# Educational Data Mining to promote active methodologies: analysis of learning patterns in Syphilis courses at AVASUS

**Abstract**

The increase in enrollment of online education in Brazil is currently a phenomenon. Between 2000 and 2013 online education in public institutions increased from 5,287 to 515,405 students. Today, only the Virtual Learning Environment of the Unified Health System (AVASUS abbreviation in Portuguese) already has more than 340,000 users, 729,000 enrollments in 173 courses. It is important to say that OERs that we design and offer at AVASUS for the Ministry of Health are aimed at contributing to online lifelong learning for healthy living and well-being. These students have increased the records of activities and actions performed in the AVASUS Databases, in which, they are already approaching the 1 Terabyte mark. The problem is that records are most often used only during the execution of a curricular component and almost always for monitoring and defining student outcomes. From this scenario, the general problem of this project arises that is to develop technologies that capture, interpret, validate and return to the student different training and updating possibilities to obtain a desired competency. For this, we analyze important data that can subsidize computational resources that allow to investigate the temporal patterns of learning of the students and the aspects related to the courses, for example, their training paths and quality of learning objects. Using Learning Analytics and Educational Data Mining (EDM) techniques, we analyze three basics aspects of learning patterns: Cognitive, Metacognitive, and Resource Management data. Through the use of EDM it was possible to discover the most commonly patterns of use presented by the students and to classify them according to the type of self-regulated learning strategy adopted. For this study, we used courses aimed at combating Syphilis, currently one of the diseases that receive the most attention by the Brazilian Ministry of Health for prevention and health promotion due to its alarming numbers of infection and epidemiology. In this sense, it is essential to research two well-known processes and guiding the formalization of courses and contents operated through AVASUS technologies: the promotion of active methodology and student engagement. Preliminary data in the selected courses point to a 40% improvement in the time taken to develop the desired competence, a reduction in the congestion rate (those that do not conclude) by 30%, and a success rate of 70% after track development based on student data. This research is still under development for the 300,000 AVASUS students.

**Key Words:** Learning Analytics, Educational Data Mining, On Line Education, AVASUS, Lifelong Education

## Introduction

E-learning has in Brazil in the last 15 years its great rise. Data from the Brazilian Society for Distance Education point to an increase of more than 500 times the number of students and new enrollments in this modality in the country, following a worldwide trend of increases in new information and communication technologies for the educational process.

E-learning has experienced a significant increase in the offer of courses in the most diverse areas, levels and formats, an aspect that coincides with the beginning of the use of Information and Communication Technologies (ICT) as mediators of this process, mainly the Internet and the Learning Management System (LMS).

The increase in enrollments in the e-learning can be verified through data from the Statistical Synopsis of Higher Education, produced by the National Institute of Educational Studies and Research (INEP abbreviation in Portuguese), which show that between the years 2000 and 2005 the e-learning increased from 5,287 admissions to 127,014, reaching 431,597 in 2011 and 515,405 in 2013.

As such, the number of enrollments is not enough, e-learning in Brazil has expanded its areas of expertise, reaching as soon as technical areas, as well as its ability to simulate, increment and communicate new elements through increasingly sophisticated and intelligent virtual environments. These environments, which were initially intended only for the availability of virtual interaction and content availability tools, have, with the advent of the analysis and mining of large data, understood the forms of teaching and analyzed the ways of learning of the students.

The Federal University of Rio Grande do Norte (UFRN abbreviation in Portuguese), has been at the forefront of the development of pedagogical materials and content available to be worked and interacted in e-learning since 2003, with undergraduate and postgraduate courses in several areas, with emphasis on Human, Exact and Biological Sciences.

As part of this effort and in partnership with the Laboratory of Technological Innovation in Health (LAIS abbreviation in Portuguese), since 2013, both SEDIS and LAIS, develop a important project to improve, amplify and re-program their virtual environments and new proposals to meet the area of health and the training of professionals focused on health education.

**The AVASUS**

Learning Management System of the Unified Health System (AVASUS abbreviation in Portuguese) was developed based on the Moodle (Modular Object-Oriented Dynamic Learning Environment). Moodle is a open source learning platform available under the GNU General Public license.

Because it is open source, Moodle can be customized and modified for the needs of those who install it. Its modular feature allows you to create plugins and themes without the need to modify its central code, also allowing integration with external applications, among other possibilities. In addition, Moodle has a great scalability, being able to support from few users, to millions (Moodle, 2019).

Taking advantage of all these characteristics, AVASUS was developed, initially in a pilot project, and is currently in version 2.0. It has an enormous amount of users, close to reaching the mark of the 400,000 users registered. The total number of courses exceeds 180 and the number of enrollments are close to 800,000, with more than 400,000 certificates issued.

Therefore, the notion of the size and importance of this environment in the training of health professionals in Brazil, serving all 27 states of the country, especially with contents related to prevention, health promotion, worker health and health care.

Since 2015, AVASUS was required to develop, qualify, implement and analyze, computational elements of comprehension of the student profile, teaching materials and the media developed especially for the set of courses related to the combat of Sexually Transmitted Infections, particularly to Syphilis, on the occasion of the project "Sífilis não!", a project also developed in LAIS and that gathers in one of its goals the mass education and communication through the mapping as well as the analysis of learning of the students contained within the itineraries of related knowledge with IST, and syphilis.

