceedings of SPIE the international society for optical engineering	16
IFIP advances in information and communication technology	13

Q5; Who are the most cited authors in the area and what are its main research themes? According to the results of the search, the researcher that more public on the subject is Pardo of the Sidney University, Australia. The main fields of research are focused on the creation of mechanisms of data capture for the analysis of the learning, fundaments, applications and trends (Table 5).

Table 5: Who are the most cited authors in the area and what are its main research themes?

Researchers	Citations	H index
Pardo	977	13
Dawson	836	14
Munoz-Merino	292	11
Gasevic	1811	23
Ferguson	361	7
Kloos	568	12
Drachsler	656	12
Ogata	1214	18
Duval	1734	22

Q6; Which is the institution that more published on the topic? The institution that more publishes on the subject is the Curtin University, Australia (Table 6).

Q7; Which is the country that more published on the topic? The majority of publications found in the review were made in EEUU.

Q8; What type of document you publish more about the subject? The types of documents that more has been

Table 6: Which is the institution that more published on the topic?

Institutions	Amounts
Curtin University	14
Open University	10
The University of Sydney	9
University of Toronto	8
Educational Testing Service	7
University of Technology Sydney	7
University of Melbourne	7
The University of British Columbia	6
University of Birmingham	6
Open University of the Netherlands	6

Table 7: Which is the country that more published on the topic?

Counties	Amounts
EEUU	178
Unided Kingdom	75
Australia	61
Canada	40
Germany	25
Spain	21
Indeterminate	18
Netherlands	17
Italy	11
France	10

Table 8: What type of document you publish more about the subject?

Document types	Amounts
Papers	154
Books	110
Conference proceedings	95
Book chapters	67
Review	16
Conferences review	4
Editorial	3
Notes	2
Total	451

published are articles and books in their great majority published by the Publishers pack and springer (Table 7).

Q9; Which is the area of knowledge that more published on the topic? The area of knowledge that registers the highest number of publications on the topic is computer sciences (Table 8).

Q10; What year was published more about the subject? With 160 publications, the year in which more it was published on the topic was 2016. Table 9 and 10 present the number of articles published per year.

In total were found 484 publications. The criteria for selecting the publications finally selected were: clarity, objectivity, relevance, novelty and that were consistent and help answer the research questions. In total, we analyzed 70 publications.

Q11; What are the main uses of the learning analytics? LA refers to the collection, analysis and presentation of data on students and their contexts in order to understand

Table 9: Which is the area of knowledge that more published on the topic?

Areas	Amounts
Computer Science	484
Social Sciences	147
Engineering	91
Math	69
Arts and Humanities	43
Decision Sciences	36
Psychology	36
Business, Management and Accounting	19
Medicine	19
Biochemistry, Genetics and Molecular Biology	7
Health Sciences	5
Neuro Sciences	5
Materials Science	4
Chemical Engineering	3
Planetary Science and Earth	3
Agriculture and Biological Sciences	2
Chemistry	2
Environmental Sciences	2
Physics and Astronomy	2
Economy, Econometrics and Finance	1
Energy	1
Total	981

Table 10: What year was published more about the subject?

Years	Amounts
2017	34
2016	160
2015	132
2014	74
2013	42
2012	20
2011	9
2010	4
2009	5
2008	0
2007	1
2006	1
2005	0
2004	0
2003	1
2002	1
2001	0
Total	484

and optimize the learning and the environments in which they produce (Drachsler and Greller, 2012). This area emerged during the last decade (Khousa *et al*., 2015), offers different types of computational support for the monitoring the students behavior, data management, patterns visualization of education for educators and students, allowing discover hidden trends, identify unknown correlations or other type of information (Cambruzzi *et al*., 2015). This combines data from institutions, predictive and descriptive models, statistical analysis, etc., to generate knowledge to students, teachers or administrators (Harrison *et al*., 2015; Aguiar *et al*., 2014). LA focuses specifically on the students and in their learning behaviors. This complies with the four characteristics of the data analysis, known as the 4 Vs: Volume, Velocity, Variety and Veracity (Asif *et al*., 2015).

