CHAPTER

53

Data analysis of students who used the MAZK Intelligent Tutor System during the pandemic period

https://doi.org/10.56238/emerrelcovid19-053

Eliane Pozzebon
Laboratory of Computational Technologies - LabteC
Federal University of Santa Catarina – UFSC.
E-mail: eliane.pozzebon@ufsc.br

Felipe Zago Canal
Laboratory of Computational Technologies - LabteC
Federal University of Santa Catarina – UFSC

Raissa Silva
Laboratory of Computational Technologies - LabteC
Federal University of Santa Catarina – UFSC

Susi Machado de Oliveira
Laboratory of Computational Technologies - LabteC
Federal University of Santa Catarina – UFSC

ABSTRACT

In this article, we present an analysis of data from an Intelligent Tutor System (ITS) created by a group of Brazilian researchers. The STI was used by approximately 19,000 users from elementary to higher education in public schools. The research methodology used was quantitative and qualitative exploratory analysis. The data set for this study was obtained through the students' interactions with the STI during remote teaching. In the data analysis, information from the students' profiles (age, frequency of access, etc.) and interactions were used, for example, the exercises were solved with objective and descriptive questions to identify the student's abilities and challenges. A total of 2,514,219 responses were analyzed in the STI until 05/31/2021, with 1,533,946 answers to objective questions and 980,273 answers to discursive questions. In the results, it was possible to observe a tendency towards the mastery of mathematics and languages and an excellent performance in 69% of the total answers, which obtained maximum scores in the questions.

Keywords: MAZK Intelligent Tutor System, students, pandemic period.

1 INTRODUCTION

The technologies with Artificial Intelligence have caused changes in the most diverse areas, including education. These changes represent challenges, risks, and opportunities to ensure inclusion and equity, qualify teachers, improve educational practices, and ensure ethics in information management (PEDRÓ, 2019) (BARRIOS-TAO, 2018).

According to Vicari (2018), the main educational systems that use these technologies are Affective Intelligent Tutoring Systems (STIs), Learning Management Systems (LMSs), Intelligent Educational Robotics, and Massive Open Online Courses (MOOCs), regard Learning Analytics (LA). However, each of these applications makes use of AI techniques in different ways.

The use of STIs in classes provides benefits to students and teachers (Bittencourt, 2021). One of the main contributions to the use of STIs is to increase the speed with which students take ownership of the proposed activities. With intelligent tutors, students can solve the activities and get ahead without the need for intervention from the teacher to clarify doubts or evaluate the result. The tutors also suggest other content for each student according to the way they solved the previous ones. The study becomes more efficient, practical, and agile. It also has the benefit of geographical freedom for both the student and the teacher who can from anywhere create and solve the activities. Students become more motivated and teachers gain time and can devote themselves to other tasks.

STI MAZZ is an intelligent tutor for teaching and learning various topics. The MAZK identified and stored the level of knowledge of the user, as well as the difficulty of the exercises that are adjusted according to the interaction of the student with the tutor. The software promotes learning adaptively and collaboratively and can be accessed on computers or from other devices such as mobile phones or tablets. In MAZK, the teacher can accompany students during the learning process synchronously through the virtual room feature. All interaction information within the MAZK is stored in the student, pedagogical and domain models (SILVA et al., 2019; MORO et al., 2019).

It is of paramount importance that the data on the learners are recorded in the STI MAZK, which presents the following structure to store the information: The learners who use the Intelligent Tutor perform activities, generating data. First, the learner performs a registration in the software, soon after viewing the home page of the Tutor, this page contains the main menu by areas, the experiences (Experiences of the apprentices called in the XP software), the successes and errors, the use of each apprentice and a ranking of all users. The learner uses the Intelligent Tutor a virtual room through a code made available by the teacher where he will interact with other colleagues and with a facial recognition reader that captures and records the emotional state of the learners. From the data generated, we can then individually understand the skills, competencies, and challenges of each learner and create strategies to improve the learning process of students.