***Project "Sífilis Não"***

It is important to understand why there is currently a project in Brazil to combat Syphilis and the training of professionals to this end, with the development of learning analytics techniques to improve the short-term performance of the practice of health professionals in the networks health care.

Syphilis is highlighted in these proposals, as this infection in the last decade has shown an increase of approximately 5,000% in their cases, according to data from the Ministry of Health. Thus, in comparison with other types of epidemics, this is an alarming fact, since in addition to having extremely high transmissibility data, its epidemiological rate is overcoming many other types of epidemics, including other Sexually Transmitted Infections (STIs), overloads the public health system in many aspects.

Comparing with the data from the year 2016, there was a 28.5% increase in the detection rate in pregnant women, 16.4% in the incidence of congenital syphilis and 31.8% in the incidence of

acquired syphilis, which increased from 44/100,00 in 2016 to 58 cases per 100 thousand in 2017. According to the Epidemiological Bulletin of Syphilis (2018), in congenital syphilis, Porto Alegre, Recife, Natal, Fortaleza, Manaus, Rio de Janeiro, Salvador, Teresina, João Pessoa, Florianópolis, Maceió, Campo Grande, Belo Horizonte, Palmas, Vitória, São Luís and Porto Velho showed a higher incidence rate than the rest of Brazil, observing that the cities of Recife, Natal, Fortaleza, João Pessoa, Porto Alegre and Teresina had a higher incidence rate of congenital syphilis than those of syphilis in pregnant women. (Image 1)

In 2016, in addition to these cities, Maceió and Aracaju also presented congenital syphilis rates higher than those of syphilis in pregnant women.
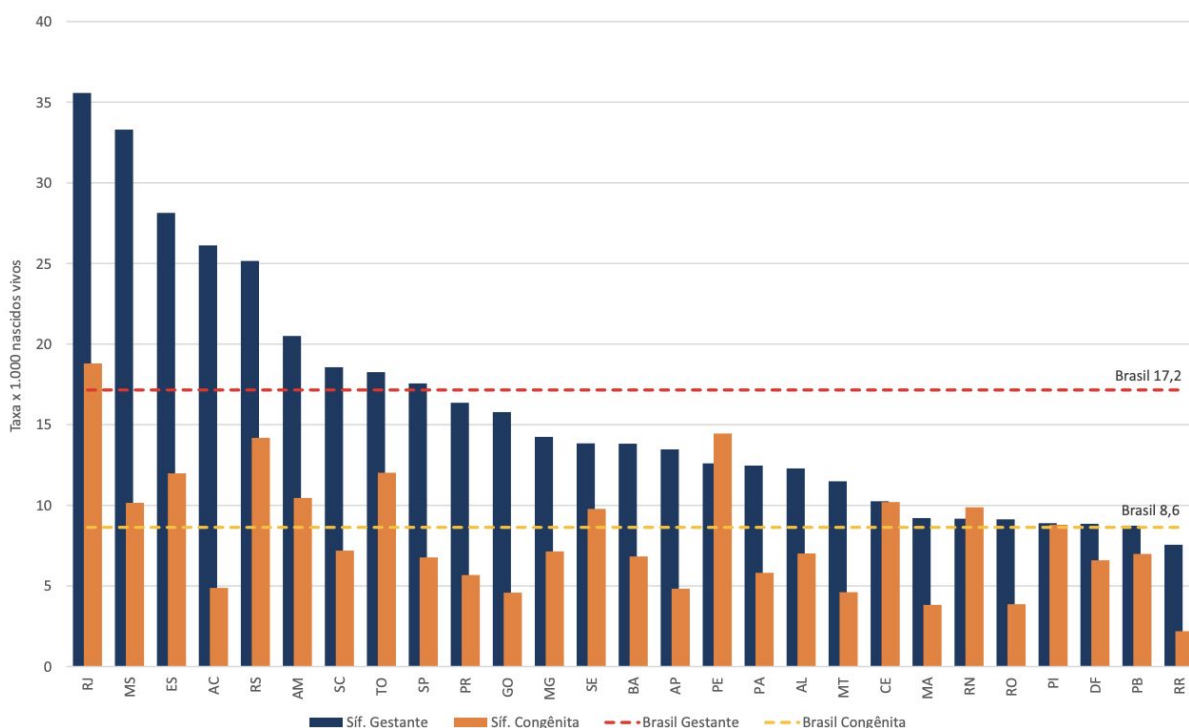


