Redin
REVISTA EDUCACIONAL INTERDISCIPLINAR
FACCAT
Núcleo de Educação On-line

Educational data mining for the profile of candidates from the Encceja in the State of Rio Grande do Sul

Mineração de dados educacionais para o perfil de candidatos da prova do Encceja no Estado do Rio Grande do Sul

Jacqueline Mayumi Akazaki[1]
Francielli Freitas Moro[2]
Liane Margarida Rockenback Tarouco[3]
Patrícia Alejandra Behar[4]
Letícia Rocha Machado[5]
Sílvio César Cazella[6]

Abstract

This article aims to identify the socioeconomic profile of candidates from the National Examination for the Certification of Youth and Adult Skills (Encceja in Portuguese) in the state of Rio Grande do Sul from an educational data mining approach. Encceja assesses students' competence in Youth and Adult Education (YAE), in order to verify and recognize the learning obtained. The database used in the process was INEP in 2017, since it is the last version released. To investigate these profiles, the clustering method was applied and the k-Means algorithm was used. As a result, it was discovered that the absence of a school close to the candidate's residence, as well as the absence of a place in a public school, and social and economic issues are not linked to school dropout. The contributions of this research are related to the reason why students interrupt their school trajectories, which is linked to the service, that is, the candidates reported that the Encceja test is the best way to reconcile studies with work.

Keywords: National Youth and Adult Skills Certification Exam; Education Data Mining; Profiling.

[1] Doutoranda em Informática na Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: jacquelineakazaki@gmail.com.
[2] Doutoranda em Informática na Universidade Federal do Rio Grande do Sul (UFRGS). E-mail: franf_m@hotmail.com.
[3] Doutorada em Engenharia Elétrica e Sistema Digitais pela Universidade de São Paulo. Professora-titular do Programa de Pós-Graduação em Informática na Educação da Universidade Federal do Rio Grande do Sul (PPGIE/UFRGS). E-mail: liane@penta.ufrgs.br.
[4] Doutora em Ciências da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS). Professora-titular da Faculdade de Educação da Universidade Federal do Rio Grande do Sul (UFRGS), do PPG em Educação (PPGEdu/UFRGS) e do PPG em Informática na Educação (PPGIE/UFRGS). Possui bolsa de Produtividade em Pesquisa (DT/Cnpq), nível I. Coordena o Núcleo de Tecnologia Digital aplicada à Educação (NUTED/Cnpq) da Faculdade de Educação (FACED), vinculado ao Centro Interdisciplinar de Novas Tecnologias na Educação (CINTED). E-mail: pbehar@terra.com.br.
[5] Pós-Doutora pelo PPG em Informática na Educação da Universidade Federal do Rio Grande do Sul (PPGIE/UFRGS). Doutora em Informática na Educação e, também, em Educação pela UFRGS. Professora-visitante no PPG em Tecnologias da Informação e Comunicação da Universidade Federal de Santa Catarina (UFSC). E-mail: leticiarmachado@gmail.com.
[6] Doutor em Ciência da Computação pela Universidade Federal do Rio Grande do Sul (UFRGS) com período de estudos na Universidade de Alberta (Canadá). Professor-associado Nível II na Universidade Federal de Ciências da Saúde de Porto Alegre (UFCSPA), onde atua junto aos PPG em Ensino na Saúde e PPG em Tecnologias da Informação e Gestão em Saúde. E-mail: silvio.cazella@gmail.com.

Resumo

O presente artigo tem por objetivo identificar o perfil socioeconômico dos candidatos do Exame Nacional para Certificação de Competências de Jovens e Adultos (Encceja) no estado do Rio Grande do Sul a partir de uma abordagem de Mineração de Dados Educacionais. O Encceja avalia a competência de alunos na Educação de Jovens e Adultos (EJA), a fim de verificar e reconhecer a aprendizagem obtida. A base de dados utilizada no processo foi o INEP no ano de 2017, visto que é a última versão liberada. Para investigar esses perfis, aplicou-se o método de *clusterização* e utilizou-se o algoritmo k-Means. Como resultados descobriu-se que a ausência de escola próximo da residência do candidato, bem como a inexistência de vaga em escola pública e, questões sociais e econômicas não estão ligadas ao abandono escolar. As contribuições desta pesquisa, estão relacionadas com a razão pela qual os alunos interrompem suas trajetórias escolares, que está ligado ao serviço, ou seja, os candidatos relataram que a prova do Encceja é a melhor maneira para conciliar os estudos com o trabalho.

