Learning Analytics as a Tool for Visual Analysis in an Open Data Environment: A Higher Education Case

Johnny Salazar-Cardona[1], David Angarita-Garcia[1](✉), and Jeferson Arango-López[2,3]

[1] Institucion Universitaria EAM, Av. Bolívar # 3 – 11, Armenia, Colombia
adavidalberto@eam.edu.co
[2] FIET, Universidad del Cauca, Calle 5 N.º 4-70, Popayán, Colombia
[3] Facultad de Ingeniería, Universidad de Caldas, Calle 65 Nº 26-10, Manizales, Colombia

Abstract. In the past few years, the data analysis in the academic field has gained interest, because higher education institutions generate large volumes of information through historic data of students, information systems and the tools used by the learning processes. The analysis of this information supports the decision making, which positively impacts the academic performance of students and teachers. In addition, if the results of the analysis are shared with the institution community, it is possible to individualize the needs of the students and professors. Thus, the professional improving process can be applied to each person.

This analysis needs to be focused on an education context such as Learning Analytics (LA) and Education Data Mining (EDM), which address the Knowledge Discovery process (KDP). Because of this, we present in this paper the implementation of LA in an Open Data (OD) environment to analyze the information of a higher education institution. The analysis is focused on the academic performance of the students from different perspectives (social, economic, family, among others). In order to improve the results, the analysis process is done in real time through Web Analytics (WA) for each member of the institution according to its needs.

Keywords: Learning Analytics · Open Data · Educational Data Mining
Web Analytics · Academic Institution

1 Introduction

The analysis of data in the academic field arose due to the adjusted in the ideology of use of Knowledge Discovery Process (KDP) and the business intelligence (BI). These processes are applied in educational institutions in order to find elements that support decision making by teachers and managers. The main objective of these processes is to continuously improve the processes of the institutions as an organization. Apart of that, it is important to identify the factors that impact students' academic performance in a positive or negative manner. This through the collection of data from their histories or

J. E. Serrano C. and J. C. Martínez-Santos (Eds.): CCC 2018, CCIS 885, pp. 55–69, 2018.
https://doi.org/10.1007/978-3-319-98998-3_5

academic trajectories and actions carried out in any system that supports the process of student training [1].

In this field of academic data analysis, we can find two scientific communities that give the guidelines to obtain the best results, LA (Learning Analytics) and EDM (Educational Data Mining). Although, some authors give a wrong approach to these concepts confusing each other. However, the approach of both communities is similar, but there are key elements that differentiate its application. The LA in its application takes as a main axis the human being for the analysis of the data, using them as a whole to make a description of their general behavior with a visual approach. For its part, EDM analyzes particular and individual data elements in an automated way, excluding the human being from the equation, without being necessary a high graphic level for the understanding of results [2].

Additionally, the LA takes as its main base the Business Analytics (BA) and the Web Analytics for its respective application in the educational field. Based on these elements, it makes use of the basic tasks of a KDP to execute them on this context, focusing on aspects of data analysis that are more effective in educational information. This provides a high level of interaction with the final user, offering intuitive graphics, search filters with dynamic reports that allow the human being to make a judgment in front of the visualized data [1].

In this study, a prior analysis was made of the approach that the technology used for the design, implementation and publication of the Dashboards would have, obtaining similar results of existing researches [3, 4]. Subsequently, with the data centralized and treated computationally, a descriptive process was applied with visual analytics, focusing efforts in the elaboration of predefined dynamic reports or Dashboards, with which the users (The academic community in general) can interact and perform their own analyzes in an Open Data environment. The data that was used in the elaboration of this prototype belongs to the students of the Software Engineering program of the "EAM University Institution", which comply with the protection regulations of the HABEAS DATA [5], and focused on analyzing different aspects as the academic performance of students throughout their career, subjects, social context, employment situation and family context [6].

The objective of this article is to present the Learning Analytics implementation process in a higher education institution with an Open Data approach, and to achieve this, the document is structured as it follows: - Background in Sect. 2 where a conceptualization of the important elements is presented and general terms; - Materials and Methods in Sect. 3, which details the materials used in the research process; - Results in Sect. 4 and finally the conclusion in Sect. 5.

2 Background

In the last few years, LA research has been conducted in different aspects of the academic sector, such as data analysis to improve teaching practice [7, 8], student motivation and collaborative work analysis in virtual environments [9], understand and improve the learning students process [10–12] and the generation of performance indicators to improve different institutional processes [11, 13], which is the objective of this investigation. Additionally, Latin America is making some implementation efforts in this area of research, with visual analytics as one of its main focuses [14], but neither applied to an environment of open data for public decision making for the academic community as in this research.

For the Learning Analytics application, it is necessary to be clear about some concepts on which this kind of methodologies is based. Now, we are going to present different key concepts that are necessary for the execution of the project.