**Image 1: Rates of congenital syphilis by State. Source: Notification System of Sinan (Sinan), updated on 06/30/2017.**

Congenital syphilis data are more serious than those of acquired transmission, since in most cases it surpasses the average of gestational syphilis and acquired syphilis, in an upward trend, accelerated and recurrent increase, as shown in the graph below:
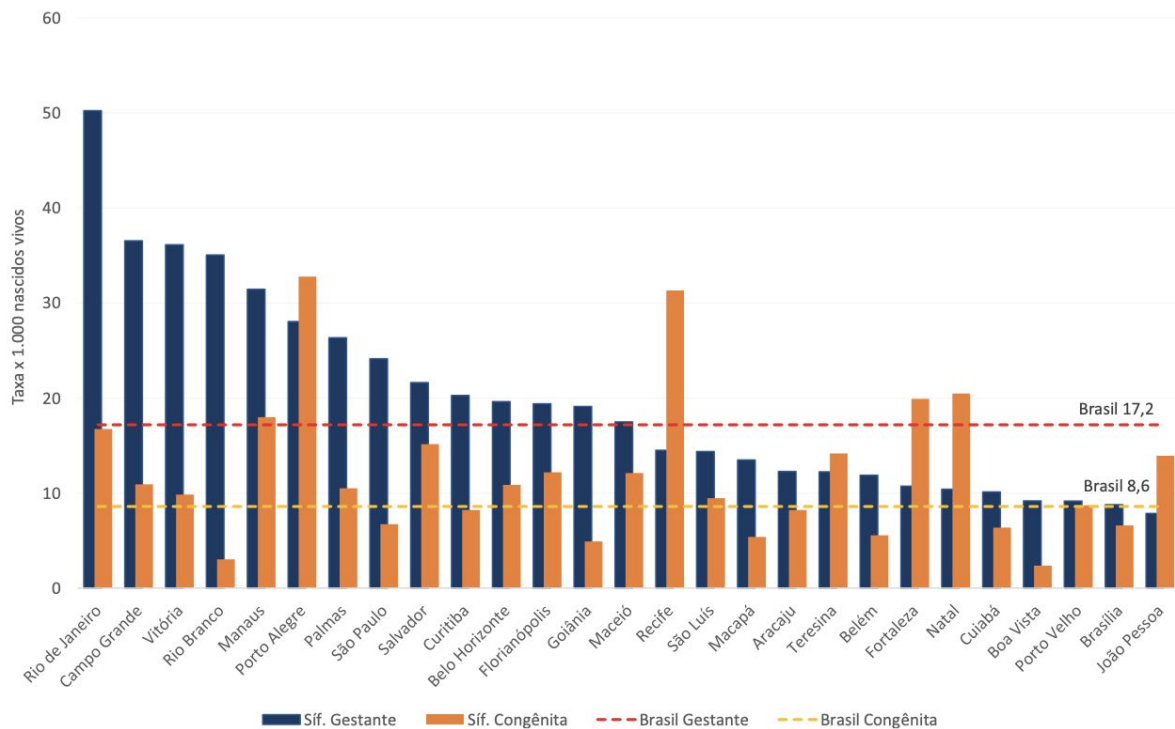
**Image 2: Growth rate of congenital syphilis compared to maternal syphilis by State. Source: Notification System of Sinan (Sinan), updated on 06/30/2017.**

According to the Epidemiological Bulletin of Syphilis(2018), when analyzing the historical series of reported cases of syphilis, 249,852 (39.9%) of them occurred in men and 376,886 (60.1%) in women, of whom 169,339 (44.9% %) reported as acquired syphilis and 207,547 (55.1%) reported as syphilis in pregnant women, showing growth in all regions of Brazil since 2010, the initial year of compulsory notifications.
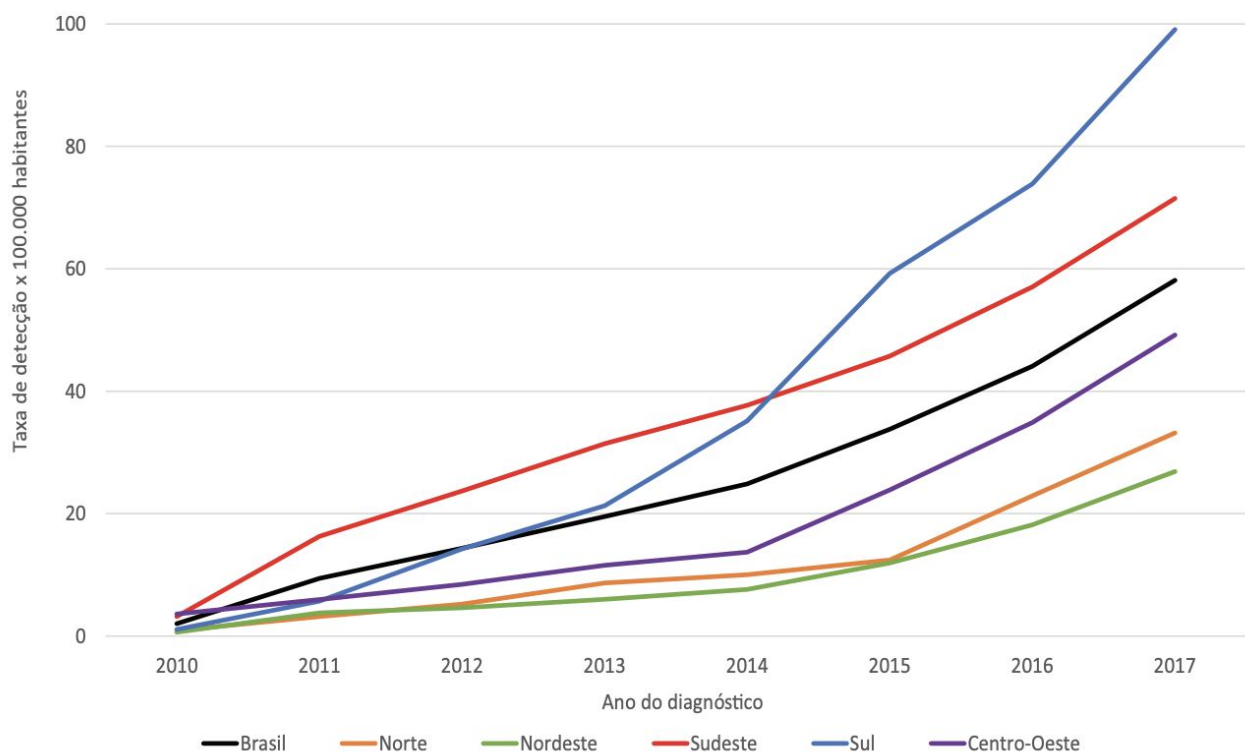