Palavras-chave: Exame Nacional para Certificação de Competências de Jovens e Adultos; Mineração de Dados Educacionais; Extração de Perfis

1. Introduction

The National Institute for Educational Studies and Research “Anísio Teixeira” (INEP in Portuguese) is responsible for survey and propagation of information about education in Brazil at all stages of teaching, through evaluations, exams and indicators. According to the Law of Directives and Bases of National Education (LDB in portuguese) Nº. 9394, of December 20, 1996, INEP created in 2002 the National Exam for Certification of Competences of Youngsters and Adults (Encceja in portuguese). Encceja is focused on verification and recognition of knowledge of Brazilian youth and adults, at levels corresponding to the completion of Elementary and Secondary Education, residents in Brazil and abroad. The exam measures the competences of students in Youth and Adult Education (YAE), building a National reference (REBOUÇAS et al., 2019).

In accordance with the LDB the minimum age to plead the certification with corresponds to the conclusion of elementary teaching, reached by the date of the tests, is 15 years old in the Elementary School and 18 years old in the High School. However, over the years 2003, 2004, 2009 and 2012, the organization of the Exam was marked by interruptions, in addition to the postponement of the application process in 2010. In this Year, important changes occurred in Encceja: the subscribe of participants began to be carried out exclusively via the internet, the socioeconomic questionnaire had a

reduced number of questions and adherence became possible only for the State Education Departments, in which 22 were inserted in the process (INEP, 2017).

With the application of Encceja, the brazilian Ministry of Education (MEC) and INEP expect to offer an opportunity to validate knowledge to young and adults living in Brazil and abroad, that for different reasons, discontinued their school trajectories, besides favoring among the participants the resumption of studies, as well as its relocation or better positioning in the labor market. From the data of Encceja 2017, managers and researchers on this theme were able to evaluate the numerical dimensions, the profiles and the trajectories of candidates, in addition to qualifying the debate on the role of Exame in the set of YAE policies developed at different levels of the Federation (INEP, 2017).

In this context, due to the large amount of Encceja data generated, it is possible to apply Educational Data Mining (EDM) techniques to identify the socioeconomic profiles of the candidates. EDM consists to the analysis of a large set of educational data, in the search for standards that can assist decision making by institutional managers and specialists (HERPICH et al., 2016 and DA SILVA PINTO, 2019). The process of EDM can be seen as an iterative cycle of formation, test and refinement of hypotheses, going through several stages, which are: data selection, pre-processing, transformation, interpretation and evaluation (ROMERO and VENTURA, 2013).

Therefore, this article presents the results of the application of EDM to extract socioeconomic profiles of the candidates who took the Encceja test in the state of Rio Grande do Sul - Brazil. For this, the INEP database for 2017 was used, since it is the last version released, made available free of charge by the Basic Education Evaluation Board. The results found in this mining made it possible to understand the profile of the candidates and, in the future, discuss the application of differential pedagogical strategies for your necessities.

The work is organized into five sections. In the next section, related work is presented. In the third, the materials and methods are described. In the fourth, the results and discussions of this study are detailed. Finally, section 5 presents the conclusions and proposals for future work.

2. Related work

Understanding how Encceja occurs in Brazil and its importance for Education brings possibilities to discuss the adoption of new public policies about Basic Education. It is possible to observe some studies in the area, such as, for example, that of Catelli Junior, Gisi and Serrão (2013), who carried out a survey of documents and existing bibliography since the creation and implementation of Encceja in Brazil from 2002, and analyzed in a quantitative and qualitatively the data. The authors discussed the need to research the reduction in enrollment in municipal and state networks, as well as the aspects of passing rates in the exam. They also stressed the importance of knowing the subject through the identification of the social condition and his interests.

