EasyChair Preprint

№ 3502

Development and Application of Academic Analytics in a School

Daniel Florez Rodriguez, Jenifer Fontalvo Aragon,
Jorge Enrique Portella Cleves and Luis Fernando Briceño Clavijo

May 30, 2020

DEVELOPMENT AND APPLICATION OF ACADEMIC ANALYTICS IN A SCHOOL

Daniel Florez Rodriguez
Jenifer Fontalvo Aragon
Luis Fernando Briceño Clavijo
Jorge Enrique Portella

daniel.florezr@cun.edu.co
jenifer.fontalvo@cun.edu.co
luis.briceno@cun.edu.co
Jorge_portella@cun.edu.co

National Unified Higher Education Corporation-CUN
Bogota - Colombia

Abstract

This document relates the data analysis that comes from academic thinking, as well as the data collection that generates potential tools. There is a growing interest in data mining and educational systems, which makes educational data mining a new growing research community, allowing the opportunity to track and store student learning activities as large data sets in online environments, making us aware of an intelligent use of the data produced by the academic environment, which allows us understand and predict the processes involved, as well as optimize the environments in which such learning occurs.

INTRODUCTION

This document relates the analysis of data that comes from academic thinking, as well as the collection of data that generates potential tools. There is a growing interest in data mining, and education systems, which makes educational data mining a new growing research community, allowing us to provide the opportunity to track and store students' learning activities as large data sets in online environments, making us aware of an intelligent use of data produced by the academic environment , which allows us to understand and predict the processes involved, as well as optimize the environments in which such learning occurs.

DEFINING THE PROBLEM

It is desired to carry out a comprehensive monitoring of a web application, which aims to be the point of communication between a school and its students, where the main factor, are the verification of virtual activities, evaluations and workshops of teachers, to mitigate the crisis that is lived by covid-19, and not lose the resources or materials of education.

GENERAL OBJECTIVES

1. Understanding how students are learning and optimizing learning processes

- Identify what elements and processes are being implemented in institutions for the learning process.

- Study student behavior in the learning environment.

- Check if the tools and processes used in learning work in the medium and long term

2. Extracting active values from Big Data using Predictive Modeling

- Consider which type of model or predictive models will be implemented for the extraction of active values.

- Check whether the extracted assets are accurate for future use.

- Understand what kind of data analysis the extracted assets are.

3. Apply Business Intelligence to education at the institutional, regional and national levels.

- Look for relevant methods so that the application of business intelligence is taken in the most productive way.

- Obtain a favorable result by applying this business strategy, allowing education to take new positions.

- Conduct a study on learning levels, to ensure that they are equitable for different populations and to technify the basic concepts in such a way that everyone is at the same level and thus move forward

Learning Analytics is an emerging field where analytical tools with a higher level of sophistication are implemented, in order to improve the level of learning and education. This technology is based on and closely linked to a number of other fields of study including business intelligence, web analytics, academic analysis, educational data mining, and action analytics. Business Intelligence is a well-established process in the business world because thanks to this decision-makers (manufacturers) can integrate strategic thinking with information technologies, being able to synthesize large amounts of data into powerful decision-making capabilities.

Web analytics is defined as the collection, presentation and analysis of reports generated from the use of a website by visitors and customers entering this site, in order to better understand and understand the effectiveness of all initiatives implemented online and other changes made to the website in an objective manner , scientific through experimentation, testing and measurement.

Stakeholders (Who?)

The Learning Analytics app can be targeted at different stakeholders, including students, teachers, mentors, educational institutions (administrators and faculty decision makers), researchers and system designers with different perspectives, objectives and expectations of Learning Analytics exercise. Students will likely be interested in how analytics might improve their grades or help them build their personal learning environments. Teachers may be interested in how analytics can increase the effectiveness of their teaching practices or help them tailor their teaching offerings to students' needs. Educational institutions can use analytical tools to support decision-making, identify potential students "at risk" to improve student success (i.e. student retention and graduation rates).

Objectives (Why?)

There are many objectives in Learning Analytics according to the particular point of view of the different stakeholders Possible objectives of Learning Analytics include monitoring, analysis, prediction, intervention, mentoring/tutoring, evaluation, feedback, adaptation, personalization, recommendation and reflection.