The objective of this article is to present the results of the analysis of the data of the learners who interacted with an Intelligent Tutor System. The data set for this study was obtained through the students' interactions with the STI during remote teaching. In the data analysis, information from the students' profiles (age, frequency of access, etc.) and interactions were used, for example, the exercises were solved with objective and descriptive questions to identify the students' skills and challenges. The article is organized into sections on STIs, related works, methodology, results, and final considerations.

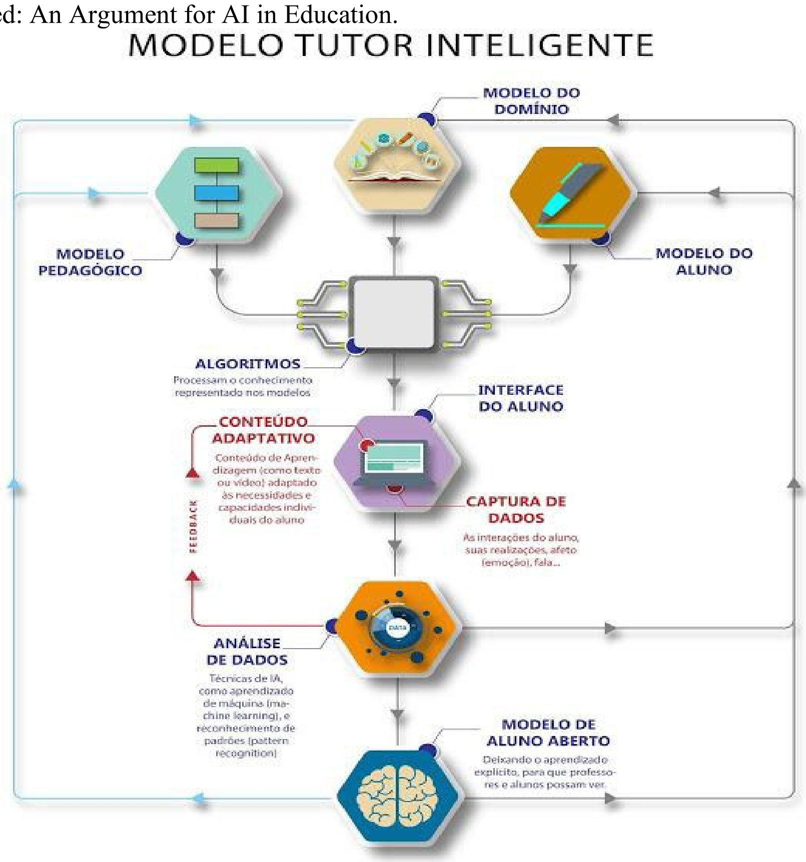
2 INTELLIGENT TUTOR SYSTEMS

For the STI to be efficient it needs to be developed containing targeted modules to cover what will be taught (domain model), as the content will be taught (pedagogical model), and for whom (the knowledge that the student already possesses, student model). And this traditional basic architecture can be found in authors such as Murray (1999); Pozzebon (2008); Santos, Menezes, and Cury (2018).

The infographic (Figure 1) shows the initial information that will nourish the STI. The student model takes into account the student's level of knowledge; By participating through interactions with the STI, the system itself levels its knowledge. According to your accesses, the system will provide opportunities for topics of interest. As Moro (2019) explains, in an STI this monitoring of access takes place through the verification of its log, where the student can access the virtual room, answer the proposed materials, and also participate in self-knowledge courses. An "X-ray" of the student's performance and level of understanding. As a support, the system offers activities in the areas that need to improve its performance.

Two types of information can be stored in the learner's model and are called static or dynamic information. Static information, according to Arcos et al (2000), are information such as the student's name, age, and gender, and is not changed by the system, and appears only explicitly in the interface. The dynamic information is that which determines the progress of the student in learning, verified by the system logs and that occurs by the interaction, as is the case of the observation of the level of knowledge and the preferences of content (JOVANOVIC, 2020).

Figure 1: Infographic developed by doctoral student Paula Toledo Palomino, from ICMC, from an image available in the book Intelligence Unleashed: An Argument for AI in Education.