Silva Junior (2015) performed a curriculum analysis, in the state of Rio de Janeiro - Brazil, about the elements related to the teaching of math present in YAE and Encceja. The author emphasizes the matrix competencies and skills used in Exam, which are characterized by language domains, phenomena comprehension, problem solving, ability to argue and make proposals. Martins and Sá (2018) systematized part of the information in the INEP database related to the 2017 edition of the Encceja Prisional in Federal District - Brazil, whith the aim of contributing to public debates on certification in the YAE. The Authors emphasize that the exam does not require previous schooling from the candidates, and it is not mandatory to attend YAE schools. Santos et al., (2019) analyze the math performance of 17.951 Ennceja participants in elementary school, according to the 2014 INEP database and compare the average performances by locations, Brazilian region, and gender. The authors Alves, Comerlato and Sant'anna (2017) developed a research report that maps YAE in the state of Rio Grande do Sul - Brazil in 2017. The aim was to realize a quantitative survey of supply and demand in educational institutions in the private, public, municipal, state and federal networks to assist in the planning of public policy actions.

In this context, a low amount of works developed in the area, and no studies related to the candidate profile of the Encceja exam using explicitly Educational Data Mining based on INEP data in the state of Rio Grande do Sul were found.

3. Materials and methods

The study aimed to identify the socioeconomic profile of the candidates of the National Exam for Certification of Competences of Youngsters and Adults (Encceja in portuguese) in the Brazilian state of Rio Grande do Sul from an Education Data Mining (EDM). This way, this section presents the research context and methodological procedures with their respective stages. The materials refer to the database and the methods are the experimentation of the use of EDM in Encceja context.

3.1 The study context

Encceja was structured in four types of the objective's tests, in addition to an essay and a socioeconomic questionnaire. This data from Encceja has been made available by INEP[7] on its website since 1995. Therefore, the target audience of this study were students from the elementary and high school levels of Youth and Adult Education (YAE) in Rio Grande do Sul - Brazil. The analyzed data correspond to the edition of Encceja 2017. In this year, the tests and the socioeconomic questionnaire were applied in Brazil and abroad. The exam can be implemented, depending on candidate, of three ways:

- **Regular National Encceja** - Exam applied in Brazil to young people without restriction of freedom;
- **Regular Foreign Encceja** - Exam applied abroad to young people without restriction of freedom;
- **National PPL Encceja** - Exam applied in Brazil to people deprived of their liberty (prisoners) and young people under socio-educational measures.

It is worth mentioning that the application of the examination abroad did not have the data disclosed, considering the possibility of identifying the participants, as only 69 people took the Encceja. In relation to regular and PPL candidates, there is a distinction in the dynamics of enrollment and filling in the socioeconomic questionnaire. Regular candidates register and complete the questionnaire online. For the PPL candidates, the registration is made by a pedagogical responsible, and the questionnaire is printed and filled out on the day of the Encceja. However, the candidate is not identified in any way and it is not possible to find out in the database to which candidate the

[7] Available in: <http://inep.gov.br/microdados>.

questionnaire belongs. For this reason, a database with information for PPL candidates was not generated, as was done for regulars (INEP, 2017). Thus, only participants without freedom restrictions in this work were analyzed (Regular National Encceja).

3.2 Methodological procedures

The methodological procedures adopted were composed of four stages, based on García et al. (2011) as described below.

3.2.1 Database description

For this work the base “MICRODADOS_ENCCEJA_NACIONAL_REGULAR” was used because it does not allow the identification of candidates and contains the registration data, responses of tests and the questionnaire of Encceja 2017 for national candidates without restriction of freedom. The data is linked to teaching quality, performance; access and permanence of students, to the economic and social context in which schools are inserted. Thus, the database has the following information:

- **Candidate Registration:** candidate registration number, year of the Encceja, age, gender, country code where the participant resides, name of the country where the candidate resides, code of state/province of residence of the participant, name of the state/province in which the participant resides;
- **Exam Data:** code of state in which the test was applied, name of the state where the test was applied, marking the participant's registration request in each exam, name of the Participating Unit responsible for exam certification;
- **Answers and Feedback of the exams:** situation of the participant in each objective test, answer of the registrant in each test and feedback, averages of the correction criteria for the essay test;
- **Socioeconomic Questionnaire Data:** information from the socioeconomic questionnaire of the participants of the Regular National Encceja.