**Image 3: Growth rate of all types of syphilis, by historical series in Brazil since 2010. Source: Notification System of Sinan (Sinan), updated on 06/30/2017.**

*Syphilis combat by courses*

From the general situation of syphilis in Brazil and the epidemiology linked to it, there is in fact a need to create and strengthen the health care and promotion network, especially among health professionals who need to support and strengthen the fight against sexually transmitted diseases with a focus on prevention, treatment and knowledge about the disease, forms of transmission and transmission.

The project of training, development and follow-up of courses directly related to the general proposal to combat syphilis. Currently 7 courses on AVASUS are focused on or related to the fight against Syphilis:

- **Primary Health Care, Family Health Strategy and Territorialization**: has a 30-hour workload and 1501 enrolled students. This module addresses the work process in family health care, reflecting on the process of territorialization, bonding, coordination of care, and the attached population;
- **National LGBT Comprehensive Health Policy**: has a workload of 45 hours and 5201 students enrolled. This module was organized in a partnership between the Ministry of Health and the State University of Rio de Janeiro through the UNA-SUS, contributing with health professionals and SUS workers to carry out their care, promotion and prevention actions with quality, guaranteeing the LGBT population access to integral health;
- **Expanded Clinic and Matrix Support**: has a workload of 30 hours and 7104 students enrolled. This module will discuss the expanded conception of health and also its organization of the work process, together with clinical management tools, as well as the concept of Matrix Support and its application by the teams of the Family Health Support Centers and the Support Centers Psychosocial, in addition to the concept of Unique Therapeutic Project;
- **Observation in the Health Unit**: has a workload of 30 hours and 649 students enrolled. It is a module that was initially offered for specialization classes in Family Health Care and was open to the general public. In it is presented to the student how the Health Unit works in the organizational structure of the Unified Health System, SUS. The student will be able to know the Health Unit as a point of attention in the Network of services, with information on organization and functioning, aiming at improving services and improving quality and access;
- **Syphilis: Have a seat here comes information!**: has a workload of 5 hours and 821 students enrolled. It is a module that brings together the ten episodes of the web series "*Senta que lá vem informação*!" Produced for the Sífilis Não Project. There are interviews and conversations that present information about Syphilis, as well as the correct way to diagnose and treat it;
- **The care of people with HIV/AIDS in Primary Care**: has a workload of 120 hours and 8527 students enrolled. The course is designed to train health professionals from the basic network for shared care of people living with HIV/AIDS, focusing on Primary Care, linking it to other points of attention of the SUS network. The course uses a problematizing methodology, whose cases will be presented in comics, to guide the various topics covered. The contents are organized in four units and permeate all aspects related to the care of people living with HIV/AIDS, from combined prevention to the use of TARV. As the infection of the patient with HIV/AIDS occurs in a similar way to that of the patient with Syphilis, this module can be correlated and its knowledge also used in this case;
- **A Virtual Visit to the 2nd International Conference on Health Innovation**: it has a workload of 45 hours and 170 students enrolled. It is a module that highlights the debates held at the 2nd International Conference on Health Innovation promoted by the LAIS, focused on the evolution and modernization of solutions for health systems and combating diseases and endemics, such as the problem of syphilis, and thereby promote the reduction of costs and investments both at the national and global levels.

In these courses the entries are continuous and without specific calendar. In addition, students do not have specific deadline to finish the courses and their workload has no daily limit, nor monthly, only full term.

The general data of the courses, until the end of June are:

| Course | Enrollments | Completed |
|---|---|---|
| A Virtual Visit to the 2nd International Conference on Health Innovation | 249 | 80 |
| The care of people with HIV/AIDS in Primary Care | 8,616 | 4,559 |
| Syphilis: Have a seat here comes information! | 930 | 513 |
| Observation in the Health Unit | 658 | 356 |
| Expanded Clinic and Matrix Support | 7,123 | 3,941 |
| National LGBT Comprehensive Health Policy | 5,232 | 3,075 |
| Primary Health Care, Family Health Strategy and Territorialization | 1,595 | 850 |
| Total | 24,403 | 13,374 |

**Table 1: Enrollments and completions from Syphilis area's courses**

Faced with the numbers of users in these courses and the cycles of enrollment, any type of analysis of course format and performance can only be performed by data mining.

In this context, Data Mining has been emerging for the investigation of educational records. With this technique it is possible, for example, to discover patterns of use in AVASUS, as well as to map the behavior patterns of a class, tutors and students. This area of research is called Educational Data Mining (EDM), which can be classified as a sub-area of Learning Analytics. EDM is defined as the area of research that has as its main focus the development of methods to explore datasets collected in educational environments.

Thus, in Learning Analytics research, a list of self-regulated learning strategies indicative of actions taken by students during their training process can be observed.

Today the researchers classify as Cognitive, Metacognitive and Resource Management. Through the use of EDM in LMS records, we found patterns of use most commonly presented by students and classified them according to the type of self-regulated learning strategy adopted.