To compose this work, we studied the modules of students in Intelligent Tutoring Systems, defined in the literature by Chrysafiadi and Virvou (2013), Pozzebon (2008), Gasparini (2013), among other authors: Overlap Model; Stereotype Model; Disturbance Model; Machine learning techniques; Cognitive Theories; Fuzzy Logic; Bayesian Networks; Based on ontologies and Hybrids.

Several related works show implementations in intelligent Tutor Systems such as a study by authors Okubo et al. (2017) who used a method to predict students' final grades by a Recurrent Neural Network (RNN); Zakzewska (2018) who presented an agent-based recommendation system; Peñafiel et al. (2018) used an application of data mining using computational techniques, such as text mining and sentiment analysis, to evaluate the open questions of online surveys conducted by university professors; Islam, Soddiqui, and Aljohani (2019) made an exploration of the various aspects of student interaction data using data mining techniques such as clustering and association to identify relevant patterns of behaviors; Uzir et al. (2020) presented a new learning analysis methodology that combines three complementary techniques – agglomerative hierarchical clustering, epistemic network analysis, and process mining. The methodology allows the identification and interpretation of self-regulated learning in terms of the use of learning strategies. Guo, Yan, and Li. (2020) presented an LA-based method, which has a structure designed and implemented to meet the problem of predicting students' attitudes toward certain blended learning classes; Jovanović et al. (2020) aimed to establish an explanatory model of student behavior by identifying patterns in online activity, offering new opportunities to identify patterns that can be easily interpreted by instructors.

3 METHODOLOGIES

The approach proposed in this work has three main entities that can be visualized: (A) data collection of an intelligent tutor system during remote classes, (B) tabulation and elaboration of graphs (c) and finally, writing about the results obtained.

The methodology used was an exploratory study to obtain information about the learners who used the STI MAZK. Investigative approaches were carried out and some interesting information about this technological resource was found. Data collection was performed in the STI MAZK database and Phyton® was used for the extraction and generation of graphs.

The STI used in this study was MAZK which has three types of users: student, teacher, and coordinator. The student/learner, when registering and initiating interactions with the system, is presented with various contents that, over time, are adapted by the tutor to their profile. In the data analysis, information from the students' profiles (age, frequency of access, etc.) and interactions were used, for example, the exercises were solved with objective and descriptive questions to identify the students' skills and challenges. A total of 2,514,219 user responses in the STI were analyzed as of 05/31/2021.

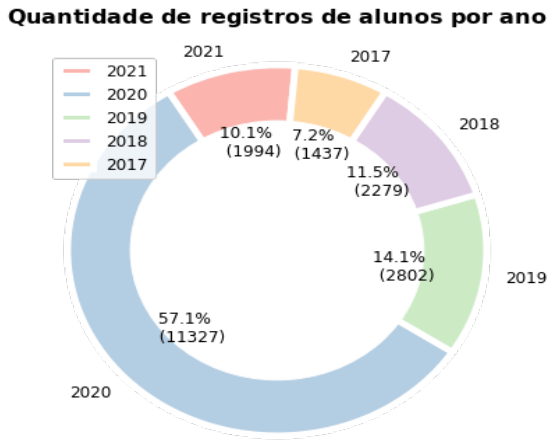
4 RESULTS

The following are the results of the data analysis of an Intelligent Tutor System (ITS) used in remote classes in Brazilian public schools.

4.1 ANALYSIS OF REGISTRATIONS AND LOGINS

According to data extracted from the MAZK intelligent tutor system, the platform currently has 19,839 registered students. Of this total, the frequency of registrations annually was analyzed, obtaining the year 2020 with the highest percentage of registrations, 57.1% as can be seen in Figure 2.

