Educational Data Mining: Analysis Based On an Intelligent Tutoring System for Teaching Algebra

Matheus Freitas de Menezes
Institute of Computing (IComp)
Federal University of Amazonas
Manaus, AM, Brazil
matheus.menezes@icomp.ufam.edu.br

José Francisco de M. Netto
Institute of Computing (IComp)
Federal University of Amazonas
Manaus, AM, Brazil
jnetto@icomp.ufam.edu.br

Arcanjo Miguel Mota Lopes
Institute of Computing (IComp)
Federal University of Amazonas
Manaus, AM, Brazil
amml@icomp.ufam.edu.br

***Abstract*—This complete article of the research category presentes proposes the application of Educational Data Mining (EDM) techniques, based on data generated by a web-based Intelligent Tutor System (ITS), developed for the teaching of algebra, which helps students to develop fundamental knowledge of mathematics, in addition to carrying out a comparative study between the level of difficulty of the proposed algebraic questions. For this, we use the knowledge discovery methodology to perform the following steps from the data: cleaning, integration, selection, transformation, mining, evaluation and presentation of information. The practical results reveal that the proposed architecture can classify the academic performance of students in each evaluation period with an accuracy of around 80%. It was also possible to identify factors related to the resolution of questions, such as the rate of correct answers and the mathematical steps to solve an exercise, classifying the level of difficulty of the questions proposed by the system, acting on the student's deficiencies, and the system. The results obtained from the data analysis can serve as a basis for decision-making, helping in the teaching process for teachers, tutors and managers to monitor academic performance, enabling the correction of problems and identifying the individual and collective difficulties present in each class. In summary, these results also provide a general roadmap on the performance of using EDM techniques in a given context.**

***Index Terms*—Educational Data Mining, Intelligent Tutoring System.**

I. Introduction

With the increasing use of technologies to support the teaching and learning process in the educational field, a large volume of data has been generated based on the different types of interaction with the system, covering mainly students, tutors, and teachers. However, much of this data has not been analyzed, establishing a significant gap, given the amount of information that can potentially be extracted from such data [1].

Consequently, data mining techniques are gaining more and more important in the education sector. Educational Data Mining (EDM) has attracted significant interest in recent years in an attempt to personalize and improve students' learning process [2]. Some other areas, teaching is discovering the potential impact of these techniques on the learning process. Certainly, these techniques can provide information to support the development of educational models that increase the efficiency and quality of [3] teaching and learning.

According to Organization for Economic Co-operation and Development (OECD), through the International Student Assessment Program (PISA), 68.1% of 15-year-old Brazilians students have a lower mathematics learning degree than that classified as basic. [4].

Taking into account the current moment in public high schools, it becomes increasingly complex for the teacher to detect individual or collective difficulties, given the abundance of variables. Often, this recognition is only noticed through the results of the final grades that are part of the evaluation method that makes up the [5] educational process.

The environment of an educational institution involves three types of actors: teacher, student, and the environment. The interaction between these three actors generates voluminous data that can be systematically grouped to extract valuable information. These data analyses allow the prediction of student performance, associate learning styles of different profiles and their behaviors, and collectively improve institutional performance [6].

Given this scenario and considering the amount of data generated by the support systems for the teaching and learning process of students [7], the project proposes to apply mining techniques in data generated by a web-based Intelligent Tutor System (ITS) called LEIA (LEarnIng Algebra), which helps students develop foundational math skills [8].

The rest of the paper is organized as follows. Section 2 presents and discusses related work. Section 3 presents the research method used while Section 4 shows the results and discussions. Section 5 presents conclusions and future work.

II. Theoretical Foundation

This Section exposes the fundamental concepts for understanding this work. The concepts of Educational Data Mining and the main works related to the research are presented.

A. Educational Data Mining

EDM is a subarea of Data Mining, that is, this line is composed of models, tasks, methods and algorithms that support the exploration of data from educational environments, to discover patterns and predictions that characterize the behaviors, practices, contents of domain, assessments and educational activities of students [9]. EDM is concerned with

developing methods to explore the unique types of data from educational environments, using methods to better understand students and the settings in which they learn [10] [11].