Volume: Amount of all kinds of data from different sources in continuous expansion (Asif *et al*., 2015). Scores obtained from tests performed on educational management platforms (Moodle, Chamilo, Canvas, e- Doceo), number of entries, hours of entry, permanence time, etc.

Variety: Different types of structured and unstructured data collected through sensors, smart phones or social networks (Asif *et al*., 2015). In virtual education the volume of unstructured data, i.e., those who are not stored in a database exceeds the volume of structured data (Asif *et al*., 2015).

Velocity: Data transfer speed (Asif *et al*., 2015). The content of the data changes due to the constant arrival of data from multiple sources, the absorption of additional data collections and the introduction of previously archived data (Scheffel *et al*., 2012).

Value: Process of discovering hidden values in large data sets with different types and rapid generation (Grau-Valldosera and Minguillon, 2011). LA allows you to analyze the students behavior, detect those needing support, predict its performance, etc.

Some research shows the potential of the learning analytics in the prediction of academic performance (Manso-Vazquez and Llamas-Nistal, 2015; Sinclair and Kalvala, 2015) and propose methods for the groups formation of new students from the analysis of indicators of previous activities with the aim of achieving groupings optimal (Sinclair and Kalvala, 2015). The way in which the data results of test is presented to the students can generate an emotional effect on them and you can become a criterion for determining whether to continue or not in the university (Sinclair and Kalvala, 2015).

Other research exploring how to collect educational data in virtual worlds which can then be used to identify behavior of students and teachers (Kim *et al*., 2016). Exploratory analyses of data from the interactions in the virtual world provided information on the perceptions and motivations of the Educational Virtual Worlds (EVW) users. After this analysis, proposed lines of action aimed at increasing the efficiency of learning and facilitation of the adoption of the EVW (Pardo *et al*., 2015).

Other studies examine the difficulties associated with the decision-making in educational contexts (Piety *et al*., 2014) due to the complexity of having involved individual members and different actors with different objectives in learning activities. The e-Participation can help resolve this complexity. Other researches dealt with the problem of how the evidence analysis of learning on mobile devices where students not only interact with their peers and with the contents but also with its context (Hecking *et al*., 2014). Reseachers define a frame for the relations analysis between the different types of interactions that take place in a stage of mobile learning and of the tasks that are pedagogically excellent for the learning activity.

Other research analyzes the extent to which conditions instructional influence the prediction of academic success and conclude that it is essential to give an account of the different ways in which the technology is adopted and applied in different contexts (Tervakari *et al*., 2013). The new approaches to the data analysis are creating new ways to understand trends and students behaviors which can be used to improve the design of learning, strengthen the students retention, provide early warning signals and help customize the learning experience (Vozniuk *et al*., 2014). Table 11 summarizes some uses of the learning analytics in the education (Menchaca *et al*., 2015).

Q12; What is the main problem of the algorithms implementation for data analysis in education? According to Vahdat *et al*. (2015), algorithms for the learning analytics have not been developed and the existing (trees/rules of decision, regression, clustering, descriptive statistics, visualization/time series/analysis of sequences association rules assembly, text mining methods, support vectors, etc.) can be useful only to developers. In addition, there is no general tools that can be applied to any educational system. Due to the foregoing, algorithms/models/tools are needed predesigned for the educational context, allowing to normalize input and output data (preprocessing-postprocessing), use semantic information and integrate the domain of the educational knowledge. In the review of the state of the art identified the different techniques used in LA, their main applications and some examples of researchs (Table 12). According to 63 the top 10 of algorithms most used by the scientists of information are Fig. 1.

Decision trees/rules: Supervised and non parametric method (there is a priori knowledge) used in the classification and regression learning. Allows you to create models to predict the value of a destination variable using decision rules learning inferred from the characteristics of the data. These allow you to make predictions of belonging to a group using statistical criteria. Some algorithms used are: ID3, C4.5, C5.0 CART, random forest and trees of conditional inference.