Figure 2: Number of student records per year. Source: Authors



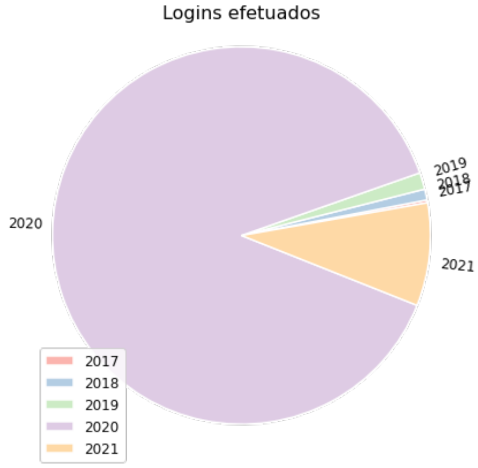
This number can be justified by the high degree of use by schools during the covid-19 pandemic (Figure 3), allowing classes to continue, even remotely. In the year 2021 it can be noticed that there was a drop in registrations, it is believed that those who used some distance learning platform continued its use. A small number of classes adopted new platforms, also for the sake of training teachers and students.

The average age of students on the platform is 18. The age that concentrates the largest number of registrations is 20 years with 2606 users, concentrating 13.13% of the total. The lowest age found was 6 years and 120 years was the highest. It is important to note that data registered by users were analyzed and may contain erroneously registered dates. Above 81 years of age 10 users are obtained and over 60 36 are obtained and these data were omitted from the following analysis for its best presentation. It can be noted that the frequency of users remains high in the range of 19 to 21 years, 3 of the 75 ages found, adding up to 29.69% of the total.

One can also observe balanced numbers without considerable peaks between the ages of 7 and 18 years. Each reserve a percentage between 3% and 5%, which together form 48.10% of the total students, about 9,523 students. These numbers are justified by the high degree of public schools having adhered to the use of STIs during the pandemic, making it possible for classes to be possible during in-person distancing from classrooms. And XXXX as the cradle of the system also justifies the great peak between the ages of 18 to 21 years, the most frequent ages of entry into the academic environment, making good use of the platform in their classes since their first years of operation.

Figure 3: Logins made

Year	Frequency
2017	2205
2018	8366
2019	12846
2020	795391
2021	79163



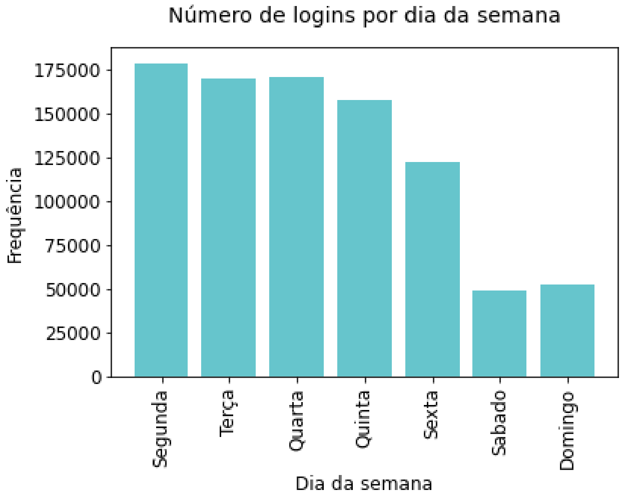
Source: Authors

We also analyzed data from students registered in 2020 to obtain a specific profile of the student who used the platform during the covid-19 pandemic. Compared to the previous overall, we see a large growth in users aged 7 to 15 years, even though the peak has continued between 20 and 21 years, previously justified.

Of the 19839 students registered, 2965 had only 1 login record on the platform, therefore, we have as students who effectively used the STI 16874 students. The average number of logins made is around 225 times per user, with the minimum number of accesses being 1 and 1135 being the maximum. Of the total logins already made, 897971 were from students. This number was analyzed by distributing it annually as well.

We also analyzed the day of the week preferred by the students to use the platform, and obtained the same visible result as in the general scope, being the Monday day of greater access by these users, as shown in Figure 4.

Figure 4: Number of logins per day of the week.