### AVASUS e Learning trails: LA models applied to learning

In AVASUS, the analysis of learning through technology, is seen in a way to customize the journey of each student according to concrete data. It is through this investment of increment of technological interpretation that the learning analytics can interpret individually the course, the difficulties and the habits of studies of the students. In addition, the educator becomes capable of making sound pedagogical decisions and reorienting practices to better address mapped needs, measuring the impact of these changes in the digital classroom and, in a broader context, the impact of large-scale education.

The first innovation of this environment was to think of content through Active Methodologies and problematization. Active methodologies and pedagogy of problematization create teaching models, which value dialogue and demystify reality. With this, the objective is the social transformation through a conscientizing and critical practice. In this context problematizing is not restricted to just presenting questions, going beyond these until arriving at the discussion of the conflicts that are part and keeps the presented problem.

The first intervention in the environment so that he could along with the content favor the student is to make it more practical and functional. As the base engine of the environment, we use the Moodle API (Application Programming Interface), with several changes in it as addition of new features as well as database configurations. Development services for adaptations and changes are contained in the following dimensions:

- **Deployment of the functionality**: sending messages using the integrated messaging API with functions of the local function library. The programming takes into account the context of system profiles, turning the back-end into a room where a student can communicate with the teacher;
- **Creation of activity modules** such as the dependent task module and mural. Each module requires the creation of registration and editing forms added to the creation of new tables and fields in the database for the persistence of the information that is passed by the users through the front end, all of this together with changes of business rule proposed by the contractor;
- **Validation of activity creation forms** using sever-side type processing languages, such as PHP and client-side, such as javascript;
- **Assignment of values to environment constants** that serve as reference parameters for functions used in modules that are contained within a module unit. A module unit can contain multiple modules and these modules must share information in common;
- **Adapting the local library** of the modules units, adapting the returns of their functions;
- **Change and create queries**, procedures and functions in the database using the SQL language;
- **Adaptation of SQL queries** to functions of module libraries and module units, as well as new procedures calls and functions stored in the database.

AVASUS learning analytics are based on Big Data and Moodle native table analyzes, with plugins developed to control or access user elements or linked to users. Basically, AVASUS uses technologies that allow you to compile and relate large amounts of data, producing relevant information based on them. It then makes it possible to make future predictions based on the previous behaviors.

For the data mining analyzes in syphilis control courses, data related to other courses were included in order to study the variables of access, permanence, use of tools and success in completing the courses or their respective itinerary. Two blocks of courses were considered. The block 1 contains current courses of the AVASUS but that are not within the block of courses to combat syphilis. Block 2, the 7 courses to combat syphilis. The objective was to monitor, recognize and analyze the learning difficulty data of students in block 1, change the media and the profile of the courses for students in block 2 and analyze the results of improvements.

The idea at this stage is to understand which improvements or changes in the users' learning profile have a direct influence on the use of groups of students and the types of activities that the students perform.