2.1 Knowledge Discovery Process (KDP)

KDP is an ideology for the implementation of specific tasks and activities that must be executed for any knowledge discovery process. From this, many methodologies have emerged according to the field to apply. Some of these methodologies are: (1) KDD (knowledge Discovery in databases), Process to discover useful knowledge of a set of structured or semi-structured data; (2) SEMMA (Sample - Explore - Modify - Assess); (3) Catalyst or P3QT; (4) CRISP-DM (Cross Industry Standard Process for Data Mining); (5) EDM (Educational Data Mining); (6) LA (Learning Analytics) among others. Of all the existing models, the first one defined was KDD in 1996 as a model for the field of research [15–17], and subsequently various approaches were defined, such as the models mentioned above. Among the most prominent models is CRISP-DM, which specializes in the industrial field, KDD that offers flexibility in the execution process and EDM-LA for the academic context, this last one was used for the execution of this research.

2.2 Open Data (OD)

Open Data is an ideology and a practice that seeks the release and public access to the data that is digitized for its free use, reuse and distribution by any interested person, provided they are attributed and shared in the same way in which it appears [18, 19]. Open data usually come from 2 main sources: the scientific field (such as data on the human genome, geographic information, climate, etc.) and the government (Accountability, transparency, crime, catastrophes, education, etc.). Regarding the governmental field, the data should be published by public and private institutions with public functions [20], in order to provide processes of transparency, participation, innovation and empowerment on the part of the citizenry or any interested third party [21]. On the other hand, in the academic sector, are the educational institutions that have the duty to apply said processes of data opening [22]. Currently Open Data is very important and is accepted as a fundamental component for making decision and the promotion of research. But compared to other areas for which data is available, the field of education has limited resources in these repositories [23]. In the case of Colombia, the few efforts that have been made do not meet the minimum quality criteria [24, 25].

2.3 Visual Analytics (VA)

Visual analytics is the point of convergence between the visualization of information and scientific visualization, centered on the analytical reasoning of the human being through visual and interactive interfaces, applying a descriptive process on the data [26]. This concept is widely used in the field of business intelligence and different areas of knowledge [27], allowing in an easy and intuitive way to understand the behavior of the data that is being visualized [28]. In addition, it offers a component of interaction with the end user to support decision making. To achieve a successful human judgment on the data used, it tries to increase the cognitive resource by using different visual elements, decreasing search times by presenting them in a compact space. This shows the data through different dimensions to understand its behavior and the identification of patterns with predefined dynamic filters that allow to evaluate the data in different contexts, supported by colors and sizes to induce knowledge [29]. Additionally, since the emergence of web 2.0, visual analytics has turned to its implementation on the web [30], in order to have Dashboards on any platform, integrating different kinds of graphics such as decision trees, bar charts, maps, and any kind of graph that allows structuring of the data.

2.4 Learning Analytics (LA)

Learning Analytics is a research community seen as an additional branch in the field of knowledge discovery processes, applied specifically to data from an educational environment. Where the concept of visual analytics is used to describe the behavior of the data generated in an academic context. These data involve elements such as levels of complexity of the courses, identification of aspects that affect academic performance, positive or negative reception of students in different topics, academic traceability, student desertion, access control to virtual education platforms, among others.

During its history, the term Learning Analytics has been debated with respect to the concept and community of Educational Data Mining [31–33], with differences such as the focus of research and the size of the data. This gap between the research data approaches and the size of the datasets is currently non-existent, due to the fact that in the different scientific disclosures of these 2 communities, the research topics are similar, although it is the applied approach to the analysis of the data which makes a difference between these. LA makes use of the concept of Business Intelligence, specifically visual and web analytics, where the main element of judgment is the human being who interprets and focuses the analyzes according to their needs in an intuitive way, performing descriptive analysis. For its part, EDM focuses on the application of semiautomatic predictive elements, where the human being loses this relevance and is not the one who issues first-hand judgments for the interpretation of results [2].

In Latin America, the implementation of Learning Analytics is still limited, where the research published in this field of knowledge is headed by Brazil, which for the year 2017 had a total of 14 articles and the participation of 16 institutions, followed by Ecuador with 7 publications, Chile with 3 and Colombia is in the 4th Latin American position with a total of 2 articles from 2 institutions [14]. These quantities show a low level if it's consider the relevance of knowledge discovery processes at the present. The few publications of the sector and of the global research trend, the predominant theme is monitoring and analysis, applying experimental processes and dominating the application of statistical processes, machine learning, social network analysis and visual analytics [34].