Table 11: Uses of the learning analytics in the education

Agents	Uses
Students	To personalize electronic learning (e-Learning); to recommend activities, resources and learning tasks to the students, so that, they improve moreover its learning; to suggest interesting learning experiences; to suggest linkage, to generate adaptable suggestions to recommend courses and pertinent debates, to improve the processes and learning patterns
Teachers	To analyze the students behavior; to detect the students who need support; to predict the students yield; to classify the students under groups; to find regular and irregular patterns; to find the errors of major frequency; to determine more effective activities; to improve the adaptation and personalización of courses; to identify possible candidates for deserting. To improve the education processes, to evaluate the academic progress, to predict of the future yield and to detect potential problems. To evaluate programs and institutions. To understand the learning behaviors in order to execute suitable interventions
Developers of courses	To evaluate and to maintain educational software; to improve the learning ability of the students; to evaluate the structure of the courses and its efficacy in the learning process; to compare informational advancing skills to be able to recommend the most useful for every specific task, etc.
Organizations	To improve the decision making processes in the higher education institutions to optimize the efficiency and to achieve specific targets; to suggest certain courses that might be valuable for every class of students; to find the most profitable way of improving the student retention; to help to select in the admission processes, etc.
Administrators	To develop the best way of organizing the institutional resources (human and material) and its educational offer; to use more efficiently the available resources; to improve the educational programs and to determine the efficacy of the approach of learning over a distance; to evaluate the teachers and the plans of study; to establish parameters to improve the efficacy of the web sites and its adaptation to the users. To improve the briefcase of courses, decision making of marketing, hirings, evaluations of performance, etc.

Table 12: Different techniques used in LA, main applications and some examples of researchs

Techniques	Uses/Applications	Examples
Prediction	To predict the performance and detect behaviors of students	28, 29 and 30
Clustering	To group similar materials or students based on their patterns of learning and interaction	31, 32 and 33
Detection of outliers	To detect students with difficulties or irregular learning processes	34, 35 and 36
Relation mining	To identify relations in standards of behavior of the students and difficulties	37, 38 and 39
Social network analyses	To interpret the structure of the relations in the activities of collaboration and interaction with hardware of communication	40, 41 and 42
Mining process	To analyze the behavior of the students	43, 44 and 45
Text mining	To analyze the content of the forums, chats, web pages and documents	46, 47 and 48
Distillation of data	To help the teachers to visualize and analyze the activities during the students and the use of the information	31, 49 and 50
Discovery with models	To identify the relations between the behaviors of the student and the characteristics or contextual variables	51, 52 and 53
Gamification	To include possibilities of playful learning to maintain the motivation for example, the integration of the achievements, experience points or ensigns as success indicators	54, 55 and 56
Machine learning	To find secret ideas in information automatically (based on models that are exhibited to new information and adapt themselves independently)	57, 58 and 59
Statistics	To analyze and to interpret quantitative information for the decision making	60, 61 and 62

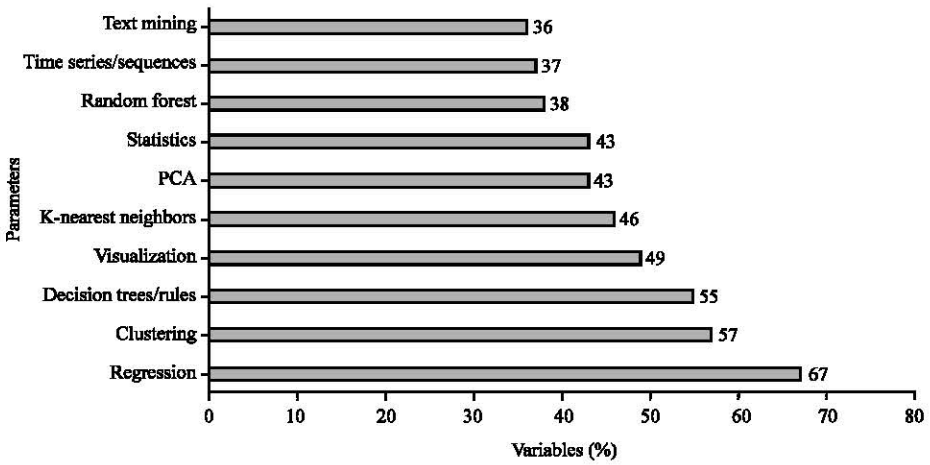

Fig. 1: Data analytics algorithms KD nuggets

ID3: It is one of the algorithms more simple to produce decision trees with categorical attributes and classes. It is based on a measurement called gain ratio (difference in the uncertainty that is obtained through a partition of the data) to build the trees. The objective is to maximize the predictive power of the tree by reducing the uncertainty in the data. Is available to users of R with the package data.tree (install.packages("data.tree"); library(data.tree).