Fig. 1. Fields related to the EDM process.

The EDM method uses related fields such as machine learning, text mining (approaches to find patterns in text in natural language) and statistics, as we can see in Figure 1. Other important influences are psychometrics (the study of psychological instruments to measure human abilities and characteristics) and web log analysis (approaches to identify user profiles and browsing patterns of website users). EDM can contribute to the assessment of learning systems, identify regular or unusual patterns, student problem-solving strategies and patterns of successful or unsuccessful collaboration, thus helping to formulate new scientific hypotheses [12].

B. Related Works

[13] applied EDM techniques to identify the learning behaviors of students in a programming class, using machine learning methods, such as classification and grouping, based on the use of Python language for data analysis, also integrating the following libraries: Panda , Numpy, Matplotlib, Seaborn, SciPy, Scikit-learn. The work verified the efficiencies of the Random Forest model, being ahead of the logistic regression and decision tree models, also managing to identify factors such as the pace of learning with synchronous and asynchronous activities. Although this study adds new insights into the application of EDM to identify student learning patterns, the questionnaires were based on self-report, which can lead to inaccuracy in responses.

[14] developed a predictive analysis to estimate the extent to which it is possible to predict the final grade point average of students in an engineering class. The work used data mining models based on Konstanz Information Miner (KNIME), which uses 6 algorithms: Probabilistic Neural Network (PNN), Random Forest, Decision Tree, Naïve Bayes, Tree Ensemble, and Logistic Regression.

To carry out the analysis, the authors used as a basis the data generated in the classes of the Covenant University in Nigeria, in seven engineering departments, namely: Information and Communication Engineering, Chemical Engineering, Computer Engineering, Mechanical Engineering, Electrical, and Electronic Engineering , Civil and Petroleum Engineering.

The results showed that among the six data mining algorithms used, a maximum precision of 89.15% was achieved. The result was verified using linear and quadratic regression models. This creates an opportunity to identify students who may graduate with poor results or may not even graduate, so early intervention can be critical [14]. Unlike this work, we seek to analyze the data generated from interactions with the ITS LEIA, addressing other variables of interest.

[15]. conduct an EDM process based on data generated from four institutions in the Indian education system, to predict student performance. The authors implemented the Apriori algorithm to extract patterns along with their associations against various sets of records. It also used the K-means model to efficiently group students into specific categories. All algorithms were implemented in the WEKA (Waikato Environment for Knowledge Analysis) tool.

The work concludes that the overall performance of students is predicted to be satisfactory after analyzing 100 students using 45 parameters. In general, the results obtained in the case study indicate that data mining using the WEKA tool is satisfactory. As a result of the analysis, students can be classified based on their academic data, family history and learning methodolog. However, the authors use a large number of different variables, differing from this work that uses a reduced number of attributes, to answer specific questions, having to apply EDM techniques that benefit the independence between the variables.

[16] present a web-based tool to support the teaching of engineering students, without the students necessarily having data science experience. The system is focused on analyzing student performance. The author's used classification models (Support Vector Machine (SVM)), clustering, and logistic regression.

The tool's usefulness was classified as satisfactory to the stated objectives, showing its potential for supporting students. Real data from higher engineering education institutions show

the tool's potential for higher education management and the validation of its application in real scenarios.

[17] presented the conduction of a high-influence variable selection process for predicting student performance in an Industrial Engineering department at the Islam Indonesia University, the authors implemented a classification analysis using Bayesian Network and Decision Tree with the aid of the WEKA tool, The work showed that the Bayesian Network model is superior to the decision tree since it presented a higher accuracy rate. It also presented the scenario where the average grade attribute proved to be the most influential in all methods. However, some attributes such as gender and origin can be biased without taking into account social aspects.

As previously highlighted, unlike the works reported, we developed analyses focused on the ITS, using only data referring to user interactions with the system, not considering social aspects related to the educational process, seeking to measure student performance and the level of the proposed questions.