Finally, LA exposes an application process and a methodology of approach, which are based on the KDP ideology (see Figs. 1 and 2) and proposes 3 stages: (1) Data collection and pre-processing, in which the data from different sources in a single data warehouse, to later apply a set of business rules that will allow to have adequate data to analyze. This stage is defined from the union of different tasks of KDP, such as the analysis and understanding of the data environment, creation of the working database, cleaning and transformation of the data; (2) Analytics and action, is the application of data analysis techniques, for its interpretation, pattern detection and from the objective of the analysis, take the respective measures together with the disclosure of results. This stage takes the actions of understanding and choosing the technique of data mining, its application, interpretation - evolution and processing of results of the KDP ideology; (3) Post-processing, focused on the refinement of the work database in order to extend the analyzes carried out, adding new sources of data, attributes and extending the previously performed analyzes to new approaches. Regarding the application model, there are 4 fundamental elements: (1) what? It refers to the set of data that will be analyzed; (2) who? Referring to whom the analysis will be destined; (3) Why? Related to the objective of the analysis; (4) how? Related to the process that is applied to achieve it, and at this point is that the application process previously treated is immersed.

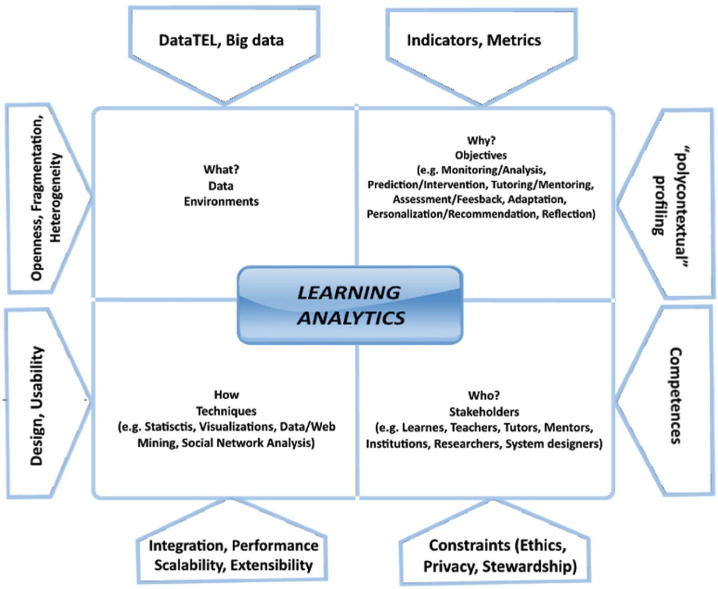

Fig. 1. Learning Analytics model [35].

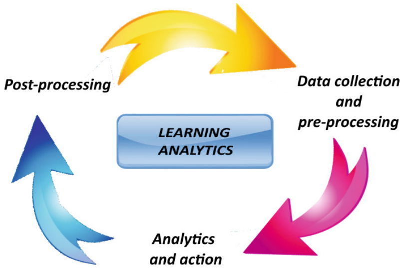

Fig. 2. Learning Analytics process [35].

3 Materials and Methods

In the following paragraphs, we will explain the different stages that were carried out in the methodology used, guiding the execution of the investigation until reaching the results stage.

3.1 Learning Analytics ¿What?

The data set that was analyzed contains the basic information of all students of the software engineering program belonging to the EAM university institution, located in Armenia, Quindío. This dataset is made up of the historical data from 2007 to 2017, including family information, origin, residence, disability and employment status of students. In addition, it was integrated with another set of different data which contains all the academic record. In total the student records are 634 with 50 attributes, and the data set of the academic record had 13282 records, with the different registers each subject taken by each student with 13 attributes, which highlights the scores of the first 3 cuts and the definitive academic space.

3.2 Learning Analytics ¿Who and Why?

The objective of the project is to provide a mechanism by which in an OD environment real-time analysis can be performed, where any person or area of the educational institution can perform their own analysis. This was focused on senior management of the institution to assess the current status of the program; to the marketing area so that they can analyze which geo-referenced locations have the largest number of students, and have knowledge about the socioeconomic strata that should be considered to increase the income of a larger population; the area of the CAP (Comprehensive Accompaniment Plan[1]), to identify patterns and behaviors of students with low performance, students with disabilities and student desertion; The "Parents' School[2]" area to focus studies and academic training programs aimed at parents of students for their

[1] Responsible area for the accompaniment and support of students with poor academic performance.

[2] Academic program focused on the integration of parents in the academic education of their children as well as the continuity of their own professional training.

continued inclusion in the academic field; to the students in general so that they can carry out descriptive processes of the program, visualizing subjects with greater positive and negative impact in their professional training process. Mainly, the previously proposed approaches seek to find patterns that affect the academic performance of the students according to the situation of the parents, family nucleus, place of residence, semester that is studying, working day, academic spaces, socioeconomic level, disability and employment situation [6].