3.2.2 Preprocessing

In this work, from the selected table (Data from the Socioeconomic Questionnaire) from the Regular National Encceja database, variables were analyzed and, initially, the data selection from the state of Rio Grande do Sul - Brazil was carried

out. These collected data were integrated and filtered in Microsoft Excel 2010[8], gathering in the same table only the records of students who participated in the test. Subsequently, with the help of one specialist from the Encceja area and three from EDM, the most important variables were classified in the table. After this procedure, it was possible to observe a few missing values for some variables, and to solve this problem, we opted for excluding these instances. Finally, the basis for elementary and high school was obtained with a total of 24,627 instances and 8 attributes, in which each instance is equivalent to a candidate who took the Exam and each attribute corresponds to one of the responses to the socioeconomic questionnaire. Table 1 presents the question and their corresponding attributes.

Table 1 – Questions of the socioeconomic questionnaire and its corresponding attributes.

Questions of the Socioeconomic Questionnaire	Attributes
What is the Certification Type?	1 (elementary school) ou 2 (high school)
How old are you?	15 years old, 16 years old, 17 years old...
Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: To get the certificate of completion of Elementary School.	0 or 1 or 2 or 3 or 4 or 5.
Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: To get the certificate of completion of High School.	0 or 1 or 2 or 3 or 4 or 5.
Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: Because it is the best way to reconcile my studies and work.	0 or 1 or 2 or 3 or 4 or 5.
The extent to which the following reasons influenced your failure to attend or drop out of regular school: There are no places in public schools.	0 or 1 or 2 or 3 or 4 or 5.
The extent to which the following reasons influenced your failure to attend or drop out of regular school: Absence of school close to home.	0 or 1 or 2 or 3 or 4 or 5.
The extent to which the following reasons influenced your failure to attend or drop out of regular school:	0 or 1 or 2 or 3 or 4 or 5.

[8] Microsoft Excel is a spreadsheet editor produced by Microsoft for computers using the Microsoft Windows operating system, in addition to Macintosh computers from Apple Inc. and mobile devices such as Windows Phone, Android or iOS.

Discrimination/prejudice based on race, gender, color, age or socioeconomic status.	

Source: elaborated by authors (2020).

Only after this stage of selection and cleaning, which involved operations such as dealing with the lack of values in some fields, checking for inconsistencies, reduction of registers, elimination of null values, duplicate data removal, and transformation for type *Attribute Relation File Format* (ARFF) files, EDM was applied as described below.

3.2.3 Educational data mining

From the information generated in the pre-processing step, it was possible to apply the clustering technique in WEKA[9]. The algorithm selected in this research was k-Means. According to Weber et al. (2013) and Silva et al. (2016) k-Means is considered the main clustering algorithm. The approach requires the user to define the number of groups they want to form. With this indication, the algorithm divides the set into disjoint parts within the established parameter, satisfying the following set of rules: (1) the elements in the same part, following the provided criteria, are close and; (2) following this same criterion, the elements of different parts are distant. The rules operated by methods to obtain this subdivision are: a random initial partition of k parts is created and in sequence, through repeated iterations the elements of the parts are reallocated to others, in order to refine the partitioning at each iteration. This means that in a new iteration, each part will contain objects that are actually close.

Therefore, the objective of EDM is to identify the socioeconomic profile of candidates who share the same characteristics, it is expected to obtain a set of groups with different combinations. The database containing 24,627 instances and 8 attributes was submitted to clustering using the k-Means algorithm. It adopted the approach of exploring the size k = 2 so that it could be made into groups divided in elementary and high school, thus, it was possible to analyze the clusters by level.

[9] It is a machine learning software that aims to computational and statistical analysis of the data provided, using data mining techniques, trying inductively, from the patterns found to generate hypotheses for solutions (WEKA, 2020). Available at: <https://www.cs.waikato.ac.nz/ml/weka/>.