Based on the mining of the AVASUS database, a series of difficulties was registered for the students, especially in the types of contents that were available for learning. In the free courses, not related to syphilis group courses, it was verified that:
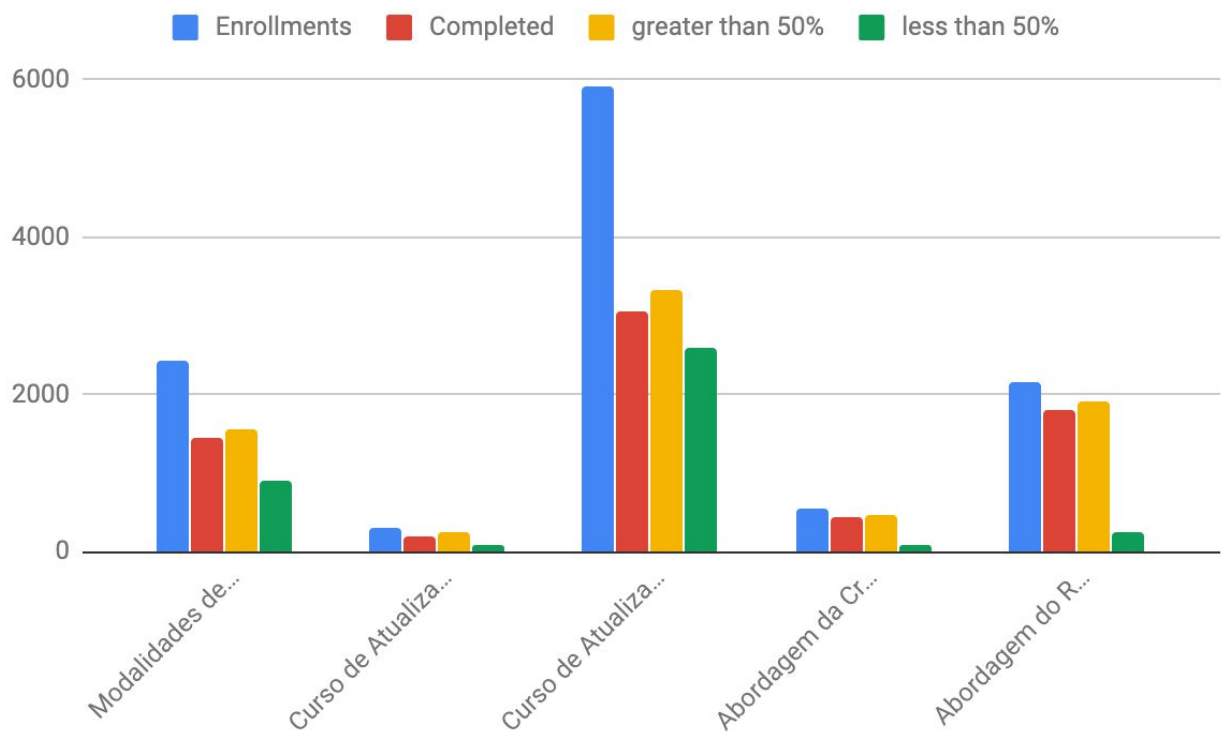
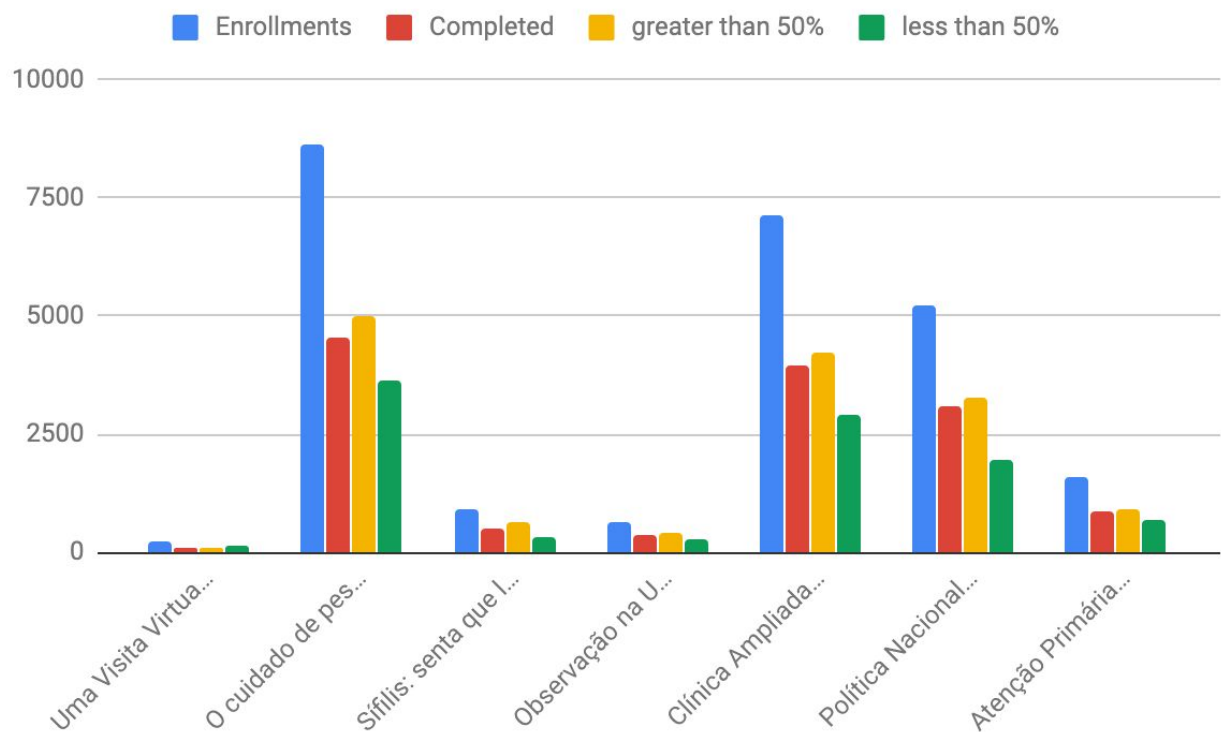**Image 4: Pre-intervention courses**