• **Monitoring and analysis:** In monitoring, the goals are to track student activities and generate reports to support decision-making by the teacher or school. Monitoring is also related to instruction design and refers to learning assessment.

• **Prediction and intervention:** In prediction, the goal is to develop a model that tries to predict student knowledge and future performance, based on their current activities and achievements This predictive model can be used to provide proactive intervention for students who may need additional assistance.

• **Tutoring- Tutoring:** Tutoring is primarily concerned with helping students with their learnings (tasks), often very domain-specific and limited to the context of a course. A tutor, for example, supports students in their orientation and introduction into new learning modules as well as instructions for specific thematic areas within a course. In tutoring processes the control is with the tutor and the focus is more on the teaching process. In contrast, tutoring goes beyond tutoring and focuses on supporting the student throughout the process, ideally throughout life, but actually limited to the time when both the mentor and the student are part of the same organization

• **Evaluation and feedback:** The goal is to support the (auto) evaluation of improvements in efficiency and effectiveness of the learning process. It's also important to get smart feedback from both students and teachers. Smart feedback provides interesting information.

• **Adaptation**: The adaptation is activated by the

Teacher/tutoring system or the educational system institution. The goal of Learning Analytics here is to tell students what to do next by adaptively organizing learning resources and educational activities according to the individual student's needs.

• **Customization and Recommendation:** In personalization, Learning Analytics is highly student-centered, focusing on how to help students decide on their own learning and continuously shape their PLE to achieve their learning goals. A PLE-based learning approach suggests a shift in the emphasis from a knowledge-based learning model to a knowledge-based learning model.

• **Reflection:** Analytics can be a valuable tool to promote reflection, students and teachers can benefit from comparative data within the same course, between classes or even between institutions drawing conclusions and (auto) reflecting on the effectiveness of their learning or teaching practice. Learning by reflection (or thoughtful learning) offers the ability to learn by returning and evaluating past work and personal experiences to improve future experiences and promote continuous learning

I USE DATA MINING IN LEARNING ANALYTICS

Learning Analytics uses educational methods of data mining to analyze large data sets. Within data mining researchers currently use several methods that are popular, these are classified into five categories, which are prediction grouping, mining relationships, discovery with models and separation of data for use in the process of human interest of which the last two mentioned are important in the education environment.

Prediction: This method involves developing a model that uses both a predicted variable and preaching variables. Predicted variables respect a specific component of the data, on the contrary, predicate variables consist of a combination of other data elements. This method can be used in education to predict the behavioral outcomes of those generated by students. The study of Baker, Gowda, Corbett - Automatic detection of student readiness for future learning is known: help use is key (Baker, Gowda and Corbett, 2011). A tool is created to automatically predict a student's future performance based on establishing positive or negative correlations between various features such as: student test results, time spent in response, time elapsed between receiving a track and writing the response, and others.

(Baker, Gowda and Corbett, 2011).

