Problem Definition: Fair Evaluation of High Schools in Brazil with machine learning

Juan David Nieto García, Giovani de Almeida Valdrighi

1 Introduction

The evaluation of schools at the national level is carried out with the main objective of guaranteeing quality education for citizens, identifying areas of improvement, alignment of educational standards and decision-making planning. However, with the increase of competition between students and between education institutions, the evaluation of schools in recent years has become a tool to "press" and "target" students and schools with lower performance, even to catalog them as *bad* or *good*. Whereas the evaluation should be used in strategic ways to identify which schools need support, grants, and projects to improve their results and get ahead of education at a social level.

In Brazil, schools are currently being evaluated through two tests that are applied to students to obtain different performance metrics. Firstly, ENEM (National High School Examination) is a test standardized by the National Institute of Educational Studies and Investigations Anísio Teixeira (INEP) that evaluates different areas and is highly known because it is used as an entrance exam for universities. On the other hand, there is the SAEB test (Basic Education Assessment System), which is also standardized and administered by INEP, its main focus is to measure and analyze the quality of basic education in Brazil.

Until 2015, the Brazilian government reported the results of each school in the ENEM, and these values were used by the media as a ranking. After discussions, the government stopped publishing this information because it understood that it was not a fair and meaningful measure, but why? One of the main reasons is that the data began to be published and used incorrectly by institutions, some schools even created special classes of elite students to obtain good results in the ENEM and use the data as advertising [1]. Thus, disparities in the educational results between rich and poor schools were notoriously observed since the schools ranked as "less educational quality" presented students with conditions of economic and social vulnerability, in contrast with students with more privileges and opportunities in the school at the top of the ranking.

Bruce Baker in his book *Educational Inequality and School Finance: Why Money Matters for America?s Students* [2] mentions that the greater the monetary investment in education that students have, the greater the tendency to have better results. This can influence the general quality of the education offered, and it is confirmed in the last ranking of the ENEM 2023 [3], in which only one of the first ten institutions is public, and the rest are private. For this reason, it is necessary to implement strategies that evaluate the schools fairly, taking into account the different factors that influence their results, in order to redefine the true meaning of the evaluation of the level of education in Brazil.

Problem Definition. The present project seeks to evaluate the schools through unsupervised and supervised Machine Learning (ML) methods in a fair way. As a starting point, we want initially to confirm with our prediction models of ENEM and SAEB scores that schools with better economic conditions are more likely to have better results, and those that do not are more likely to be in the last places of the ranking. Secondly, our goal is to present a fair methodology that is able to separate schools into groups of similar economic conditions (unsupervised) and, within each group, create a model to score the schools based on their results at the exams (supervised).

2 Related Work

Previous works have already used ML in the context of education in Brazil. ENEM is performed yearly with a large population of students, and the complete and anonymized data is published in the following year. This motivated the development of a large corpus of work done in the development of ML models for the prediction of the ENEM results based on the socioeconomic characteristics of the student [4], [5], [6]. Generally, these works aim to identify the relationship between the social condition of the students and the result in the exam, so the predictor model is not the final product but the analysis of how this model performs its predictions. Stearns et al. [4] identified that the most important features are spatial position, age, and household income. Banni et al. [5] in a similar way identified the relation of the parent's educational level, household income, race, and age in the ENEM results. In a similar context, machine learning has already been used to predict school dropouts based on the socioeconomic factors of students.

School dropout is a critical problem in the social and economic development of the country. Despite the many projects employed to reduce it, identifying susceptible students is vital to the effectiveness of these efforts. [7] modeled the probability of students dropping out of high school based on their characteristics and the characteristics of the school. By employing significance tests, it was able to identify the important variables related to the dropout rates. Similarly, [8] used the information of first-semester undergraduate students to predict the dropout probability before graduating. These studies demonstrate the usefulness of machine learning models for education.