III. Method

The objective of this work is to analyze the data generated by the ITS LEIA for the teaching of algebra, applying data mining techniques, to extract information that contributes to the evolution of the educational performance of the system.

To achieve the objectives, the solution proposed in this work developed, of a nature applied to teaching and learning, focusing on the ITS LEIA, the model for carrying out this work was inspired by the knowledge discovery approach, in order to extract, visualize and understand the generated data. [18].

the architecture proposed for the development of the knowledge discovery process is presented in Figure 2, detailing the sequential process of the following steps:

- Data cleaner to remove noise and inconsistent data.
- Data integration where multiple data sources can be combined.
- Data selection where data relevant to the analysis task is retrieved from the database.
- Data transformation where data is transformed and consolidated into forms appropriate for mining, performing summary or aggregation operations.
- Data mining an essential process where intelligent methods are applied to extract patterns from data.
- Pattern assessment to identify patterns that represent knowledge based on measures of interest.
- Knowledge presentation where visualization and knowledge representation techniques are used to present information.

The data applied in the knowledge discovery process were stored in the ITS LEIA database, these data were obtained through an experiment, where 20 students from a public school used the system, trying to solve a set of questions proposed by the ITS. From this, the process steps can be applied, starting with data cleaning where irrelevant variables were selected and excluded based on the information to be extracted, for example, at this moment the ITS incorrectly counted the resolution time of each question, then the variable that represents the time was excluded from the mining, the other relevant data were integrated into the process.

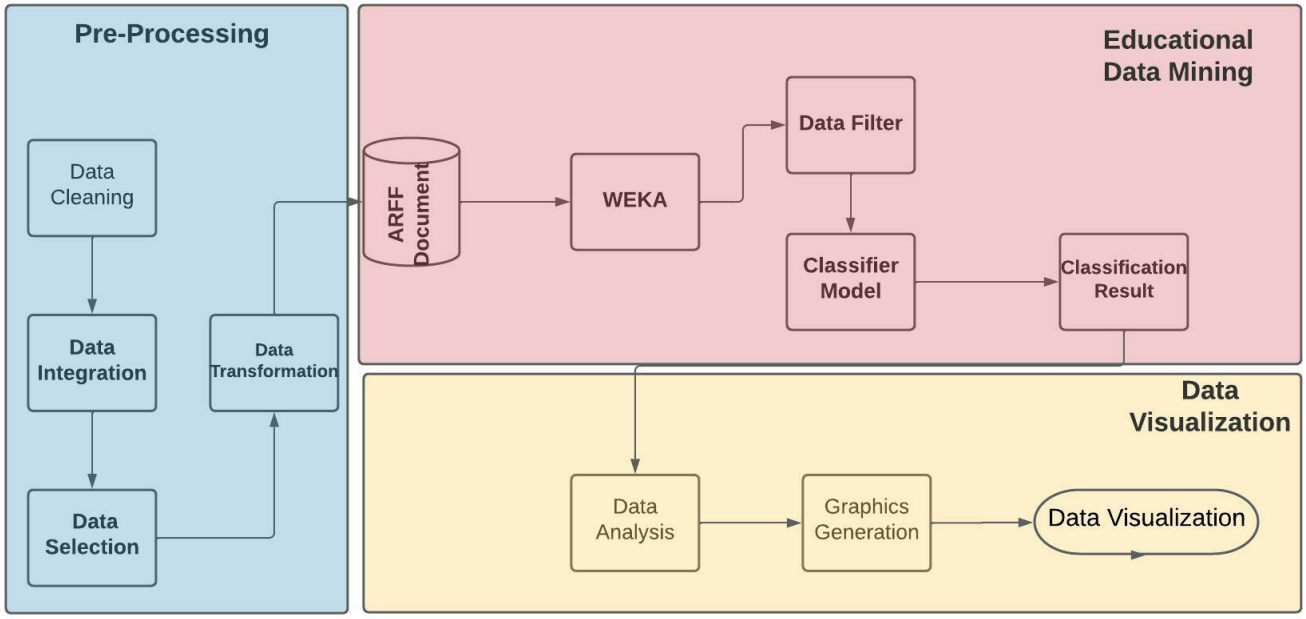

Fig. 2. Research knowledge discovery process.