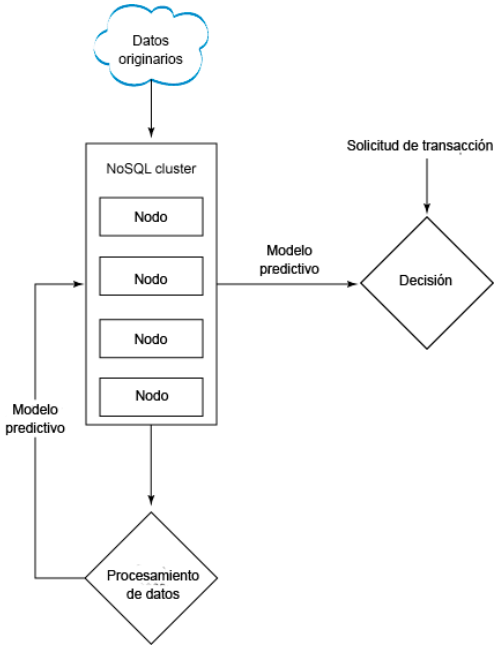

Fig. 1 Structure based on the predictive model

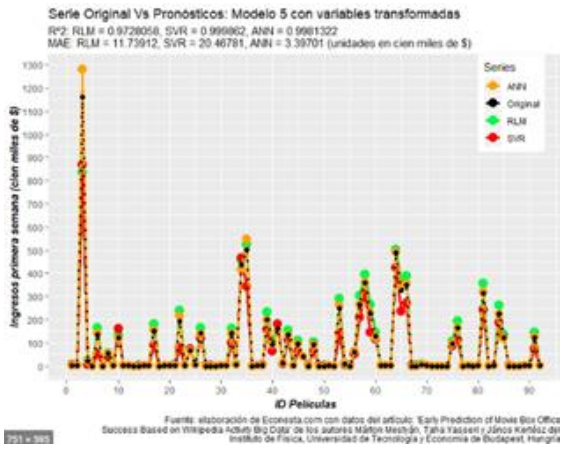

Fig. 2 Linear regression

Grouping: This method involves finding a set of data points that make up a logical group together. For this reason the observation reveals the resulting form. The method consists of constructing patterns that identify grouping data after certain similarities. For the model to provide quality predictions, the similarities within the class must be maximized and similarities between classes minimized. We could say that the implementation of this method in secondary education can give an orientation to the students of grouping according to the learning style of each student either (auditory, visual, practical - kinesthesia) based on behavioral analysis.

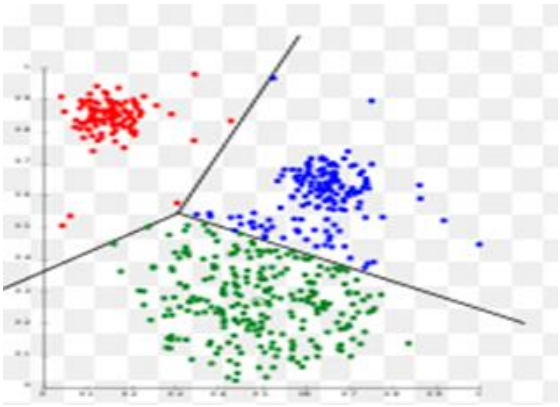

Fig. 3 Clustering of data by clustering

Relationship mining: The relationship mining method focuses on the goal of discovering relationships between variables in a set composed of a large number of variables, Relationship mining forms can include learning which variables are related to a single variable or discovering the strongest relationship between two variables. It is a method that uses algorithms to find association rules to detect, for example, mistakes made by students in solving a series of exercises. Based on the associations made, a certain behavior of the student can be predicted based on the hypothesis of solving the problem from which it starts.

Discovery with models: It is one of the most implemented and used procedures for data exploitation in the education environment. It is considered in the implementation of a previously validated model as a component in analyses that use prediction grouping or knowledge engineering. Knowledge engineering uses human reasoning for model development. In this way, information on educational materials that contribute the most to the progress of education can be obtained.

Distillation of data for human judgment: This method includes statistics and visualization techniques that help and allow people to understand data analysis. The method is the basis for creating many useful tools that provide clear analysis that can be quickly understood by unrelated users. This method has two different identification and classification objectives. the importance of distilling data for identification when data visualization allows for easy identification of known patterns that can be difficult to formally express. The analysis

provides real-time learning about learning performance indicators.

II DEVELOPMENT TO IMPLEMENT DATAMINING

The different structures and methodologies in the management of data, make the development of datamining, easier to implement in companies as well as educationally.

- On the other hand, thanks to the significant extraction of large data processes, it is possible to identify and reveal business ideas through the different behavioral modeling and correlations that allow the user to interact with predictions that finally solve, a myriad of real-life problems thus providing a higher value advantage.
- In the educational field, the analyses that are done exhaustively for data collection help to give, accurate measurements that to the future allow to know exactly the behavior with respect to the academic performance of the data.
- Through the social networks between schools, it is possible to create an analysis, with the preference decisions of each user, which help in a concluding the tastes of each person, and the analysis of the social group where they frequent or surround themselves most.
- With datamining, such complete applications can be made, that they will allow the construction of courses, qualification methodologies, organization of events, communications with teachers and students, and even the planning and analysis of each process and delivery.
- With the construction of automated codes and processes, it is possible for the user to identify with the interface of the applications, that in one way or another a communication is more user-friendly is why additional packages are easily installable.
- With the help of data processing, it is possible to communicate with the files and databases, which facilitates the quick modification of any errors or updates that are in turn required.