Source: Authors

4.2 STUDENT'S SKILLS AND CHALLENGES

The platform has an individual analysis of the student's skills and challenges that can be verified as soon as they log in to the platform. In this visualization are contemplated the tags that the same has already had contact within the platform and make an average of the amounts of hits and errors of the same displaying this in a graph, so you can have a follow-up of how your learning is going and pay attention to some discipline or material that should have more attention. This section will be discussed the analysis of skills and challenges in a general scope of students.

A total of 2,514,219 responses were recorded on the platform until 05/31/2021, with 1,533,946 answers to objective questions and 980,273 answers to discursive questions. The objective and descriptive questions were analyzed separately:

4.2.1 Objective questions

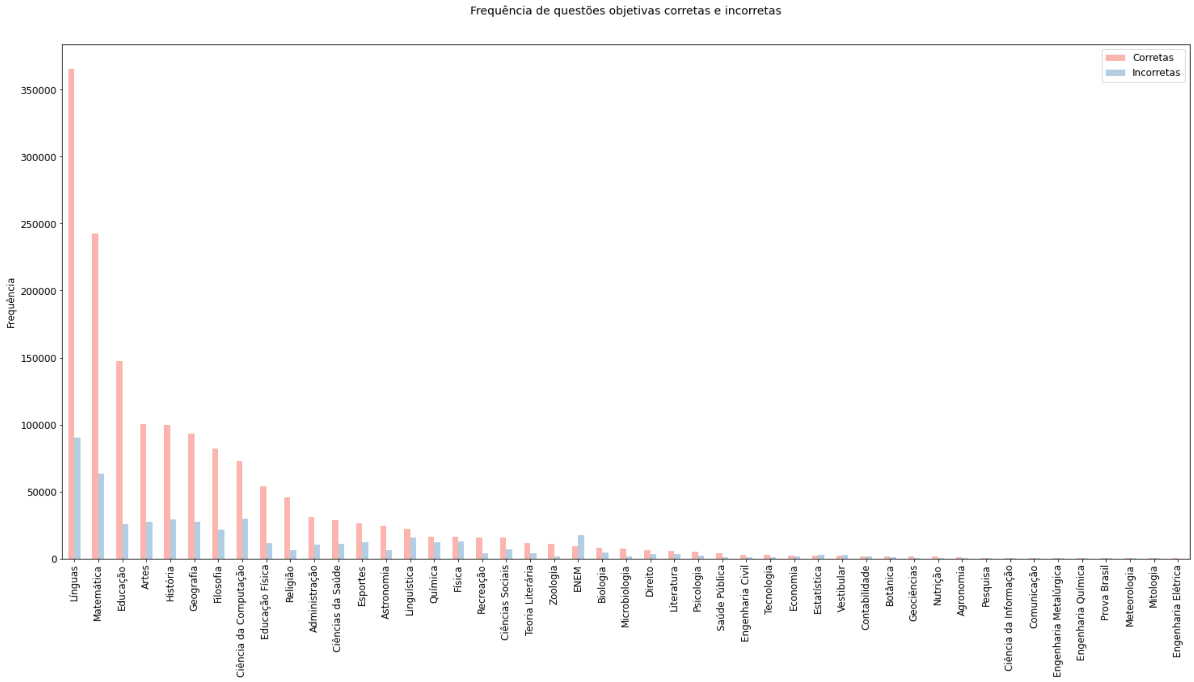
Of the total of 1,533,946 answers to objective questions recorded on the platform, 1,216,961 correct answers, and 316,985 incorrect answers were obtained, demonstrating a great use by the students since 79.33% of the answers to the objective questions were assertive.

It is also possible to verify to which areas of knowledge these questions belong with the analysis of the tags registered in the material. The tags are markings used on the platform as a means of identifying the materials created by the teacher. Until the date of the analysis, there were 788 tags registered, and in times of pandemic, that is, from the year 2020 141 new tags were registered.

It is also important to note that tags can have subtags, that is, a tag can have child tags, as in the case of mathematics which has the child tags 'Differential and integral calculus' and 'Derivatives' for example. In the scope of this analysis, the child tags for their main tags were suppressed, that is, the questions marked with the tags of 'Differential and integral calculus' and 'derivatives' were taken into account as their main 'mathematics'.