After the execution of the cleaning and integration process, the relevant data were selected and transformed to be properly applied in the data mining process, transporting to the type ARFF (Attribute Relation File Format), format used by WEKA, being the tool used in this work to carry out the EDM process, the WEKA package has a huge variety of algorithms from different approaches, acting after statistical analysis, seeking to identify patterns and generate hypotheses [19]. In Figure 3 we can see one of the files used in EDM.

```
@relation sti

@attribute student {s1, s2, s3, s4, s5, s6, s7, s8

@attribute question  {q1,q2,q3,q4,q5,q6,q7,q8,q9,q

@attribute acuracy {0,1}

@attribute level {easy, normal, hard}



@data
s1  q1  1   easy
s1  q2  1   easy
s1  q3  1   easy
s1  q6  0   easy
s1  q4  1   easy
s1  q15 0   easy
s1  q5  1   easy
s1  q6  0   easy
s1  q6  0   easy
s1  q6  0   easy
s1  q6  0   easy
s2  q4  1   easy
s3  q1  1   easy
s5  q1  1   easy
s4  q1  1   easy
s6  q1  1   easy
s7  q1  1   easy
```

Fig. 3. Data in ARFF format.

After the transformation step, the data were entered in the WEKA tool, to analyze the best mining algorithm for this context and the most appropriate filter, to obtain the best possible use.

After performing the data mining step, the extracted information was analyzed and organized to identify patterns looking for evidence that is related to the students' learning process to improve the ITS LEIA.

IV. Results

Based on the data set with the attributes described, some experiments were carried out, to evaluate the performance of the students, measuring the level of difficulty of the questions and comparing with the classification adopted by the system, in addition to understanding the path that the student goes through until he hits or misses an activity. Below we can understand the variables used in the EDM process.

- Question: Represents the algebraic questions available in ITS LEIA.
- Student: Represents the students who participated in the experiment, using the ITS LEIA.
- Accuracy: Represents the errors and successes in each activity.
- Level: Represents the difficulty level of each question.

The metric used to evaluate the results of the experiments is the Accuracy. Defined as the ratio between the number of students and questions correctly classified by the algorithms in their respective class and the total number of students considered in the study.

A. Rate of correct steps for each question

In experiment 1, we sought to identify the rate of correct steps to get each question right, serving as an example, to solve question 1, student 1 needed two correct steps to reach the conclusion, that is, 100% of correct steps in the question 1.

We use the Naïve Bayes technique [20], being one of the most popular data mining algorithms, its efficiency comes from the possibility of independence between the variables, even though this can be violated in many datasets [21] . Through the algorithm we obtained 80.4665% of correctly classified instances and 19.5335% of incorrectly classified instances.

We also apply the resample filter, in each experiment, in order to try to increase the quality and performance of the classifier of each experiment, since it acts in the balancing of the classes, not allowing the classification to be biased, considering only a few attributes. In Table I we can identify the correct rate for each question.

Questions q14, q19 and q20 were disregarded in the mining process, since few students were able to answer such questions, due to the high level of difficulty. Taking into account the data in the table, we noticed that only questions q4 and q11 received a small hit rate.

Below 35%, demonstrating a clear difficulty of the students in the development of the mathematical steps.

TABLE I
RATE OF CORRECT STEPS FOR EACH QUESTION

Question	Hit rate
q1	81.15%
q2	77.14%
q3	92.95%
q4	31.03%
q5	46.53%
q6	82.35%
q7	58.57%
q8	88.53%
q9	88.37%
q10	84.93%
q11	18.18%
q12	47.36%
q13	50%
q15	70.37%
q16	66.66%
q17	53.84%
q18	68.57%

Another 6 questions obtained a rate of correct steps above 80%, namely: q1, q3, q6, q8, q9 and q10. Indicating a

greater proficiency of the students on the questions. The other questions were between 35% and 80%, indicating an average rate. In Figure 4 we can observe the distribution of errors and correct answers among the questions, where 1 represents the correct answer and 0 represents the error.