C4.5: It works in a similar way to the ID3 but it uses the profit relation as a division criterion. Accepts the partition in numeric attributes that divided into categories. The value of the division is selected with the aim of reducing the entropy for the attribute. In contrast to the ID3, the C4.5 allows the pruning which is basically the low part up to the simplification of the tree to avoid the overfitting of the training information. Is available to users of R with the package RWeka (install.packages ("RWeka"); library(RWeka)).

C5.0: It is an enhanced version of the C4.5 includes push and winnowing. The objective of push is to increase the reliability of the predictions by performing the analysis iteratively and adjusting weights of the observations after each iteration. Winnowing refers to the removal of useless for the main analysis attributes. Is available to users of R with the package C50 (install.packages("C50"); library(C50)).

CART: Uses different statistical criteria to decide on the divisions of the tree. The attribute to make the partition is selected with the Gini index as a decision criterion. In classification trees, the Gini index is calculated as:

$$1 - \sum_{j=1}^{c} p_j^2 \quad (1)$$

where, p is the probability of each possible partition on the attribute. Is available to users of R with the package rpart (install.packages("rpart"); library(rpart)). CART allows to predict numeric results.

Random forest: Use the set of learning (combination of trees cart) to improve the classification through a voting principle. This uses an similar algorithm to CART. Differences include the use of packaging and selection of predictors randomly in each partition of the tree. The packaging is intended to reduce the impact of the measurement error (noise) in the data and therefore, avoid the overflow. It is available to users of R with the package randomForest (install.packages("randomForest"); library (randomForest)).

Conditional inference trees: In contrast to the previous, this depend of the statistical significance in the selection of attributes on which the division is realized. This algorithm first searches the attributes that predict significantly the class in a hypothesis test that can be selected in the function call. Is available to users of R with the package partykit (install.packages(c("Formula", "partykit"))library(Formula);library(partykit)).

Regression: Method used in prediction whose objective is to try to predict an output of real value (numeric value) of some variable. The types of regression models can be: simple linear regression, multiple or straight of regression. The first is used to examine as an attribute affects another, the second to assess as multiple attributes affect another. Mathematically linear regression models can be expressed as:

$$Y_t = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \cdots + \beta_p X_{p1} + \varepsilon \quad (2)$$

Where:

Y_t = The dependent variable
$X_1, X_2, ..., X_p$ = The independent variables
$\beta_1, \beta_2, ..., \beta_p$ = Parameters
ε = The error

The models of simple linear regression analyzes the relationship between a predictor (cause) and a criterion (result). There are two very important parameters that are the result of a regression analysis, the intercept (average value of the criterion when the predictor is 0) and the value of the slope (indicates how many units on average, the criterion changes when the predictor increases by one unit). The multiple linear regression models, measure the impact of several predictors on a single criterion instead of only one in simple regression. The value of the observations can be calculated as the intersection over the slope coefficient multiplied by the predictive factor over the residue.

Clustering: Unsupervised learning method used to analyze data that do not include pre-defined classes. The data instances are grouped by maximizing the similarity between classes equal and minimizing the similarity between different classes. The clustering algorithm allows you to identify and group the instances that are very similar as opposed to ungroup the instances that are less similar to each other. A cluster is a group of cases, observations individuals or other units which are similar to each other in the characteristics considered. The clustering algorithms using measures of distances between the cases in order to create homogeneous groups of cases. Some of the methods used are: hierarchical (aglomerativo), allocation of points, cohesion, K means, grouping population, algorithm BFR (Bradley Fayyad Reina). Is available to users of R with the package cluster.

Descriptive statistics: They allow you to describe quantitatively a compilation of information and summarize through a single number (mean, median, mode, probability distributions, covariance, correlation, etc.) to a data set.

Visualization: Technique used to communicate data or information through the codification of visual objects (points, lines or bars) contained in the graphics. The objective is to communicate information in a clear and efficient manner to the users through graphic media Some of the methods used are sampling, shuffling, sorting, maze generation, using vision to think.