Knowing this, in the graph below you can see the numbers of correct and incorrect answers to objective questions according to the main areas listed by the tags, in this case, 47 areas.

Figure 5: Frequency of correct and incorrect objective questions.



Source: Authors

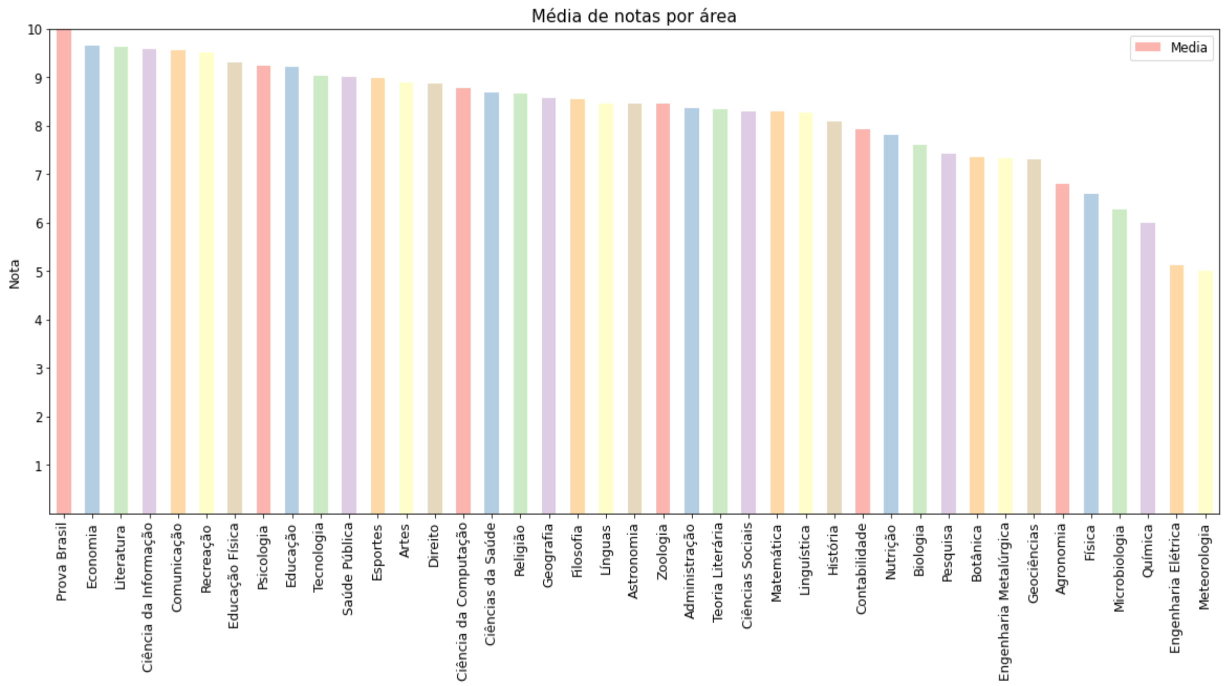
4.2.2. Descriptive Questions

As of 05/31/2021, a total of 980,273 responses to discursive questions have been recorded on the platform. Of this total, about 297 were removed due to invalid data, which makes it impossible to use them in a general context, thus leaving 979,976 records to be analyzed and inferences made.

Of the total, 324,921 responses from students obtained some descriptive feedback from teachers, 33% of the responses were recorded, and the rest were only with feedback from the grade. The grades in turn have two forms of attribution, one is the possibility of automatic grade generated by the STI and another is the one registered by the teacher. Of course, even if there is an automatic grade, the teacher can choose whether or not it is the definitive one for each student. The automatic grades are generated from the comparison between an expected response that was previously registered by the teacher and the response obtained from the student through the artificial intelligence of the system. About 83,386, 8.5% of the registered questions were configured to assign automatic grades by the platform using these expected answers, and of these, 14,324, only 17.17% were accepted by the teacher as a definitive grade for the student.