**Image 5: Learning analytic intervention courses**

The relationship between enrollees and graduates after indicating the syphilis courses for profiles of workers more linked to professional practice, provided the lowest distance between enrollees and graduates decreasing evasion and increasing the level of the graduates by more than 60%.
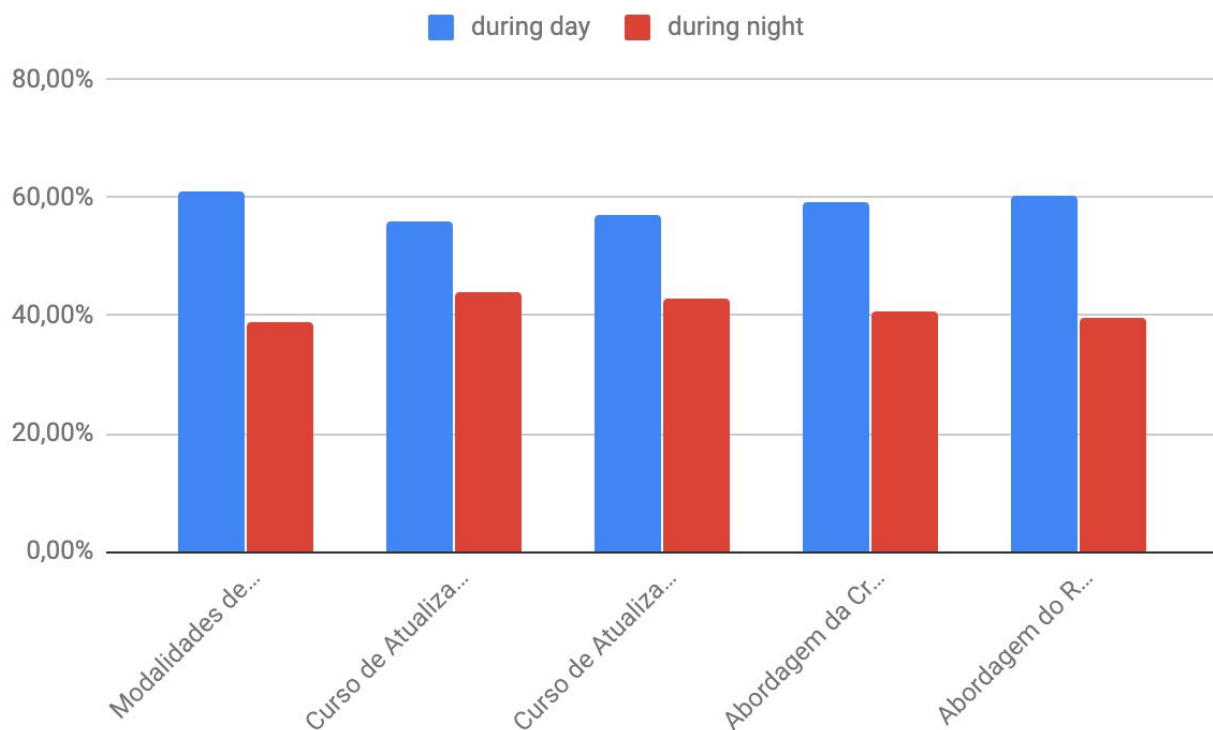
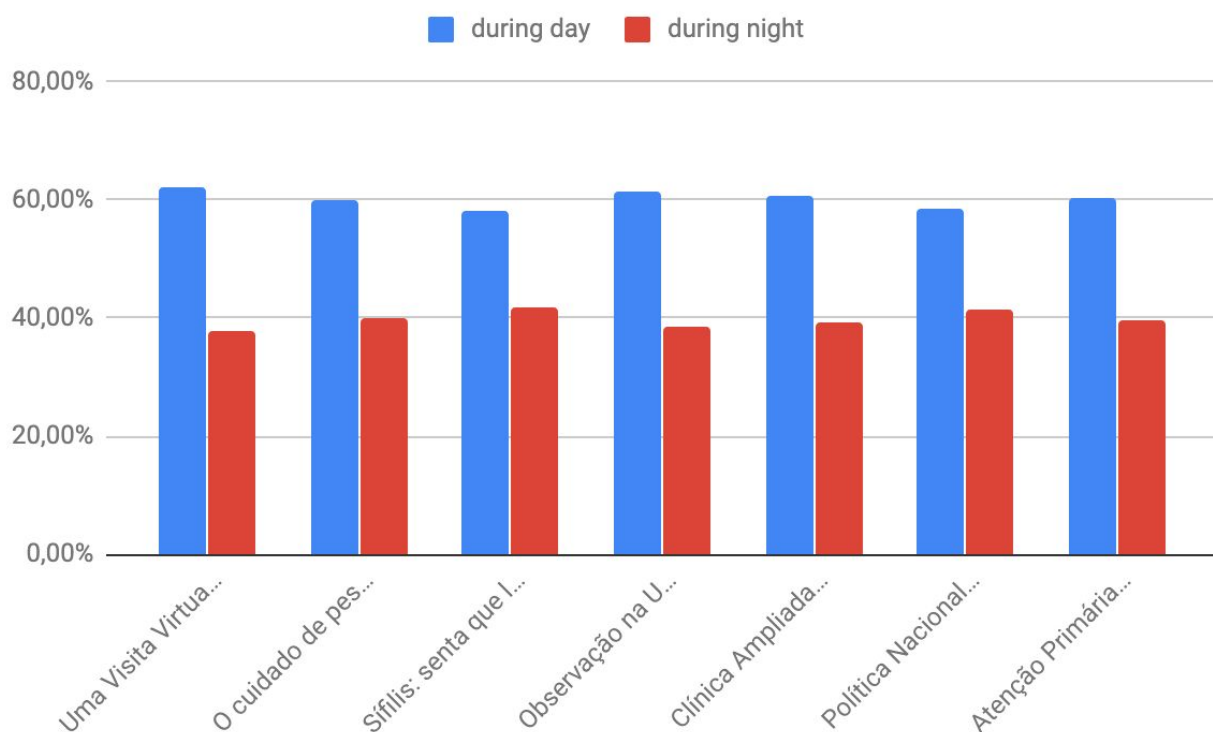**Image 6: Pre-intervention courses - access**



**Image 7: Courses after the intervention - accesses**

The courses in block 1 tended to have less distance between students' accesses by day and night because they do not find specific contents that favor their respective performances. In the courses in block 2, students further mark their entries at specific times because they have specific methods and workload balanced enough to draw a recursive study, so they access more during the day than the evening, even considering the practical content

Finally after the analysis of the more than 24 thousand students enrolled in block 2 compared to the first block reveal a curious fact. If we analyze the achievement in terms of completion, in the relation between material diversity and completion levels, the courses in block 1 has 60.8 of the students in conclusion, while in the courses of the other block there are 54.8. Even in the data of achievement, considered between more and less than 50%, the courses in block 1, presents more students with completion of more than 50% of the courses, reaching 59.7% of the students while in block 2, the quantity of students with achievement greater than half reaches a little more than 53%. However, when crossing two other variables, the average completion time with the quantity of materials available, even considering the time load, the courses in block 2 perform better than that of block 1, both in decreasing the average time to perform the course, and the permanence of the user on the platform. The average completion from the courses from the block 2 was 8,72 days while in the block 1 was 30 days, even with less hours and content. This indicates that even in courses with more content, but based on indicative content by profile and in the diversity of transmigration, students tend to complete the courses in a shorter average time than students who do not have material diversity. This causes students to access courses less than 50% faster, but they give up or interrupt their courses, increasing the average completion time in case of block 1 reaching more than 100 days for some scenarios, whereas in block 2 the largest time was 28 days in average time.