Time series: They allow to analyze the behavior of a variable in the time, facilitating decision-making. Take as a basis historical data to obtain a statistical description of the existing links in the past and present, they allow considering temporal information, estimating parameters and provide accurate timely data on future values. Some of the techniques used are ARIMA, RNA, SVM, Genetic algorithms for the forecasting with time series.

Association rules: Technique to discover how items are associated with each other. There are three common ways to measure the association, support, confidence, lifting. There are three algorithms for searching for association rules in databases: Apriori, Partition and Eclat.

Emsemble methods: Technique to combine multiple learning algorithms weak in an attempt to build a learning algorithm more powerful, combine multiple scenarios to form a best case scenario. Some of the techniques used are boosting, bagging, combination methods, diversity, ensemble pruning.

Text mining: The text analysis or text mining is the natural and essential extension of predictive analytics and data science. Some of the methods used are classification associations discovery, grouping.

Support vectors: Automatic learning supervised algorithm that can be used for both the challenges of classification or regression. However, it is mostly used for classification problems. In this algorithm, we represent graphically each data item as a point in the n-dimensional space (where n is the number of features) with the value of each feature to be the value of a given coordinate. It then performs the classification through the search of the hyper-plane which differentiate the two classes very well. The support vectors are simply the coordinates of individual observation. Machine support vector is a border that is best segregates the two classes (hyper-flat/line).

Q13; What is the main problem of the algorithms implementation for data analysis in education? The users have to provide the values adapted to the parameters to obtain good results and models, therefore they must possess a certain knowledge quantity specialized to be able to find the correct configuration. The users have to select the method or specific algorithm that they want to use of a wide range of methods and available algorithms of data analysis.

The majority of the algorithms for the analysis of data need to be configured before they are executed and are not designed for users who are not experts in programming.

Q14; What are the possible solutions to the problem? To implement/to adapt/to develop algorithms predetermined for the educational context to simplify the configuration and execution to not expert users. Also, software must be constructed for the educational analysis of information with an intuitive interface with visualization facility to realize its significant results for educators and designers of e-Learning, also, it is important to develop specific pre-prosecution software in order to automate and to facilitate all the tasks or functions of pre-prosecution that at present the users of educational data analysis must do manually.

Q15; What real cases of learning analytics application have been reported in higher education?

Purdue University: Implemented a system of signs and interventions to improve the academic students success which increased retention and graduation rates. Signals system was intended to help students understand their progress early enough, so that, they were able to seek help.

Maryland University: In this university, there was a growing use of Virtual Learning Environments (VLE) but there was no evidence that these improve learning. He was a project of institutional research for correlations between data of the VLE and the students scores and to know how to use better predictions to support students. The research also, allowed analysis of possible interventions and how to identify effective educational practices from the data.

New York Technology Institute: They developed their own predictive model in collaboration with staff from expert advice to identify at-risk students with a high degree of accuracy. The objective was to increase the retention of students in the first year of their studies through the creation of a model to identify those most in need of support and to provide information about the situation of each student that could help teachers in their work.

Marist College: Developed a predictive model of open source code to generate early alerts which has been transferred to different institutions, allowing you to define intervention strategies to help students at risk.

Cowan University: They implemented the C4S technology which automatically identifies students who need support. This technology allows teachers have contact with a large number of students and manage a series of interventions for each student.

New England University: They developed a dynamic, systematical and automated process that captures the students welfare state which capture as base three main components: e-Motion (emoticons to express the emotional state, happy, neutral, unhappy, unhappy), the Vibe (Text boxes that meet the student comments) and the el AWE (Automated Engine that analyzes the information proceeding from the corporate systems and suggests risk behaviors).

Open University: It implemented a strategic analysis program of learning to improve the students success by means of the incrustation of the decision making based on the evidence on all the levels. This university has come developing its institutional aptitudes to strengthen the analysis of learning which are organized concerning three main axes: availability of information, analysis and creation of knowledge and processes that influence the success of the students.

Nottingham Trent University: It implemented processes to improve the retention, to increase the belonging sense and to improve the institutional results.