- You can generate the export of results for analysis, in a simple way and in various formats..

- The availability of the graphical environment makes user interaction more enjoyable, in a way that is more understandable at every step or operation that is required

- With measured use and working hand-in-hand with the internet, it is intended that in the database students have accessibility to large

- With measured use and working hand-in-hand with the internet, it is intended that in the database students have accessibility to large

III BUILDING A MODEL

- It is indispensable for data construction, to have clear knowledge about data mining.

- It must be clear where the root of the data comes from, i.e. its origin, because it will help to identify what is being worked and you want to build it.

- With utility tools based on the help of complex problems that the user has, or the program seeks the machine learning of the person who uses it, tools such as, Rstudio and Rattle help statistical computing and graphs that include a console and text editor that concisely support the execution of the code.

- Then, we proceed to the preparation of the data, which is constituted in the debugging or cleaning of the data, its organization and finally its processing.

- Models are constructed with the help of decision-making in a way that allows data to be classified and aggregated into models that interact in development.

Id	Nombre de tarea	Duración	Comienzo	Fin	Predeces
1	**TFG - Educational Data Mining & Learning Analytics**	**85 días**	**jue 27/02/14**	**mié 25/06/14**	
2	**Plan de Trabajo**	**6 días**	**jue 27/02/14**	**lun 10/03/14**	
3	Definición y Objetivo	6 días	jue 27/02/14	lun 10/03/14	
4	**Tratamiento de Datos**	**25 días**	**mar 11/03/14**	**lun 14/04/14**	2
5	Selección y Extracción de Datos	3 días	mar 11/03/14	jue 13/03/14	
6	Exploración de los Datos	6 días	vie 14/03/14	vie 21/03/14	5
7	Limpieza de Datos	8 días	lun 24/03/14	mié 02/04/14	6
8	Transformación de Datos	8 días	jue 03/04/14	lun 14/04/14	7
9	**Procesamiento de Datos**	**26 días**	**mar 15/04/14**	**mar 20/05/14**	
10	Selección del Modelo	6 días	mar 15/04/14	mar 22/04/14	
11	Elaboración del Modelo	10 días	mié 23/04/14	mar 06/05/14	10
12	Verificación y Prueba de Contraste	10 días	mié 07/05/14	mar 20/05/14	11
13	**Liurament Final y Presentación Virtual**	**14 días**	**mié 21/05/14**	**lun 09/06/14**	
14	Preparación Entrega Final	5 días	mié 21/05/14	mar 27/05/14	
15	Preparación Presentación Virtual	5 días	mié 28/05/14	mar 03/06/14	14
16	Revisión	2 días	mié 04/06/14	jue 05/06/14	15
17	Referencias y Bibliografía	2 días	vie 06/06/14	lun 09/06/14	16
18	**Tribunal**	**2 días**	**mar 24/06/14**	**mié 25/06/14**	13
19	Tribunal	2 días	mar 24/06/14	mié 25/06/14	

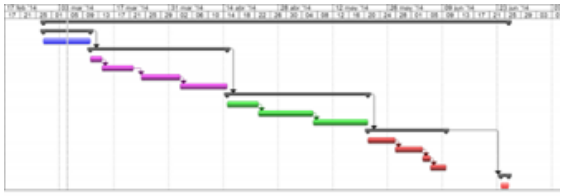

Fig. 4 Planning diagram

- A larger-scale project was considered among Educational Data Mining & Learning Analytics, which aims to obtain information from the different learning environments that generate information about students and also help the target treatment of data. The project consists of the fundamental pillars for the creation of methodologies that, combined with machine learning, form a complex and very complete application for the educational process.

- Work plan, contemplates the design of a working group that verifies that the tasks in the work plan are fulfilled by each member of the group, it should clarify all the processes that are already proposed, and by making decisions proceed to the processing of data.

- Data processing, according to the previous step, starts to segregate all information not relevant to TFG development, and starts to classify the data information for its respective processing.