## Conclusion

To perform this research, some structured data mining algorithms were used, grouping not only by student profile but also functional clusters. The intervention libraries installed in the AVASUS considered dynamic groupings, as there were improvements in levels of achievement, permanence, average time and student evaluation, mining libraries considered the best and worst scenarios simulating data of better use and better use of the student considering, professional profile, sex and age.

The proposed method should be improved and incorporated into the AVASUS system procedure in the next 12 months to ensure that the described variables are safe for the treatment and improvement of student profiles and improvement in established performance.

The methods of learning analytics point to better evaluation - 100% of the courses in block 2 had a maximum evaluation performed by the students according to the material, system, availability and support of the system - and the best rate of success of the users.

However, when related to block 1 with block 2, the data of completion of course and achievement above of 50% in a shorter time, are in all the better simulations. This occurs both because there is a numerical difference in the blocks of courses, that is, block 2 has more courses than block 1, or because the elements of transmigration tend to be more effective when the time series considers a longer time interval.

In this way, the mining algorithms for cross-checking the data of learning analytics, even if they are data separation and clustering to classify them, although they do not have elements of effective database library for intervention and process automation, already point to a diagnosis in the type of student, support, subject and materials needed for communication and a global education of a professional inserted in methods of lifelong learning.

## References

AFSHAR, M. H. et al. (2017). *Exploring the efficiency of harmony search algorithm for hydropower operation of multi- reservoir systems: A hybrid cellular automat-harmony search approach*. In: SER, J. D. (Ed.). Harmony Search Algorithm. Singapore: Springer Singapore.

ALIA, O. M.; MANDAVA, R.; AZIZ, M. E. (2011). *A hybrid harmony search algorithm form ribrain segmentation*. Evol. Intell.

AVASUS (2019). Retrieved from https://avasus.ufrn.br

GARDELLA, C (2003). *Sexually transmitted infections in pregnancy: treatment options. Current Treatment Options in Infectious Disease*. Switzerland, Springer Nature Switzerland.

LINDEN, R (2008). *Algoritmos Genéticos*. 1. ed. Rio de Janeiro: Brasport.

MACHADO, Aydano P. et al (2010). *Mineração de Texto em Redes Sociais Aplicada à Educação a Distância*. 2010. Retrieved from
http://pead.ucpel.tche.br/revistas/index.php/colabora/article/view/132

MAHDAVI, M.; ABOLHASSANI, H (2009). *Harmony k– means algorithm for document clustering*. Data Min. Knowl.

MEENA, Y. K.; SHASHANK; SINGH, V. P (2012). *Text documents clustering using genetic algorithm and discrete differential evolution*. Int. J. Comput. Appl.

Ministério da Saúde. (2017 May 30). *Ambiente Virtual de Aprendizagem do SUS (módulos Educacionais) – (AVASUS).* Retrieved from
http://www.saude.gov.br/artigos/674-assuntos/trabalho-e-educacao-na-saude/40535-ambiente-virtual-de-aprendizagem-do-sus-modulos-educacionais-avasus

Ministério da Saúde (2017). *Boletim Epidemiológico da Sífilis 2017*. Retrieved from
http://www.aids.gov.br/pt-br/pub/2017/boletim-epidemiologico-de-sifilis-2017

Ministério da Saúde (2018). *Boletim Epidemiológico da Sífilis 2018*. Retrieved from
http://www.aids.gov.br/pt-br/pub/2018/boletim-epidemiologico-de-sifilis-2018

Moodle. (2019 June 29). *About*. Retrieved from https://docs.moodle.org/37/en/About_Moodle

NASCIMENTO, J. P.; VIEIRA, M. G (2016). *Os desafios da institucionalização do ensino superior na modalidade a distância: a visão dos gestores de uma universidade federal*. Revista Ensaio: Avaliação e Políticas Públicas em Educação.

PICHILIANI, Mauro (2008). *Data mining na prática: regras de associação*. Retrieved from
http://imasters.uol.com.br/artigo/7853/sql_server/data_mining_na_pratica_regras_de_associacao

ROCHA, Marcus V.; DA COSTA, Mateus Conrad B.; DOS SANTOS NETO, Pedro de Alcântara (2002). *Busca por Frases em Bancos de Dados Textuais*. UFMG. Retrieved from
http://homepages.dcc.ufmg.br/~nivio/cursos/pa02/seminarios/seminario3/seminari3.html

STEWART, C (2017). *Learning analytics: shifting from theory to practice*. Journal on Empowering Teaching Excellence.

WILLIAMS, C. J.; MATT, J. J.; O'REILLY, F. L (2014). *Generational perspective of higher education online student learning styles*. Journal of Education and Learning.