Taking into account the total number of students on the platform until the date of this analysis, 9,427 students had access to descriptive questions, of which 502 had a zero grade for not having answered any question and about 5% of the total questions still do not have an assigned grade. Of 94% of discursive questions answered by students, the frequency of grades was analyzed by measuring their performance as a whole on the platform. To facilitate the visualization, the decimal places were discarded and only full notes from 0 to 10 displayed in the chart below were considered. It is then seen an excellent performance in 69% of the total answers, which obtained the maximum score in the questions.

Figure 6: Average grades per area.



Source: Authors

It was also considered relevant to analyze and present the average grades by area to which they belong (Figure 6), being able to measure the performance of the students in a more specific way. It can be observed that as in the grades in general, the average per area also has a great result. In all areas, the average was above 5, with 15% of the areas between grades 5 to 7 and 85% above grade 7.

The areas with the lowest averages were Meteorology with 5, Electrical Engineering at 5.12, Chemistry with 5.98, Microbiology with 6.27, Physics at 6.58, and Agronomy at 6.80. As a curiosity, the 3 areas with the best performance were Prova Brasil with 10, Economics with 9.64, and Literature with 9.63.

5 FINAL CONSIDERATIONS

This article presents the analysis of data from an Intelligent Tutor System (ITS) with approximately 19 thousand users from elementary to higher education in public schools and found that the use of Intelligent Tutors in the school context has become an ally of teachers and learners, despite the challenges. As competitiveness increases and challenges increase, the inclusion of technology in and out of the classroom is essential to ensure that learners are prepared to meet the demands of today's professions and especially those that have not yet been created. "The professions of the future."

The current generation of students was born accustomed to the presence of new technologies, but that does not mean that it is easy to engage students and teachers in the processes of digital transformation, this being the biggest challenge faced. So, the school should not see technology as a negative aspect, one should see the positive side of how to work it in teaching to provide better results in learning. Many times, we look at the challenges, and the difficulties, but we cannot see the good side, of how it will be worth overcoming the challenges. And the MAZK Smart Tutor contributes as the numbers in this article show, because being intuitive and easily accessible, it motivates learners to perform more and more tasks within their base of activities, in addition to providing intelligent individualized assistance and promoting social presence.

In 2020, the percentage of registrations increased greatly due to the pandemic and remote classes. The average age of students is from 18 to 21 years, the most frequent ages of entry into the academic environment, making good use of the platform in their classes since their first years of operation. We also analyzed the students' preferred day of the week for STI use and obtained Monday, the day of greatest access by these users.

In the STI analysis, there is an individual analysis of the student's skills and challenges that can be verified as soon as he logs in to the platform. In the analysis, it was possible to verify that the vast majority of the answers were in the area of languages, mathematics, and education, and also that generally, the number of errors does not exceed the correct ones, except for the areas 'Enem' and 'Vestibular'. Another item that drew attention was the two ways of assigning grades in the exercises, one is the possibility of an automatic grade generated by the STI and the other is the one registered by the teacher. Of course, even if there is an automatic grade, the teacher can choose whether or not it is the definitive one for each student. About 83,386, 8.5% of the registered questions were configured to assign automatic grades by the platform using these expected answers, and of these, 14,324, only 17.17% were accepted by the teacher as a definitive grade for the student.

As for related works, it was observed that authors Okubo et al. (2017) used a method to predict students' final grades from the stored record data while in STI MAZK students had access to their performance, skills, and challenges based on the data stored in the system. Zakzewska (2018) assumes that student groups have already been created and that they consist of learners of similar characteristics, differentiating from STI MAZK where the creation of groups is random. Peñafiel et al. (2018) used sentiment analysis that was not used in this study. Islam, Soddiqui, and Aljohani (2019) used similar data mining techniques such as clustering and association to identify relevant student patterns.

As a future work, analyze the data of the pedagogical module to identify the pedagogic strategies used in the STI and correlate with the learning outcome.