3.3 Learning Analytics ¿How? – Data Collection and Pre-processing

The process of data centralization and pre-processing was applied from an ETL process [36], in which the data from the original data source were extracted from the consumption of a web Service, and subsequently a set of business rules were applied in order to calculate and organize data that did not meet a minimum quality, in order to answer the established questions. Some of the business rules applied to have an optimal data set were: Calculation of latitude and longitude for geo-referenced visualization of students from residence addresses, treatment of badly named subjects and notes out of range, calculation of time that it takes the students to complete their studies, last semester completed - last year attended by each student for the detection of student desertion, elimination of attributes in which a high percentage had null values and that were not necessary to answer the questions posed, and corrections of some attributes with non-standardized values such as gender, professions of parents, cities of residence (integrating neighboring towns, from where students move daily), levels of education of parents and monthly economic income [37]. It is noteworthy that the analyzed data comply with the HABEAS DATA protection standards [5]. Finally, when the data was found clean they were centralized on the Windows Azure platform [38], which provides a quick access mechanism on the web and integration with multiple data analysis platforms. In this centralization, a data warehouse with multiple dimensions was designed, which will allow an optimal performance in the long term when the amount of data is even greater [39] (see Fig. 3).

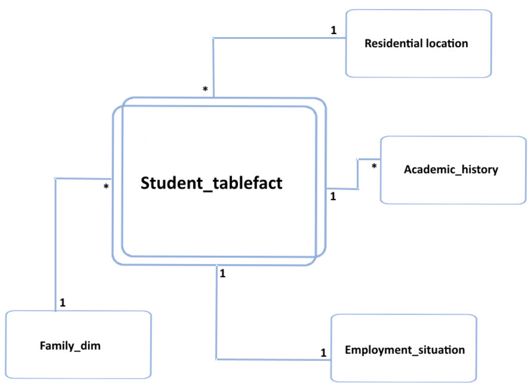

Fig. 3. Data warehouse model

3.4 Learning Analytics ¿How? – Analytics and Action

In the initial stages of the project execution, there was carried out a systematic review process to determine in a higher education institution, what would be the most adaptable and efficient kind of tool that could be applied in an academic setting, resulting in the Power Tool Microsoft BI [4]. With this tool a configuration was established for the consumption and visualization of data in real time, in order to perform analyzes with updated data for the entire university community.

Subsequently, we proceeded to design and implement each of the Dashboards that would allow the university community to perform their own analysis, and these reports provide the possibility of downloading the data at any time, in order to apply external analysis processes, as predictive analysis processes applying the EDM model with any of its different fields of action [40]. With the different Dashboard designed, it was sought to focus the analysis to answer the questions and objectives. The first one was the analysis of the academic performance of the students by the day, level and semester. For this Dashboard variables were used as: Definitive note of the 3 cuts per semester along with its definitive, and its visualization through the years of existence of the software engineering program, as well as through the different academic semesters. It is also allowed to see the behavior of each of the academic spaces and its final note, the total number of students who have seen each course, and the application of filters per academic day (Day - Night), city of residence, level (technical, technological or university), academic term (1–2), distribution by gender and a key performance indicator with the general academic average. It should be noted that all Dashboards implemented are completely dynamic, so when selecting any element in said report, all the graphics and data of the other sections in it will be affected.

The Dashboard that allows analysis by social stratum or physical disability has variables such as final academic space, total students per year, distribution of students by socioeconomic stratum, distribution of students by kind of disability, indicator with the total number of students, map with geo-referenced visualization of all students, and filters such as municipality, kind of disability and kind of document. Regarding the analysis provided focused on the family nucleus of students has variables such as final academic space, average of each cut and final per year, distribution of the total siblings of students, distribution of total students by socioeconomic stratum, total students, general academic average, geo-referenced location of the students, and specific filters by HPE (Health Provider Entity), municipality, gender, number of dependents, own home, type of document, year and marital status (see Fig. 4).

Regarding the analysis provided focused on the family nucleus of students, has variables such as final academic space, average of each cut and final note per year, distribution of the total siblings of students, distribution of total students by socioeconomic stratum, total students, general academic average, geo-referenced location of the students, and specific filters by HPE (Health Provider Entity), municipality, gender, number of dependents, own housing, type of document, year and marital status. An example of the generated Dashboards can be seen in Fig. 4.

Regarding the analysis provided focused on the employment situation, the variables were available as definitive for academic space, distribution by socioeconomic stratum, distribution by dependence economy (dependent, independent, family), average per