However, these large and complex models can also make unfair decisions and promote bad decisions from schools. Cathy O'Neil, in the book Weapons of Math Destruction [9], presents diverse real-life situations in which the use of data-driven models resulted in discrimination, unfairness, and feedback loops where decisions were made with the intent to improve the models' predictions, but not the real measure that it was supposed to represent. One of these examples is the data-driven ranking of colleges created by U.S. News magazine in 1988, which uses proxies of school excellence in an algorithmic fashion. Due to the relevancy that it obtained, a college obtaining a high position in the U.S. News ranking was of great importance. It could be the most important factor in attracting the best students and professors. However, due to the methodology used, many colleges decided on policies that were positive for their ranking but not for the students. One example is that a low acceptance rate was a variable used to measure excellence. By rejecting more students, the colleges could improve their position, a negative outcome for the students.

As previously mentioned, Brazil's government published the mean results of ENEM for each school until 2015, resulting in a similar ranking of schools. Even though there is no algorithm or model generating the scores (it is just a mean value of the students' results), this score was not representative of the school's quality and also resulted in negative feedback loops. The measure of school quality, despite being very important, can result in many problems when the ethical and fair aspects are not taken into consideration. This work tries to present a step in the correct direction.

3 PLANNING

The initial step of development is going to be the cleaning and analysis of the data sources. The values of ENEM and SAEB are published yearly and include socioeconomic variables of students and information regarding the schools' infrastructure and its economic condition. Despite the results of ENEM being available at a student level, the data is anonymized and does not present the names and addresses of students. To be anonymized, it also does not present the name of the school of the student, so it is not possible to aggregate students by school. This problem could be avoided in our work because it is available information of the type of the school, such as if it is a public or private school or the administrative level. The results from the students could be aggregated by the different types of schools in each city. This large quantity of data needs to be analyzed, studying what each information represents and what is useful for evaluating schools.

After data cleaning, our work can be separated into two steps. In the first step, a regression model is going to be used to predict the results of ENEM based on the student's socioeconomic variables, and another model to predict the scores in SAEB based on the school variables. Different models are going to be considered, such as linear models, nearest neighbors, decision trees, and feed-forward networks, each one with hyperparameter tuning to obtain the best R^2 metric (the variance of the target variable explained by the model prediction). After the fit, the models are going to be carefully evaluated using fairness metrics. Separating schools into groups of public or private schools (other groups could also be defined based on other attributes), the performance of the model is going to be evaluated within it group. Despite not being the standard, fairness metrics for regression have already been studied [10]. Jointly with the fairness metrics, interpretability techniques are necessary to identify which features the model is using to make its predictions, such as counterfactual explanations [11].

The second step is the development of predictor models defined for groups of schools. Using unsupervised clustering algorithms, it is possible to separate schools based on relevant infrastructure characteristics. By comparing schools of the same infrastructure, a more fair score is obtained. Unsupervised techniques generally are not evaluated with metrics, as we do not have the ground truth groups, however, schools of different groups can be compared to identify what are the attributes that separate them. Next, regression models are going to be fit using samples of each group of schools to predict their SAEB scores. An interesting idea could be to replace the target SAEB scores with the difference between the school SAEB score and the mean score of its group and use this new target in a unique model for all groups. This idea needs to be evaluated using the same fairness metrics and the interpretability presented in the first step so that it is possible to compare if the new model is an improvement from the initial one.

During development, we hope to envision addressing each of the trustworthiness aspects in the following way:

1) **Competence:** extensive analysis and data engineering, as well as different models, will be considered to improve performance.
2) **Integrity:** privacy and fairness will be considered throughout the development process, using suited techniques to correct possible problems.
3) **Predictability:** the learned model will be studied with interpretability techniques.
4) **Selflessness:** Tte goal at the end of the model is social good, identifying schools in need of support and exemplary schools of great results.