- In the processing of Data, the classification and creation of new methodologies is initiated, which help to structure and carry out in a more organized way the complexity and development of the project.

- The virtual presentation and final delivery, depends on its effectiveness, when all the steps described above can be done in a correct way.

- Court, is the data mining, which determines which processes should be followed to a predictive method for the best and/or updating of the project

. V PRODUCTS OBTAINED

- ✓ In R code, which are basically based, on the statistical analysis approach. This language is widely used in scientific research, and very popular in the fields of machine learning.

- ✓ Thanks to machine learning and data mining, different studies related to the creation of complementary techniques are created.

- ✓ Logs or logs in each data mining process help you see clearer results of algorithm execution.

- ✓ the graphical interface, we seek to make the explanation of the results more understandable.

- ✓ Files provide an important tool because it helps to discard the analysis of errors resulting in the investigation.

- ✓ The memory of the project, turns out to be the base compilation of everything done, for decision-making, which finally results for the project being improvements for its virtual presentation.

Fig. 5 Extracting data

With the study of the data you can carry out different activities and acts to be able to carry out different tasks and be able to apply Data mining

VI SAW THINGS THAT CAN BE DONE WITH DATAMINING

INFORMATION - DATA + STATISTICS

It is based on the rule of well-collected data and well-applied statistic.

Data mining in its global concept is nothing more than the analysis and study of large data warehouses, which, in its primary function, is creating business ideas based on decision-making, with a predictive method that helps organizations solve complex problems as well as the organization and structure of their data; thus giving statistical results in an exact way that seek to make the best decisions for the operability of any project.

Generally data mining is a very complex, organizational process so accurate that it helps analyze data from different points of view in order to debug data into useful segments, and of great value, and importantly so that it ultimately helps the increase and profits of the company, and even giving it the opportunity to reduce costs , and has tools that allows users from different dimensions to analyze the data and summarize the identified co-relationships.

These techniques often seek to give light clearly the relationships of the data that, in turn between events, become hidden information. A simple example, but at the same time would be to find out the relationship between buying exercise products and consuming energizers, on specific Friday afternoon days in gyms. With this example, it is clearly sought to illustrate, the need of the field that is required for the application of Data Mining, in this case it should be clarified that only a specialist who has comfortable knowledge for the analysis of his clients, is also able to interpret the raw correlation of the data, which allows to see clearly the portrait of the people who go to the gym and even interpret what they consume most in a given time.

Another function that Data Mining fulfills is to work with large quantities, billions of views, with massive processing information without any inconvenience. On the other hand, it allows us to work with a large number of predictive processes that serves us as a great utility to select the variables, this means selecting the most useful ones within the data volume.

The Data Mining user usually determines the following 4 types of relationships:

1. Classes: The most complex views are randomly assigned to default groups. The process of classifying the data is nothing more than, assigning a data group to previously set sets, so that the probability of misclassifying data is minimized.

For example, a typical data classification problem is to divide a bank transaction database into a given module into possible solid groups and in turn classify them into values such as good or bad.

2. Clusters: According to the default criteria, they begin to be built in similar observations for the final deduction of each selected group.

The process or function of clustering consists in the grouping, organization of the datasets, in that order of ideas, is based on subdividing the dataset into exclusionary groups, so that each person in the group, is much closer to the other object, and the groups different from the latter, as far away as possible for its restructuring and organization , to all available variables and objects.

A very common and simple example for clustering application is the classification of your inventory in the logistics sector. For example, an electronic equipment repair corporation wants to be included in the drone market, but first conducts a marketing study to find out if there are groups of customers who concurrently use drones. In this example, for its breadth, there would be no default customer groups.

3. Associations: the views and observations generated in the project, fulfill a bald objective, therefore, are used for the identification of associations between objects. The finding of associations is different from relationships casually. Causal relationships are much more complex to find than associations because they are not of equivalent type, therefore they do not have to see causality between these processes.

4. Sequential Patterns: The identification of trend patterns and behaviors is carried out. An interesting and clear example serious, intensively the expression of different arrays that help the attack of different terminal diseases.