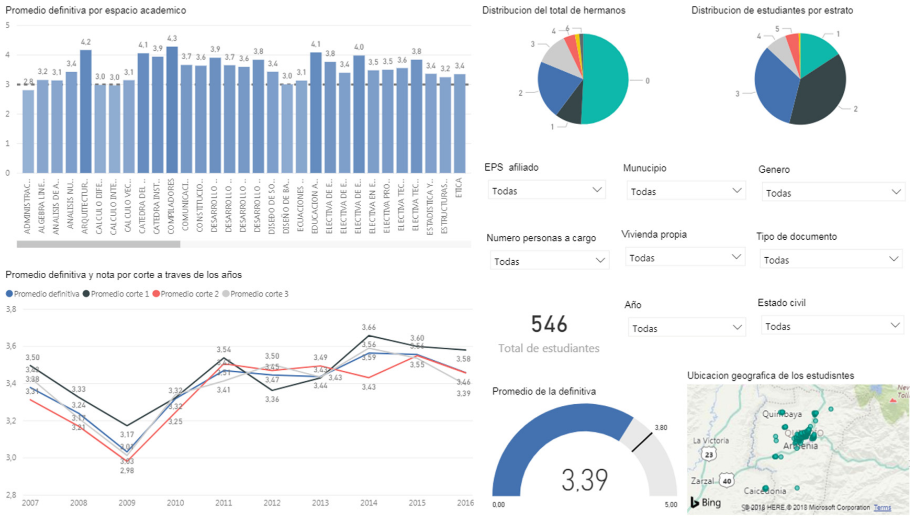

Fig. 4. Dashboard of the family nucleus of the students

semester, total of dependent and independent students per year, total of students in general, and filters such as civil status, gender, city of residence, position in the company as a dependent and independent.

The Dashboards in which it is not sought to analyze the academic performance as the main axis, but seeks to assess the context in which the student develops in their day to day as their family situation, can cross variables such as the distribution of fathers and mothers of the students by profession and academic level, academic performance of the students by said formation of the parents and the application of filters by years and gender visualizing the total of students.

Finally, for the evaluation of student desertion, the analyzed variables were as average per academic space, total students per year and per semester, general academic average, student distribution by gender, last year attended by students, total students by stratum, total number of students enrolled, total number of students in course, and the provision of filters by program, year, last semester taken by a student and duration in culminate the race (here can be filter by seeing only deserters).

3.5 Learning Analytics ¿How? – Data Collection and Post-processing

All data and dashboards implemented were made available on the web portal of the EAM university institution, so that the entire academic community can access them and can perform their own analysis, as well as everything is in the Windows Azure ecosystem, all the data that is displayed is updated periodically, offering fully updated and treated data.

4 Analysis of Results

In the next paragraphs, we will explain the different results obtained, dividing them by academic performance, family nucleus and student desertion.

4.1 Students' Academic Performance

By conducting an exploratory process in the different implemented dashboards, some relevant behaviors were identified in the students of the software engineering program, which allow to focus specific points that will improve the students' academic performance: (1) In general terms, women have a better academic performance than men in the course of their training as software engineers, although it is a preferred career for men corresponding to 83.44% of the total student body; (2) It is possible to visualize the constant improvement of students through their professional training, but this behavior changes in 5th semester, this may be due to the fact that it is the semester where the change of training from technical level to technologist is generated, and subsequently to university (see Fig. 5); (3) It is possible to demonstrate that the academic spaces that have an average in the final mark of the students below 3 are: Administration of databases, Physics I, Geometry, Programming Logic, Mathematics, Discrete Mathematics, Applied Mathematics I, Applied Mathematics II, Pre-calculus and Principles of software engineering, being these academic spaces with high demand for logical skills; (4) It is evident that around 72% of the students are from strata 2 and 3, in addition there is a clear relationship between the socioeconomic stratum and academic performance (see Fig. 6), where the highest socioeconomic stratum is the lowest; (5) Half of the students in the program are only children; (6) Only 10% of students have a job in a dependent or independent way, the other 90% depends on their family; (7) Students who work as freelancers have better communication skills, and this is clearly reflected in the grades of the academic spaces where communication skills are necessary, far exceeding the general average; (8) Students with a physical or motor disability or who are married at the time of beginning their professional training process have an academic performance higher than the general academic average. (9) A representative factor in the analyzed sample is the impact that the socioeconomic strata have on the academic performance of the students and the possible student desertion. For this reason, awareness-raising strategies should be generated to encourage and value the process of professional training in students in order to level these differences between strata. One element that may be of interest for increasing the income of students to the engineering field of the Software area is the increase of effective dissemination in the highest strata, because these are not being well received, this being reflected in the low number of students to a greater socioeconomic stratum.

4.2 Family Nucleus of the Students

In the exploratory process carried out, it was identified that: (1) Although the students have fathers and mothers with postgraduate training, they tend to have a lower academic performance when their father has this kind of training, but when the mother has the performance, the student's academic level is above average (see Fig. 7); (2) When