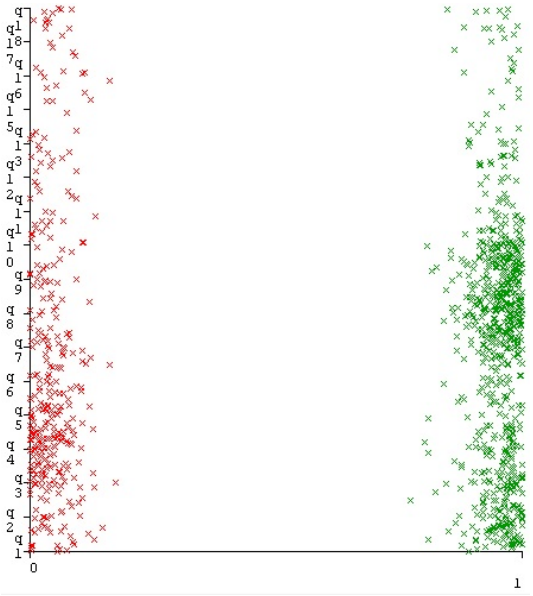

Fig. 4. Graphical representation of errors and successes between steps

Table II displays the accuracy rate of mathematical passes for each student participating in the experiment, demonstrating that half of the students managed to reach a rate of success greater than 70%, in addition, only 2 students reached less than 50 % hits.

TABLE II
STEP SUCCESS RATE PER STUDENT

Students	Step success rate
student1	91.66%
student2	89.47%
student3	88.23%
student4	94.73%
student5	39.74%
student6	87.14%
student7	89.58%
student8	34.61%
student9	44.30%
student10	58.73%
student11	62.00%
student12	76.08%
student13	58.62%
student14	85.71%
student15	42.85%
student16	72.50%
student17	64.38%
student18	58.06%
student19	59.64%
student20	78.43%

B. Setting the level of questions

To define the level of the questions, we used the ease index [22], developed by the National Institute of Educational Studies and Research Anísio Teixeira (INEP), to be applied during the National Student Performance Exam. This index classifies the questions in five levels of difficulty through the correct rate, being: very easy, easy, medium, difficult and very difficult [23]. Table III displays the rate corresponding to each level.

TABLE III
INEP CLASSIFICATION MODEL

Hit Rate	Classification for ENADE
>0,86	Very easy
0,61 - 0,85	Easy
0,41 - 0,60	Normal
0,16 - 0,40	Hard
<0,15	Very hard

One of the research objectives is to compare the difficulty level assigned to each question with its real level, based on the ease index, since clearly the students felt difficulties in some questions classified as easy. The ITS LEIA distributes the questions among three categories: easy, normal and hard, based on the size of the equations, the more elements the equation had, the greater its level of difficulty.

To perform the comparison, we used the Naive Bayes classification algorithm again, with the purpose of obtaining the correct rate of each question, then categorizing within the standards of the ease index and performing the comparison between the methods. Through the algorithm we obtained 78.6164% of correctly classified instances and 21.3836% of incorrectly classified instances. In Figure 5 we can see a graph that details the distribution of questions among the levels of difficulty proposed by ITS LEIA.