Examples of Data Mining application

Some areas of application of Data Mining are:

- Help with complex decision-making issues. Examples: Banking Offices, Finance Entities, Insurance Offices, Companies in Digital Marketing Processes,
- Industrial processes such as the reuse of raw materials and generation of conversion in different kinds of products.
- Scientific Research Examples: the discovery of new cures against diseases, the rapid absorption of dead cells, which help fight cancer back.
- Support for the design of databases of any company, both in infrastructure and methodologies.
- Improvement in the Consulting business area, helps decision-making and results improvement.

VII DATA ANALYSIS TOOL

Based on the above information, one of the tools used for the construction, modeling and analysis of data that has implemented these characteristics since its inception and today increased its level of productivity.

VIII PHASE DEVELOPMENT OF GOOGLE ACADEMIC ANALYTICS

Phase I: Data Collection (Collection)

Data feed is the source of revenue from any mobile app or from any connected internet point we require. All information received goes through a process which is initially sent to google Analytics in the form of packages known as Hits or interactions through the so-called tracking snippet.

The function of the tracking code is to send to google Analytics all information covered on page and screen that the user is viewing; another notable aspect is the collection of information about the browser such as version, language, operating system, etc. It should be noted that one of the pillars of Google is confidentiality, that is why it has some privacy considerations such as:

Google reports don't store or report personal information. Access to your data, you can not have google, if you do not have your prior notice or permission. It has multiple ways to share information with google and providers**.**

Phase II: Information Processing (processing)

This phase involves the transformation of information into data, so that they can be categorized and are highly relevant.

To perform this procedure, first of all, it is necessary to re- structur the data, they must be organized into two groups, sessions and users, then the snippet brings together the information collected by google adwords, adsense, search console and any other source external to Google. There are 2 methods to import data into Google Analytics: Custom data import: Implementing the Google Analytics API will allow us to import new features and dimensions into all reports that are generated.

Cost Data Import: Your goal is to report to Google Analytics money or amounts that are not spent on ads other than adwords. To comply with this method we must upload a file that includes the source and the media of the campaign.

Phase III: Configuration: (Configure)

At this stage it is necessary, go hand in hand with the reports, when it is up to them the customization

that we have included in the past and that works now, for example, we may have created a filter to exclude traffic from URLs. Finally, the information goes through a process called aggregation, in which the main function is to prepare all the information, so that it can then be analyzed; this procedure manipulates database tables, for their respective organization and storage, so that when necessary we can access quickly and effectively, and best when required. After the previous step we store again in the database and in this way it can no longer be modified again, for this reason we will never be able to alter the data of the Analytics history.

Phase IV: Reporting

Finally, at this stage, we access all the data in the form of reports, this process is considered a form of readable and organized combinations for analysis of metrics and dimensions.

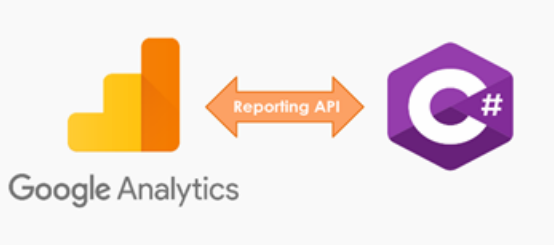

Fig. 6 Reporting of google analytics and code.

When requesting more complex data, google by default will use a process called sampling or also known as sampling, which will eventually allow you to return the quick information. We conclude with a journey that will open concepts of metrics and dimensions, that will allow us to differentiate, and classify the quantitative of the qualitative.

Fig. 7 Google Analytics process phases

CONCLUSIONS

In conclusion, it can be said that:

- Learning Analytics has attracted a lot of attention in recent years as educational institutions and researchers increasingly a potential learning analytics has to support students' learning processes as approaches to Learning Analytics share various areas of various methodologies and create movement from data to analysis to action to form clear learning and help students in the future. Learning Analytics is a modeled learning today it can be evidenced that after bringing together different learnings and tactics Learning Analytics managed to be something more compacted and directed to a specific audience helping to strengthen gaps in different aspects of the educational environment.

- Today we can see Analytics Academy as part of our lives, and it becomes an indispensable tool for the preparation, reporting and monitoring of both web applications as well as mobile applications. Moreover, many of the services offered by google analytics are completely free, making it the most used application by companies to do statistical analysis for quantitative topics and analysis to users for qualitative topics.