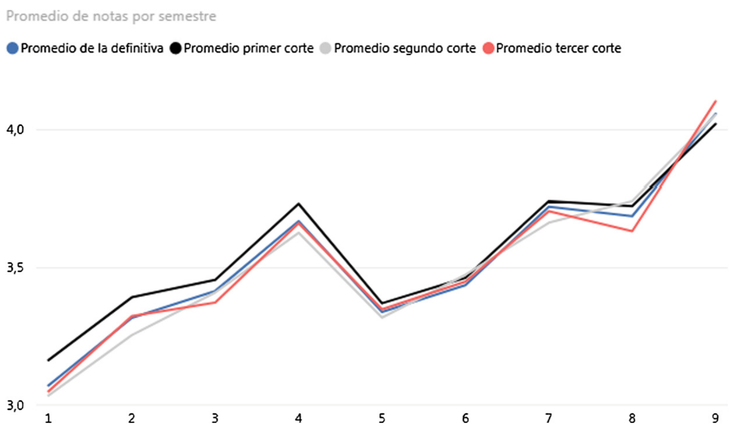

Fig. 5. Academic average by semester

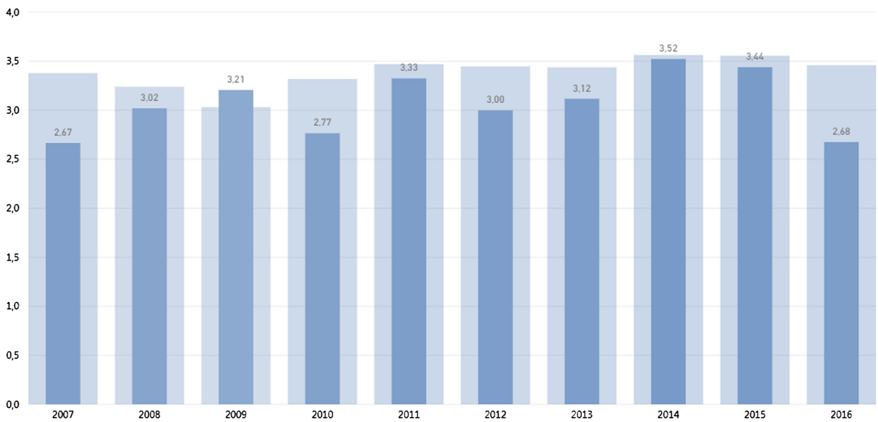

Fig. 6. Academic average by social stratum 5–6 in comparison with the general average

the father of a student is abroad or works as an independent, the student's academic performance is low, while if the father works as a teacher the academic performance tends to be higher; (3) In a high percentage the parents of the students of the software engineering program are retired, and of these none have postgraduate training, while in the retired mothers although there are not many, they have postgraduate training.

With the above, element that influences the academic performance of students are their parents, which was verified with the results obtained, and, therefore, should create awareness strategies for parents to positively influence their children in the process deformation. In addition, the diversity of the professional practice of the parents was also identified, which can be seen as an opportunity for inclusion through programs such as "School of parents" for the integration of these with their professional training and that of their children.

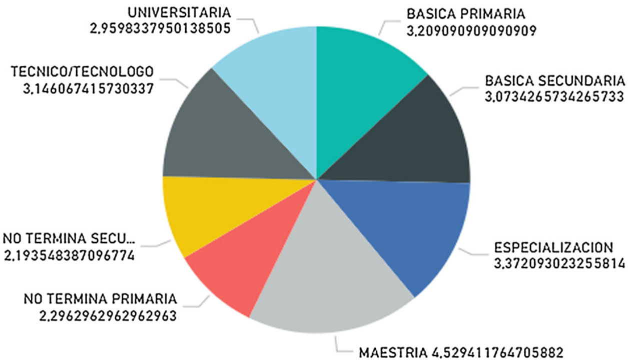

Fig. 7. Academic average bi academic level of the mother's student

4.3 Student Desertion

The most relevant results obtained from the exploratory process carried out in the student dropout Dashboard are: (1) although in the program the great majority are men, women are the ones who desert the most. If taken individually, 67% of women are dropouts, compared to 59% of men; (2) In all academic semesters there are dropouts, but the lower the semester the greater the amount of desertion; (3) The distribution of desertion for stratum 5 and 6 is 75%, for stratum 4 it is 54%, 60% for stratum 3, 61% for stratum 2 and 55% for stratum 1; (4) Women, although they drop out more than men, have an academic performance above 3 when they make this decision, while men have an average below 3 when they drop out, (5) Academic spaces with worse academic performance of dropouts they are those related to mathematics and programming (see Fig. 8).

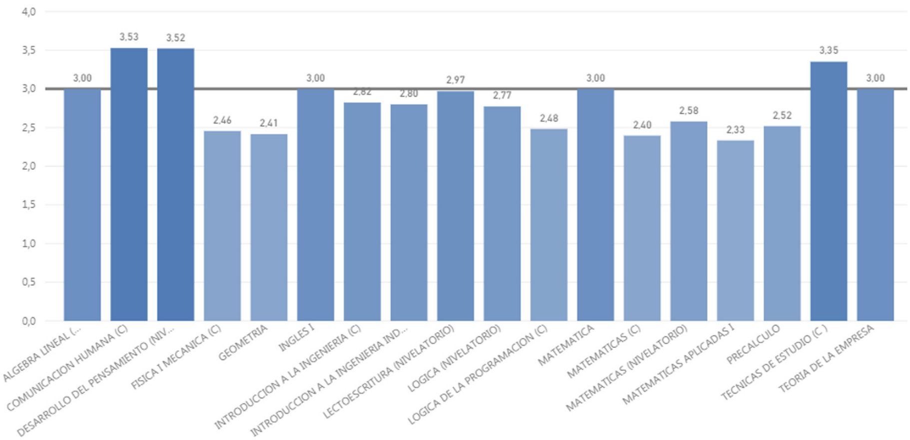

Fig. 8. Academic average by academic spaces and student desertion on the first semester.