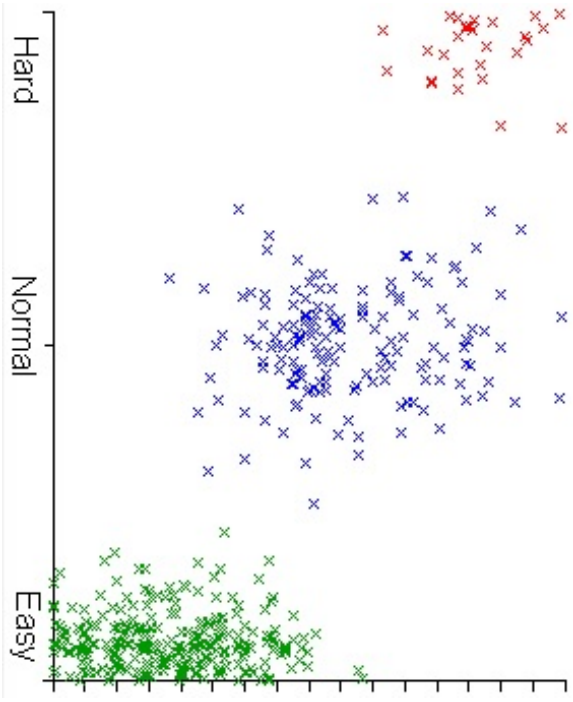

Fig. 5. Distribution of questions among the levels proposed by the ITS

After classifying the data, we obtained the hit rates for each question, with this we were able to make a comparison between the classifiers of difficulty of questions. In Table IV we can observe the rate of correct answers and the categorization of the questions.

TABLE IV
COMPARATIVE TABLE BETWEEN THE CLASSIFICATIONS

Question	Hit Rate	Classification - INEP	Classification - ITS
q1	0.97	Very Easy	Easy
q2	0.69	Easy	Easy
q3	0.63	Easy	Easy
q4	0.48	Normal	Easy
q5	0.27	Hard	Easy
q6	0.36	Hard	Easy
q7	0.65	Easy	Normal
q8	0.63	Easy	Normal
q9	0.40	Hard	Normal
q10	0.48	Normal	Normal
q11	0.05	Very Hard	Normal
q12	0.25	Hard	Normal
q13	0.30	Hard	Hard
q15	0.57	Normal	Easy
q16	0.55	Normal	Hard
q17	0.71	Easy	Hard
q18	0.28	Hard	Hard

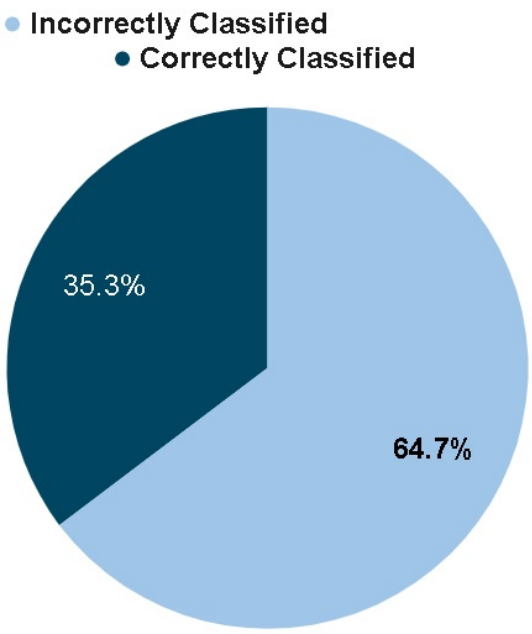

Fig. 6. STI method hit rate

Taking into account the hit rate of the steps taken to solve a question, as we saw in Session 4.1, we can see some substantial differences between the types of hits, exemplifying through q6 where the hit rate of steps obtained was 0.82 and the correct rate of questions was 0.36, that is, even with a high number of steps, few students managed to finish the question, this difference is also justified by the dropout at the time of resolution, when the student passes the question, the ITS counts as a abandonment, classifying as an error.

The other example is question q17, which in relation to the success rate of steps reached an index of 0.53, but regarding the success rate of the question itself, it was more successful, reaching 0.71, this is justified due to the fact that students manage to solve the problem even if they miss some steps, probably due to the help of the ITS, at the moment the system does not keep the data related to the use of the help resource, being impossible to detect if the help was useful or not, this functionality will be implemented in the future.

As for the comparison between the classification models, some considerations must be made, such as the adequacy of the categories so that they are equal in number, since the facility index method has 5 sets and the ITS classification contains only 3, so the Very Easy and Easy categories, corresponded to ITS Easy. Consequently, the Very Hard and Hard types match the ITS Hard, the Normal category remains the same. In Figure 6, we can observe the comparison between correct and incorrect classifications.