3.2.4 Post-processing

At this stage, knowledge extracted by the k-Means algorithm in the clustering task must be understandable and useful for the decision-making process supporting the planning of new strategies.

In the next section, the results of the socioeconomic profile of the candidates from Encceja are described and discussed.

4. Results and discussion

In choosing the size k = 2 in the k-Means algorithm, divided into two categories, TP_CERTIFICACAO = 1 which corresponds to elementary school and TP_CERTIFICACAO = 2 which refers to high school. Of the total number of candidates in the base (24,627) who took the Encceja in 2017, the majority (18,655) took the exam to obtain the certification of Elementary Education (cluster 0), consequently, the minority (5,972) did for High School (cluster 1), as can be seen in Figure 1.

This information can be explained because, according to INEP (2017) with the creation of National High School Exam (ENEM in portuguese), Encceja is no longer mandatory for high school students. After that, a certification started to be carried out nationwide with the results of ENEM.

Figure 1 shows the average age of 28 years for the participants (NU_IDADE) who took the Elementary School test (cluster 0) and 22 years (cluster 1) for High School.

The data are in accordance with the LDB, in which the minimum age is 15 years for Elementary School and 18 years for High School level, age reached until the date of the tests.

Regarding the questions analyzed, corresponding to Q22, Q23, Q28, Q42, Q43 and Q49, chosen by the specialists who participated in this research in the pre-processing stage, the answers were classified by the degree of importance and influence, determined by an increasing scale from 0 (least important) to 5 (most important), shown in Table 2.

Figure 1 – Clustering in Weka.

```
=== Clustering model (full training set) ===


kMeans
======

Number of iterations: 3
Within cluster sum of squared errors: 25392.0

Initial starting points (random):

Cluster 0: 1,28,1,5,5,5,0,0,0
Cluster 1: 2,46,0,0,5,0,0,0,0

Missing values globally replaced with mean/mode

Final cluster centroids:
                                Cluster#
Attribute         Full Data          0          1
                  (24627.0) (18655.0)   (5972.0)
=============================================
TP_CERTIFICACAO           1          1          2
NU_IDADE                 28         28         22
Q22                       5          5          0
Q23                       5          5          5
Q28                       5          5          5
Q42                       0          0          0
Q43                       0          0          0
Q49                       0          0          0


Time taken to build model (full training data) : 0.12 seconds

=== Model and evaluation on training set ===

Clustered Instances

0      18655 ( 76%)
1       5972 ( 24%)
```

Source: elaborated by authors (2020).

Table 2 – Questions of the socioeconomic questionnaire and answers.

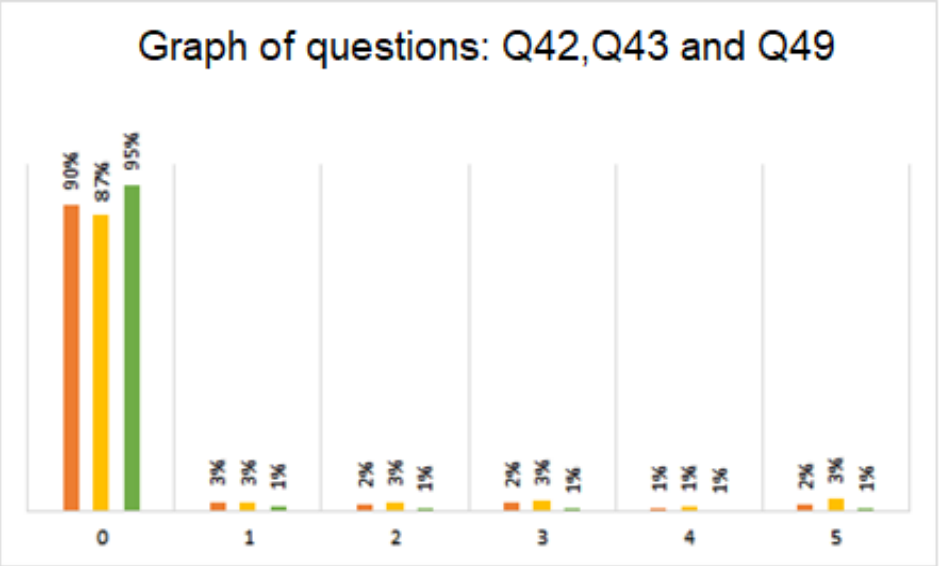
Questions	Answers
Q22) Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: To get the certificate of completion of Elementary School.	Of the total number of candidates (24,627) who took Encceja, the majority (18,655) considered the Exam for the certification of elementary education (cluster 0) *very important*.
Q23) Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: To get the certificate of completion of High School.	All participants (24,627) consider Encceja *very important* for medium level certification (cluster 0 and 1).
Q28) Indicate the degree of importance of the reasons that led you to participate in ENCCEJA: Because it is the best way to reconcile my studies and work.	All candidates (24,627) indicated that it is *very important* to participate in Encceja, because through the Exam it is possible to reconcile work with studies (cluster 0 and 1).