- A Data Mining system allows us to make predictions about complex variables, which make it possible to have a statistical system that allows us to perform different alarms, that allow us to identify the future or what will happen, important to note that Data Mining is considered as the last stage method of the introduction of quantitative data.

REFERENCES

[1] AlShammari, I. A., Aldhafiri, M. D., & Al-Shammari, Z. (2013). A meta-analysis of educational data mining on improvements in
learning outcomes. College Student Journal, 47(2), 326-333.

[2] Baker, R., & Ocumpaugh, J. (2015). Interaction-based affect detection in educational software. In R. Calvo, S.

[3] D'Mello, J. Gratch, & A. Kappas (Eds.), The Oxford handbook of
affective computing (pp. 233–245). New York: Oxford University
Press.

[4] Baker, R., Corbett, A., Koedinger, K., y Wagner, A. (2004). el comportamiento fuera de la tarea en el aula tutor cognitiva. Actas de la Conferencia sobre 2004 Factores Humanos sobre sistemas de
computación - CHI '04.

[5] likstein, P. (2013). Multimodal learning analytics. Proceedings of the 3rd International Conference on Learning

[6] Analytics and Knowledge (LAK '13), 8–12 April 2013, Leuven, Belgium. New York: ACM

[7] Dietz-Uhler, B. y Hurn, J. E. (2013). Using learning analytics to predict (and improve) student success: A faculty perspective. Journal of Interactive Online Learning, 12(1), 17-26

[8]Siemens, G. y Baker, RSD (2012, abril). Análisis de aprendizaje y
minería de datos educativos: hacia la comunicación y la colaboración.En Actas de la 2da conferencia internacional sobre análisis de aprendizaje y conocimiento (pp. 252-254).

[9] Romero, C. y Ventura, S. (2007). Minería de datos educativos: una encuesta de 1995 a 2005. Sistemas expertos con aplicaciones, 33 (1),
135-146.

[10] (2019). 2019 14th Iberian Conference on Information Systems and Technologies (CISTI), Information Systems and Technologies (CISTI),
2019 14th Iberian Conference On, 1–5

[11] Triana, D. y Portela, J.(2020).Cava novel solution.
Análiticas de Aprendizaje.
Recuperado el 28 de marzo de 2020. Disponible en:
https://es.slideshare.net/secret/ntgI39BQsWo2zb

[13] Seufert, S., Meier, C., Soellner, M., & Rietsche, R. (n.d.). A
Pedagogical Perspective on Big Data and Learning
Analytics: A Conceptual Model for Digital Learning Support. Technology, Knowledge and Learning, 24(4), 599– 619.

[14] Recuperado el 28 de marzo de 2020. Disponible en: https://link-springer-com.bibliotecavirtual.unad.edu.co/article/10.1007/s10758-019-09399-5

[15] Peña-Ayala, A. (2017). Learning Analytics: Fundaments, Applications, and Trends : A View of the Current State of the Art to Enhance E-Learning. Cham, Switzerland: Springer. Recuperado el 28 de marzo de 2020. Disponible en: http://search.ebscohost.com.bibliotecavirtual.unad.edu.co/login.aspx?=true&db=edsebk&AN=1470947&lang=es&site=eds-live&scope=site

[16] KhanAcademy. (2016). KhanAcademy. Recuperado el 28 de Marzo de 2020. Disponible en: https://es.khanacademy.org/about

[17] Klass data. (2014). SmartKlass. Recuperado el 28 de Marzo de 2020. Disponible en: http://klassdata.com/smartklass-learning-analytics-plugin/

[18] imagen recuperada de https://www.econesta.com/wp-content/uploads/2018/09/pred_mod5.Rplot_.jpeg

[19] imagen recuperada de https://www.google.com/url?sa=i&url=https%3A%2F%2Fes.wikipedia.org%2Fwiki%2FAn%25C3%25A1lisis_de_grupos&psig=AOvVaw2YlwqlR8rw8ZaTsxuFz88q&ust=1587260663079000&source=i

mages&cd=vfe&ved=0CAIQjRxqFwoTCLDyo57t
8OgCFQAAAAAdAAAAABAZ