After adjusting the nomenclatures, it is noted that most of the questions do not have the same level of difficulty, 11 questions did not fit the classification established by the ITS, namely: q4, q5, q6, q7, q8, q9, q11, q12, q15, q16, and q17. Among these, the questions q5 and q6, categorized as Easy by the ITS, stand out, but based on the INEP classification, they are classified as Hard, exposing a discrepancy in the labeling, as well as in q11, considered a Normal question in the ITS, however it has only 0.05 in the hit coefficient, being evaluated as Very Hard.

V. CONCLUSION

Assimilating the particularities of each student in detail has proved to be a great challenge for researchers and professionals in the educational field. In view of this, efforts have been made by the scientific community in the development of technological solutions that provide relevant information to help the management of the [24] teaching process.

In order to understand some educational aspects of interactions between student and system. In this work we propose the application of EDM techniques, based on data generated by the web-based ITS LEIA, developed for the teaching of algebra, which help students to develop fundamental knowledge of mathematics, seeking to identify patterns that contribute to the educational process.

For this, we used EDM techniques employing classification algorithms, reaching more than 80% of accuracy, even without a large volume of data, since the experiment had only 20 participants, but it was enough to generate interesting results, being possible to classify and compare the difficulty level of the questions used in the ITS.

It was also possible to verify some important information, such as the relationship between the number of correct steps and the coefficient of correct answers for questions, we can see that some questions with a high rate of correct steps do not present a high rate of correct answers for the question,

indicating that in some exercises the students managed to get most of the steps right but could not reach the final result.

The results obtained from the analysis carried out indicate that the method used in the EDM process can be used in future research, even if these works have a more significant amount of data.

The research had some limitations, such as the system's inefficiency in storing some important attributes, such as counting the number of tips used by students, so it was not possible to identify whether the tips given by the ITS were useful or not. Therefore, new possibilities for analysis must arise as the system updates the proposed needs.

In future works, we will seek to expand the amount of data to be analyzed, based on new experiments and updates in the way the system stores data and the implementation of new functionalities, allowing the extraction of a wide variety of information. We also intend to carry out a comparative study between EDM techniques, in order to report the advantages and disadvantages related to the use of each technique, identifying the best performance.

ACKNOWLEDGEMENT

We thank for Coordination for the Improvement of Higher Education Personnel (CAPES).

REFERENCES

[1] H. Rabelo, A. Burlamaqui, R. Valentim, D. Rabelo, and S. Medeiros, "Utilização de técnicas de mineração de dados educacionais para predição de desempenho de alunos de EaD em ambientes virtuais de aprendizagem" in *Brazilian Symp. on Comput. in Educ.*,Vol. 28, No. 1, pp. 1527-1537, Nov. 2017.

[2] M. Injadat, A. Moubayed, A. B. Nassif, and A. Shami, "Systematic ensemble model selection approach for educational data mining," *Knowledge-Based Systems*, vol. 200, p. 105992, Jul.2020.

[3] H. Aldowah, H. Al-Samarraie, and W. M. Fauzy, "Educational data mining and learning analytics for 21st century higher education: A review and synthesis," *Telematics and Informatics*, vol. 37, pp. 13–49, Apr. 2019.

[4] INEP; "Relatório Brasil no PISA 2018" in *INEP*, pp .113-116, 2020

[5] EV. Silva, JF. Netto, and RA. Souza, "VLA Dashboard: Um Mecanismo para Visualização do Desempenho de Estudantes de Matemática no Ensino Médio," *RENOTE*, vol. 16, no. 2, Dec. 2018.

[6] A. Dutt, M. A. Ismail, and T. Herawan, "A Systematic Review on Educational Data Mining," *IEEE Access*, vol. 5, pp. 15991–16005, 2017.

[7] E. Costa, J. Aguiar, and J. Magalhães, "Sistemas de Recomendação de Recursos Educacionais: conceitos, técnicas e aplicações," in *II Jornada de Atualização em Informática na Educação*, V. 1, N. 1, pp. 1-29, 2013.