5 Conclusions and Future Work

This document presents the different elements that were taken into account for the implementation of a visual analytics process in an Open Data environment of a higher education institution, based on the Learning Analytics approach. These descriptive analyzes focused mainly on the academic performance of the students, student desertion and their family context from structured data, pre-processed, transformed and centralized in a Data warehouse under the Windows Azure ecosystem and analyzed with the different available elements of the Power BI tool.

Statistical patterns were identified, which can be used by different areas of the university institution analyzed, such as the Software Engineering academic program, the CAP, the Marketing area or any individual of the university community. Among the elements identified, the impact that students have on their passage from a technical level to technological, and later the University. For this, strategies must be generated that allow the change between the different academic levels to be carried out normally. Additionally, it was evident that the most difficult academic spaces for students in general are related to the use of logic to solve problems, so efforts should be focused to further develop this skill in students, in order to increase these indicators and thus avoid student desertion, which was also compared with the low academic performance of this kind of subjects.

This research demonstrates the possibility of integrating visual analytics in a higher education institution, where all the analysis mechanisms are made available to the academic community in general, releasing all non-sensitive data so that any interested can rely on them for decision making, applying filters and visualizing the behavior of the data in real time without violating any kind of privacy and always looking for the continuous improvement of the academy, both at the level of the training processes of the students and the institution seen as an organization.

In order to continue refining the implementation of data analysis in the academic sector, data should be integrated not only from a particular program, but from the entire university community, in addition EDM elements should be added as predictive processes, in order to find even deeper patterns that allow generating new strategies for the strengthening of university institutions.

Acknowledgements. We thank to the research area of the EAM University Institution and the Software Engineering program, which facilitated the whole process of project execution.

References

1. Baepler, P., Murdoch, C.J.: Academic analytics and data mining in higher education. Int. J. Sch. Teach. Learn. **4**, 11 (2010)
2. Siemens, G., Baker, R.: Learning analytics and educational data mining: towards communication and collaboration. In: Proceedings of the 2nd International Conference on Learning Analytics and Knowledge, Vancouver, British Columbia, Canada, pp. 252–254. ACM (2012)

3. Gounder, M.S., Iyer, V.V., Al Mazyad, A.: A Survey on business intelligence tools for university dashboard development. IEEE (2016)
4. Salazar Cardona, J.A., Angarita Garcia, D.A.: Evaluación y selección de herramientas de analítica visual para su implementación en una institución de educación superior. IngEAM, vol. 4, pp. 1–20 (2017)
5. Cote Peña, L.F.: Hábeas Data en Colombia, un trasplante normativo para la protección de la dignidad y su correlación con la NTC/ISO/IEC 27001. Magíster en Derecho, Facultad de Derecho, División de Ciencias Jurídicas y Políticas, Universidad Santo Tomás Seccional Bucaramanga, Bucaramanga, Colombia (2016)
6. Vásquez Velásquez, J., Castaño Vélez, E., Gallón Gómez, S., Gómez Portilla, K.: Determinantes de la desercion estudiantil en la Universidad de Antioquia. Facultad de ciencias econimicas. Centro de investigaciones economicas, Universidad de Antioquia, p. 43 (2003)
7. Tanaka, M.: C3.js D3-based reusable chart library, Mayo 2014. http://c3js.org/
8. Einhardt, L., Aires Tavares, T., Cechinel, C.: Moodle analytics dashboard: a learning analytics tool to visualize users interactions in moodle. In: XI Latin American Conference on Learning Objects and Technology (LACLO) (2016)
9. Echeverría, L., Benitez, A., Buendia, S.: Using a learning analytics manager for monitoring of the collaborative learning activities and students' motivation into the Moodle system. In: IEEE 11th Colombian Computing Conference (CCC) (2016)
10. Charleer, S., Vande Moere, A., Klerkx, J.: Learning analytics dashboards to support adviser-student dialogue. IEEE Trans. Learn. Technol. (2017)
11. Morais, C., Alves, P., Miranda, L.: Learning analytics and performance indicators in higher education. In: 12th Iberian Conference on Information Systems and Technologies (CISTI) (2017)
12. Ruiz Ferrández, M., Ortega, G., Roca Piera, J.: Learning analytics and evaluative mentoring to increase the students' performance in computer science. In: IEEE Global Engineering Education Conference (EDUCON) (2018)
13. Van der Stappen, E.: Workplace learning analytics in higher engineering education. In: IEEE Global Engineering Education Conference (EDUCON) (2018)
14. Lemos dos Santos, H., Cechinel, C., Carvalho Nunes, J.A.B., Ochoa, X.: An initial review of learning analytics in latin America. In: 2017 Twelfth Latin American Conference on Learning Technologies (LACLO), Argentina (2017)
15. Fayyad, U., Piatesky-Shapiro, G., Smyth, P.: From data mining to knowledge discovery in databases. IA Mag. **17**(3), 37–54 (1996)
16. Fayyad, U., Piatesky-Shapiro, G., Smyth, P.: The KDD process for extracting useful knoledge from volumes of data. Commun. ACM **39**, 27–34 (1996)
17. Fayyad, U., Piatesky-Shapiro, G., Smyth, P.: Advances in Knowledge Discovery and Data Mining. MIT Press, Cambridge (1996)
18. Open Knowledge Foundation: Open Data Handbook Documentation: Release 1.0.0 [online], vol. 1, Cambridge (UK) (2012)
19. Salazar, J.: Marco de referencia para la implementación del mapa de ruta establecido en los lineamientos nacionales de apertura de datos del sector público y su integración con procesos de descubrimiento de conocimiento e inteligencia de negocios. Master en Ingenieria Computacional, Facultad de Ingenierias, Universidad de Caldas, Manizales, Colombia (2015)
20. El Congreso de la República, "Ley 1712 del 2014. Literal J, artículo 6. Definiciones," Colombia, Ed., ed. Colombia: Diario Oficial 49084 de marzo 6 de 2014 (2014)