Q42) The extent to which the following reasons influenced your failure to attend or drop out of regular school: There are no places in public schools.	All candidates (24,627) indicated as *"not very important"* the lack of places in public schools, reason for not having attended or abandoned regular school (cluster 0 and 1).
Q43) The extent to which the following reasons influenced your failure to attend or drop out of regular school: Absence of school close to home.	All candidates (24,627) mentioned the absence of a school close to their home as *"not important"* as a reason for not having attended or dropped out of regular school (cluster 0 and 1).
Q49) The extent to which the following reasons influenced your failure to attend or drop out of regular school: Discrimination/prejudice based on race, gender, color, age or socioeconomic status.	All candidates (24,627) considered social and economic issues as *"not important"* as the reasons given for not having attended or dropped out of regular school (cluster 0 and 1).

Source: elaborated by authors (2020).

The analysis generated in the EDM process for questions Q42, Q43 and Q49 was extended, as the clusters for these questions obtained 0 (zero) as an answer. In this context, through statistical analysis it was noted that not all candidates considered the relationship between the questions asked "not important", but as the percentage is very low and not significant to be part of a cluster. The k-Means algorithm performed the clustering by extreme values 0 or 5, as can be seen in Graph 1. The orange bars correspond to the answers of Q42, the yellow ones to Q43 and the green ones to Q49.

Graph 1 – Analysis of questions Q42, Q43 and Q49.



Source: elaborated by authors (2020).

In this way, it can be seen that all students analyzed consider Encceja important for the certification of High School, even though its application is no longer mandatory,

being the responsibility of ENEM. It was found as a result that one of the factors that lead candidates to take Encceja is linked to the reconciliation of work and study. And, it was noted that there is no relationship between dropping out of regular school due to the lack of vacancies, nor due to the absence of a school close to the residence or the fact of discrimination or prejudice related to race, gender, color, age or socioeconomic.

In this context, it was possible to observe that these last three questions (Q42, Q43 and Q49) chosen by the specialists, both in the area of Encceja and in EDM, had no influence on the candidate choices for doing Encceja.

5. Conclusions

As contributions of this work, it was found that social and economic issues, such as discrimination or prejudice related to race, gender, color and age, are not related to school dropout. It was also identified that the lack of a place in a public school or the absence of a school close to the residence does not influence school dropout. According to INEP (2017), there are several reasons that lead students to interrupt their school trajectories, as found in this research, one of these reasons is linked to the service, in which some candidates reported that Encceja is the best way to reconcile studies with the job.

Based on the profile of the candidates found, it is possible to qualify the debate on the role of Encceja in the set of YAE policies, decide the best strategies that can increase the number of subscribers to the test and trace the trajectory of the candidates for certification. These results represent a contribution for managers in decision making in relation to the next Exams.

However, it should be noted that, as it is an experiment, the results of this work are limited to the data and methodology of the analyzed region, but it does not invalidate its use for other educational scenarios, with the necessary adaptations. In this way, as a future work, it is proposed to analyze the tests applied in Brazilian regions, as well as the application of other educational data mining techniques to evaluate and validate the results.