[8] A. Lopes, J. Netto, R. Souza, A. Mourao, T. Almeida e D. Lima, "Improving Students Skills to Solve Elementary Equations in K-12 Programs Using an Intelligent Tutoring System", in *IEEE Frontiers in Education Conf. (FIE)*, 2019.

[9] A. Peña-Ayala, "Educational data mining: A survey and a data mining-based analysis of recent works," *Expert Systems with Applications,* vol. 41, no. 4, pp. 1432–1462, Mar. 2014.

[10] R. S. Baker and P. S. Inventado, "Educational Data Mining and Learning Analytics," in *Learning Analytics,* Springer New York, pp. 61–75, 2014.

[11] S. George, RS. Baker and SJ Ryan d, "Learning analytics and educational data mining: towards communication and collaboration," in *Proc. 2nd. Int. Conf. on learning analytics and knowledge*, pp. 252-254, 2012.

[12] O. Scheuer and B. M. McLaren, "Educational Data Mining," in *Encyclopedia of the Sciences of Learning,* Springer US, pp. 1075–1079, 2012.

[13] H.-C. Hung, I.-F. Liu, C.-T. Liang, and Y.-S. Su, "Applying Educational Data Mining to Explore Students' Learning Patterns in the Flipped Learning Approach for Coding Education," *Symmetry*, vol. 12, no. 2, p. 213, Feb. 2020.

[14] A. I. Adekitan and O. Salau, "The impact of engineering students' performance in the first three years on their graduation result using educational data mining," *Heliyon*, vol. 5, no. 2, Feb. 2019

[15] G. S. Gowri, R. Thulasiram, and M. A. Baburao, "Educational Data Mining Application for Estimating Students Performance in Weka Environment," *IOP Conf. Series: Mater. Sci. and Eng.*, vol. 263, Nov. 2017.

[16] M. A. Prada et al., "Educational Data Mining for Tutoring Support in Higher Education: A Web-Based Tool Case Study in Engineering Degrees," *IEEE Access*, vol. 8, pp. 212818–212836, 2020.

[17] A. U. Khasanah and Harwati, "A Comparative Study to Predict Student's Performance Using Educational Data Mining Techniques," *IOP Conf. Series: Mater. Sci. and Eng.*, vol. 215, p. 012036, Jun. 2017.

[18] J. Pei, J. Han, and H. Tong, "Data Mining: Concepts and Techniques," *Elsevier Science Technology Books*, 2021.

[19] K. Alpan and G. S. Ilgi, "Classification of Diabetes Dataset with Data Mining Techniques by Using WEKA Approach," in *2020 4th Int. Symp. on Mult. Studies and Innovative Tec. (ISMSIT)*, Istanbul, Turkey, Oct. 2020. IEEE.

[20] GH. John and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers," in *11th Conf. on Uncertainty in Artif. Intell*, pp 338-345, 1995.

[21] S. Chen, G. I. Webb, L. Liu, and X. Ma, "A novel selective naïve Bayes algorithm," *Knowledge-Based Systems*, vol. 192, p. 105361, Mar. 2020.

[22] INEP, "Sistema Nacional de Avaliação da Educação Superior: Relatório Síntese de Área Ciência da Computação (Bacharelado/Licenciatura)". *Federal Government of Brazil*, 2017.

[23] M. A. P. Lima, L. S. G. Carvalho, E. H. T. d. Oliveira, D. B. F. d. Oliveira, and F. D. Pereira, "Uso de atributos de código para classificar a dificuldade de questões de programação em juízes online," *Revista Brasileira de Informática na Educação*, vol. 29, pp. 1137–1157, Sep. 2021

[24] E. Gottardo, C. A. A. Kaestner, and R. V. Noronha, "Estimativa de Desempenho Acadêmico de Estudantes: Análise da Aplicação de Técnicas de Mineração de Dados em Cursos a Distância," *Revista Brasileira de Informática na Educação*, vol. 22, no. 01, p. 45, May. 2014.