Universidad de Wollongong: They implemented processes that link the use of social networks to the teaching practices (SNAPP: Social Networks Adapting Pedagogical Practice). These processes allow to analyze discussions in line, discover patterns in forums which are moderated by an instructor, the interventions realized by the instructor can have a significant impact in the experience of the students learning.

Open Universities of Australia: To be institutions that provide education to distance, adapted a system to perform customized planning for each student which proposes alternative modules and virtual support according to the needs of the same.

CONCLUSION

Currently, strategic and systematic management of large and complex databases is allowing to solve problems arising in the education. The use of different computational techniques for the data analysis; computational statistics, data mining, research of operations, applied mathematics, optimization, statistical, machines learning, modeling strategic, among others can add value to the data and generate advantages in the academic and business world.

Due to the vertiginous information increase in the educational sector and to the intensive rate of growth of the virtual education, it becomes necessary to increase the number of qualified experts in data analysis, especially in latin America countries. In the education, the learning analytics is constituted like a tool of big utility to improve the academic quality and the education learning processes.

REFERENCES

Aguiar, E., N.V. Chawla, J. Brockman, G.A. Ambrose and V. Goodrich, 2014. Engagement vs performance: Using electronic portfolios to predict first semester engineering student retention. Proceedings of the 4th International Conference on Learning Analytics and Knowledge, March 24-28, 2014, ACM, New York, USA., ISBN:978-1-4503-2664-3, pp: 103-112.

Asif, R., A. Merceron and M.K. Pathan, 2015. Investigating performance of students: A longitudinal study. Proceedings of the 5th International Conference on Learning Analytics and Knowledge, March 16-20, 2015, ACM, New York, USA., ISBN:978-1-4503-3417-4, pp: 108-112.

Cambruzzi, W.L., S.J. Rigo and J.L. Barbosa, 2015. Dropout prediction and reduction in distance education courses with the learning analytics multitrail approach. J. UCS., 21: 23-47.

Conde, M.A. and A. Hernandez-Garcia, 2015. Learning analytics for educational decision making. Comput. Hum. Behav., 47: 1-3.

Cruz-Benito, J., R. Theron, F.J. Garcia-Penalvo and E.P. Lucas, 2015. Discovering usage behaviors and engagement in an educational virtual world. Comput. Hum. Behav., 47: 18-25.

Drachsler, H. and W. Greller, 2012. The pulse of learning analytics understandings and expectations from the stakeholders. Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, April 29-May 2, 2012, ACM, New York, USA., ISBN:978-1-4503-1111-3, pp: 120-129.

Freitas, S., D. Gibson, D.C. Plessis, P. Halloran and E. Williams *et al*., 2015. Foundations of dynamic learning analytics: Using university student data to increase retention. Br. J. Educ. Technol., 46: 1175-1188.

Fulantelli, G., D. Taibi and M. Arrigo, 2015. A framework to support educational decision making in mobile learning. Comput. Hum. Behav., 47: 50-59.

Gasevic, D., S. Dawson, T. Rogers and D. Gasevic, 2016. Learning analytics should not promote one size fits all: The effects of instructional conditions in predicting academic success. Internet Higher Educ., 28: 68-84.

Grau-Valldosera, J. and J. Minguillon, 2011. Redefining dropping out in online higher education: A case study from the UOC. Proceedings of the 1st International Conference on Learning Analytics and Knowledge, March 01, 2011, ACM, New York, USA., ISBN:978-1-4503-0944-8, pp: 75-80.

Harrison, S., R. Villano, G. Lynch and G. Chen, 2015. Likelihood analysis of student enrollment outcomes using learning environment variables: A case study approach. Proceedings of the 5th International Conference on Learning Analytics and Knowledge, March 16-20, 2015, ACM, New York, USA., ISBN:978-1-4503-3417-4, pp: 141-145.

Hecking, T., S. Ziebarth and H.U. Hoppe, 2014. Analysis of dynamic resource access patterns in a blended learning course. Proceedings of the 4th International Conference on Learning Analytics and Knowledge, March 24-28, 2014, ACM, New York, USA., ISBN:978-1-4503-2664-3, pp: 173-182.

Hernandez-Garcia, A., I. Gonzalez-Gonzalez, A.I. Jimenez-Zarco and J. Chaparro-Pelaez, 2015. Applying social learning analytics to message boards in online distance learning: A case study. Comput. Hum. Behav., 47: 68-80.