REFERENCES

Arcos, j. L.; muller, w.; fuente, o.; orúe, l.; arroyo, e.; leaznibarrutia, i.; santander, j.l. (2002) “integration of virtual environments and it’s for sugery training.” In: conference international intelligent tutoring system, 5. 2000, montreal, ca, berlin, de: springer, p. 43-52.

Barrios-tao, h.; díaz, v.; guerra, y. (2019) “artificial intelligence and education, challenges and disadvantages for the teacher”. Arctic journal, v. 72, n. 12, p. 30- 50.

Bittencourt, w. N. (2018) “a utilização do tutor inteligente mazk no processo de ensino-aprendizagem”. Dissertação (mestrado) - ufsc, ppgtic, araranguá-sc.

Guo, c., yan, x., and li, yan. (2020). “Prediction of student attitude towards blended learning based on sentiment analysis”. In proceedings of the 2020 9th international conference on educational and information technology (iceit 2020). Association for computing machinery, new york, ny, usa, 228–233. Doi: 10.1145/3383923.3383930

Islam, m. S. And aljohani. N.r.(2019). “Identifying online profiles of distance learning students using data mining techniques”. In proceedings of the 2019.

Jovanović, j. Dawson, s. Joksimović, s. And siemens, g. (2020). “Supporting actionable intelligence: reframing the analysis of observed study strategies”. In proceedings of the tenth international conference on learning analytics & knowledge (lak ’20). Association for computing machinery, new york, ny, usa, 161–170. Doi: 10.1145/3375462.3375474

Moro, f. F. Et al, (2019) “the use of mazk intelligent tutor in the process of teaching and learning geography applied in elementary education”. In: workshop on advanced virtual environments and education, out. Florianópolis-sc.

Murray, t (1999). “Authoring intelligent tutoring systems: an analysis of the state of the art. In: international journal of artificial intelligence in education”, v.10, p. 98-129. Disponível https://telearn.archives-ouvertes.fr/hal-00197339. Acesso em: 18 jun. 2021.

Okubo, f. Yamashi, t. Shimada, a. And ogata. H. (2017). “a neural network approach for students’ performance prediction”. In proceedings of the seventh international learning analytics & knowledge conference (lak ’17). Association for computing machinery, new york, ny, usa, 598–599. Doi: 10.1145/3027385.3029479

Pedró, f. (2019). “Artificial intelligence in education: challenges and opportunities for sustainable development”. Unesco. 2019. Disponível em: https://unesdoc.unesco.org/ark:/48223/pf0000366994 acessado em 02 de julho de 2021.

Pozzebon, e. (2008) “um modelo para suporte ao aprendizado em grupo em sistemas tutores inteligentes”. 2008. Tese (doutorado em engenharia elétrica), programa de pós-graduação em engenharia elétrica, centro tecnológico, ufsc, florianópolis, br.

Peñafiel,m.vásquez, s. Vásquez, d., zaldumbide, j. And luján-mora, s. (2018). “Data mining and opinion mining: a tool in educational context”. In proceedings of the 2018 international conference on mathematics and statistics (icoms 2018). Association for computing machinery, new york, ny, usa, 74–78.

Silva, v. I. (2019) “um modelo para a utilização da metodologia ativa aprendizagem baseada em casos no sistema tutor inteligente mazk”. 2019. Dissertação (mestrado em tecnologias da informação e comunicação), ppgtic, cts, ufsc.

Vicari, r.m. (2018) “tendências em inteligência artificial na educação no período de 2017 a 2030”. 1. Ed. Brasília: diretoria de tecnologia e educação. Serviço nacional de aprendizagem industrial, serviço social da indústria. Https://www2.fiescnet.com.br/web/uploads/recursos/d1dbf03635c1ad8ad3607190f17c9a19.pdf acessado em 02/07/2021.

Uzir, n. Ahmad, d. Gašević, j. Jovanović, w. Matcha, l. L. And fudge a. (2020). “Analytics of time management and learning strategies for effective online learning in blended environments”. In proceedings of the tenth international conference on learning analytics & knowledge (lak ’20). Association for computing machinery, new york, ny, usa, 392–401. Doi: 10.1145/3375462.3375493