4 LIMITATIONS

The proposed methodology is going to use repurposed data, i.e., data originally used for the evaluation of students and their entrance to universities is going to be used with the intent of measuring school excellence. This metric is a proxy of school quality, as other factors can impact the performance of students rather than the school: their background, available time for studying, etc. For that reason, the analysis and conclusions made in this work need to be made carefully to obtain a fair evaluation of schools.

This work is an initial step in the development of an automatic system of school evaluation. The knowledge from specialists in education is of great importance for the validation of the methodology and conclusions. School managers, teachers, and students are also very important to participate in discussions regarding the models and its implications, as they are the group impacted by the outputs of the model. With that in mind, the methodology must be implemented in such a way that it can be updated with new data and the addition or removal of features, so it can easily improve from future feedback.

5 Author's Statement

Juan David Nieto García is a Ph.D. student in Computer Science who, who since his undergraduate had several accomplishments that broadened his knowledge in his field. He believes that education goes beyond numbers that label a student and schools, education must be for life, to develop skills that allow us to face the world in the different fields of our lives.

Giovani de Almeida Valdrighi is a Ph.D. student in Computer Science who, since young, had great opportunities in government projects of education. He participated in Math Olympiads and obtained support for a high-quality basic education. He believes in fair opportunities for everyone based on their needs, believes that the free market rules can lead to undesirable social outcomes, and believes in the value of rigorous science developed for the social good.

References

[1] g1. (2017) 'ranking' do enem por escolas deixará de existir; entenda a mudança. [Online]. Available: https://g1.globo.com/educacao/enem/2017/noticia/ranking-do-enem-por-escolas-deixara-de-ser-divulgado-diz-mec.ghtml

[2] B. Baker, *Educational Inequality and School Finance: Why Money Matters for America's Students*. 8 Story Street First Floor, Cambridge, MA 02138: Harvard Education Press., 2018. [Online]. Available: https://eric.ed.gov/?id=ED594085

[3] (2023) Melhores escolas do brasil no enem 2023. [Online]. Available: https://enem.net/melhores-escolas-do-brasil-no-enem-2023

[4] B. Stearns, F. Rangel, F. Firmino, F. Rangel, and J. Oliveira, "Prevendo desempenho dos candidatos do enem através de dados socioeconômicos," in *Anais do XXXVI Concurso de Trabalhos de Iniciação Científica da SBC*. Porto Alegre, RS, Brasil: SBC, 2017. [Online]. Available: https://sol.sbc.org.br/index.php/ctic/article/view/3244

[5] M. Banni, M. Oliveira, and F. Bernardini, "Uma análise experimental usando mineração de dados educacionais sobre os dados do enem para identificação de causas do desempenho dos estudantes," in *Anais do II Workshop sobre as Implicações da Computação na Sociedade*. Porto Alegre, RS, Brasil: SBC, 2021, pp. 57–66. [Online]. Available: https://sol.sbc.org.br/index.php/wics/article/view/15964

[6] R. L. C. S. Filho, "Modelo de análise e predição do desempenho dos alunos dos institutos federais de educação usando o enem como indicador de qualidade escolar," in *PhD Tesis UFPE*, 2017.

[7] F. S. Machado, "Análise e previsão da evasão escolar do ensino médio através de dados públicos," *Escola de Matemática Aplicada, Fundação Getúlio Vargas*, 2019.

[8] A. M. Barbosa, E. Santos, and J. P. P. Gomes, "A machine learning approach to identify and prioritize college students at risk of dropping out," *Sociedade Brasileira de Computação*, 2017.

[9] C. O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. USA: Crown Publishing Group, 2016.

[10] H. Narasimhan, A. Cotter, M. Gupta, and S. Wang, "Pairwise fairness for ranking and regression," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, no. 04, 2020, pp. 5248–5255.

[11] I. Stepin, J. M. Alonso, A. Catala, and M. Pereira-Fariña, "A survey of contrastive and counterfactual explanation generation methods for explainable artificial intelligence," *IEEE Access*, vol. 9, pp. 11 974–12 001, 2021.