References

ALVES, E.; COMERLATO, D.; SANT'ANNA, S. **Relatório de pesquisa 2017: mapa da educação de jovens e adultos no Estado do Rio Grande do Sul**. 2017. Available in: <https://www.lume.ufrgs.br/bitstream/handle/10183/173114/001060717.pdf?sequence=1>. Consulted: 19 Nov. 2020.

CATELLI JUNIOR, R.; GISI, B.; SERRAO; L.F. Encceja: cenário de disputas na EJA. **Revista Brasileira de Estudos pedagógicos (online)**. Brasília, v.94, n.238, p.721-744, 2013.

DA SILVA PINTO, G.; JÚNIOR, O.; DE BARROS COSTA, E. Identificação dos fatores de melhorias no IDEB pelo uso de mineração de dados: um estudo de caso em escolas municipais de Teotônio Vilela-Alagoas. [S.l.]. **RENOTE,** v.17, n 3, p.183-193, 2019.

GARCÍA, E.: ROMERO, C.; VENTURA, S.; CASTRO, C. A collaborative educational association rule mining tool. **The Internet and Higher Education.** Elsevier, v.14, n.2, p.77-80, mar. 2011.

HERPICH, F.; NUNES, F.; CAZELLA, S.; TAROUCO, L.M. **Mineração de Dados Educacionais: uma análise sobre o engajamento de usuários em mundos virtuais.** Anais dos Workshops do V Congresso Brasileiro de Informática na Educação (CBIE 2016), Uberlandia/MG, 2016. Available in: <http://dx.doi.org/10.5753/cbie.wcbie.2016.910>. Consulted: 30 Sep. 2020.

INSTITUTO NACIONAL DE ESTUDOS E PESQUISAS EDUCACIONAIS ANÍSIO TEIXEIRA. **Microdados do Encceja 2017**. Brasília: Inep, 2017. Available in: <http://portal.inep.gov.br/microdados>. Consulted: 30 Sep. 2020.

MARTINS, W.; SA, H.C. Uma síntese do contexto do Exame Nacional para Certificação de Competências de Jovens e Adultos para Pessoas Privadas de Liberdade e jovens sob medida socioeducativa (Encceja PPL) nas unidades prisionais do Distrito Federal. [S.l.]. **Cadernos RCC**, v.5., n.3, p.225-233, 2018.

REBOUÇAS, I.; SANTOS, F.; PEREIRA, M.; SILVA, R.; PINTO, V. A Trajetória do exame nacional para certificação de competências de jovens e adultos. *In:* OLIVEIRA, S. (Org.). **A Educação de Jovens e Adultos em Tempos de Incertezas.** São Paulo: Editora Soul, p.1-374, 2019.

ROMERO, C.; VENTURA, S. **Data mining in education**. [S.l.]. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, v.3, n.1, p.12-27, 2013.

SANTOS, M.; ALMEIDA, G.; SOUSA, L.; SILVA, S. Performance in Mathematics of young and adults in lower Secondary Education in ENCCEJA. [S.l.]. **Research, Society and Development,** v.8, n.7, p.1-18, 2019.

SILVA JÚNIOR, W. **Uma Análise Curricular da matemática dos programas Encceja, nova EJA e PEJA no estado do Rio de Janeiro.** Masters Dissertation – Programa de Pós-Graduação em Ensino de Matemática da Universidade Federal do Rio de Janeiro (UFRJ). Rio de Janeiro: UFRJ, 2015.

SILVA, J.C.; RAMOS, J.; RODRIGUES, R.; SOUZA, F.; GOMES, A.S. Análise do engajamento de estudantes com base na Distância Transacional a partir da Mineração de Dados Educacionais. [S.l.]. **RENOTE**, v.14, n.1, p.1-11, 2016.

WEBBER, C.G.; ZAT, D.; LIMA, M.F. Utilização de algoritmos de agrupamento na mineração de dados educacionais. [S.l.]. **RENOTE,** v.11, n.1, p.1-10, 2013.

WEKA. **Waikato Environment for Knowledge Analysis** (WEKA). Available in: <https://www.cs.waikato.ac.nz/ml/weka/>. Consulted: 08 May 2020.