Khousa, E.A., Y. Atif and M.M. Masud, 2015. A social learning analytics approach to cognitive apprenticeship. Smart Learn. Environ., 2: 1-23.

Kim, J., I.H. Jo and Y. Park, 2016. Effects of learning analytics dashboard: Analyzing the relations among dashboard utilization, satisfaction and learning achievement. Asia Pac. Educ. Rev., 17: 13-24.

Lonn, S., S.J. Aguilar and S.D. Teasley, 2015. Investigating student motivation in the context of a learning analytics intervention during a summer bridge program. Comput. Hum. Behav., 47: 90-97.

Manso-Vazquez, M. and M. Llamas-Nistal, 2015. A monitoring system to ease self-regulated learning processes. IEEE. Iberoam. Mag. Learn. Technol., 10: 52-59.

Menchaca, I., M. Guenaga and J. Solabarrieta, 2015. Project-Based Learning: Methodology and Assessment Learning Technologies and Assessment Criteria. In: Design for Teaching and Learning in a Networked World, Conole, G., T. Klobucar, C. Rensing, J. Konert and E Lavoue. (Eds.). Springer, Cham, Switzerland, ISBN:978-3-319-24257-6, pp: 601-604.

Nafria, J.M.D., J.A. Cendon and L.P. Alonso, 2015. Building up eParticipatory decision-making from the local to the global scale: Study case at the European higher education area. Comput. Hum. Behav., 47: 26-41.

Pardo, A., N. Mirriahi, S. Dawson, Y. Zhao and A. Zhao *et al*., 2015. Identifying learning strategies associated with active use of video annotation software. Proceedings of the 5th International Conference on Learning Analytics and Knowledge, March 16-20, 2015, ACM, New York, USA., ISBN:978-1-4503-3417-4, pp: 255-259.

Piety, P.J., D.T. Hickey and M.J. Bishop, 2014. Educational data sciences: Framing emergent practices for analytics of learning, organizations and systems. Proceedings of the 4th International Conference on Learning Analytics and Knowledge, March 24-28, 2014, ACM, New York, USA., ISBN: 978-1-4503-2664-3, pp: 193-202.

Romero, C. and S. Ventura, 2010. Educational data mining: A review of the state of the art. IEEE. Trans. Syst. Man Cybern. Appl. Rev., 40: 601-618.

Scheffel, M., K. Niemann, D. Leony, A. Pardo and H.C. Schmitz *et al*., 2012. Key Action Extraction for Learning Analytics. In: Technology Enhanced Learning, Ravenscroft, A., S. Lindstaedt, C.D. Kloos and D. Hernandez-Leo (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-33262-3, pp: 320-333.

Sinclair, J. and S. Kalvala, 2015. Engagement Measures in Massive Open Online Courses. In: Learning Technology for Education in Cloud, Uden, L., D. Liberona and T. Welzer (Eds.). Springer, Cham, Switzerland, ISBN:978-3-319-22628-6, pp: 3-15.

Tervakari, A.M., J. Marttila, M. Kailanto, J. Huhtamaki and J. Koro *et al*., 2013. Developing learning Analytics for TUT Circle. In: Open and Social Technologies for Networked Learning, Ley, T., M. Ruohonen, M. Laanpere and A. Tatnall (Eds.). Springer, Berlin, Germany, ISBN:978-3-642-37284-1, pp: 101-110.

Vahdat, M., L. Oneto, D. Anguita, M. Funk and M. Rauterberg, 2015. A Learning Analytics Approach to Correlate the Academic Achievements of Students with Interaction Data from an Educational Simulator. In: Design for Teaching and Learning in a Networked World, Conole, G., T. Klobucar, C. Rensing, J. Konert and E Lavoue (Eds.). Springer, Berlin, Germany, ISBN:978-3-319-24257-6, pp: 352-366.

Vozniuk, A., A. Holzer and D. Gillet, 2014. Peer assessment based on ratings in a social media course. Proceedings of the 4th International Conference on Learning Analytics and Knowledge, March 24-28, 2014, ACM, New York, USA., ISBN:978-1-4503-2664-3, pp: 133-137.