21. Florez Ramos, E.: Open data development of countries: global status and trends. In: IEEE ITU Kaleidoscope: Challenges for a Data-Driven Society (ITU K), 27–29 November 2017
22. Heise, A., Naumann, F.: Integrating open government data with stratosphere for more transparency. J. Web Seman. **14**(Jan), 3 (2012)
23. Akhtar Khan, N., Ahangar, H.: Emerging trends in open research data. In: IEEE 9th International Conference on Information and Knowledge Technology, pp. 141–146 (2017)
24. Mahecha Moyano, J.F., López Beltrán, N.E., Velandia Vega, J.A.: Assessing data quality in open data: a case study. In: 2017 Congreso Internacional IEEE De Innovacion y Tendencias en Ingenieria (CONIITI) (2017)
25. Salazar Cardona, J.A., Gomez, C.H., López Trujillo, M.: Knowledge discovery process in the open government Colombian model. In: 2014 9th Computing Colombian Conference (9CCC) (2014)
26. Wong, P.C., Thomas, J.: Visual analytics. IEEE Comput. Graph. Appl. **24**, 20–21 (2004)
27. Chitvan, M., Nayan, C., Ajayshanker, S.: Scope and challenges of visual analytics: a survey. In: International Conference on Computing, Communication and Automation (ICCCA2017) (2017)
28. Mitchell, J., Ryder, A.: Developing and using dashboard indicators in student affairs assessment. New Dir. Stud. Serv. **142**, 71–81 (2013)
29. Scholtz, J., Ebert, D., Elmqvist, N.: User-centered evaluation of visual analytics. Synthesis Lectures on Visualization, p. 71 (2017)
30. Yu, S., Wu, L., Zhang, X.: Research on equipment knowledge representation based on visual analytics. In: IEEE 13th International Conference on Semantics, Knowledge and Grids, pp. 208–212 (2017)
31. Goldstein, P.J., Katz, R.N.: Academic analytics: the uses of management information and technology in higher education. ECAR Research Study, vol. 8 (2005)
32. Oblinger, D.G., Campbell, J.P.: Academic Analytics. EDUCAUSE White Paper (2007)
33. Norris, D., Baer, L., Leonard, J., Pugliese, L., Lefrere, P.: Action analytics: measuring and improving performance that matters in higher education. EDUCAUSE **43** (2008)
34. Bodily, R., Verbert, K.: Review of research on student-facing learning analytics dashboards and educational recommender systems. IEEE Trans. Learn. Technol. **10**(4), 405–418 (2017)
35. Amine Chatti, M., Lea Dyckhoff, A., Schroeder, U., Thüs, H.: A reference model for learning analytics. Int. J. Technol. Enhanced Learn. **4**, 318–331 (2012)
36. Mukherjee, R., Kar, P.: A comparative review of data warehousing ETL tools with new trends and industry insight. In: 2017 IEEE 7th International Advance Computing Conference (IACC) (2017)
37. Howard, H.: Knowledge Discovery in Databases (2012). http://www2.cs.uregina.ca/~dbd/cs831/index.html
38. Di Martino, B., Cretella, G., Esposito, A., Giulio Sperandeo, R.: Semantic representation of cloud services: a case study for microsoft windows azure. In: 2014 International Conference on Salerno Intelligent Networking and Collaborative Systems (INCoS), Italy (2014)
39. Carreño, J.: Descubrimiento de conocimiento en los negocios. Panorama **4**, 59–76 (2008)
40. Anoopkumar, M., Zubair, R.: A Review on Data Mining Techniques and Factors Used in Educational Data Mining to Predict Student Amelioration, pp. 1–12. IEEE (2016)