education sciences

MDPI

Article

On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout

Elvis Melo [1], Ivanovitch Silva [1,2,*], Daniel G. Costa [3], Carlos M. D. Viegas [2] and Thiago M. Barros [4]

[1] Postgraduate Program in Electrical and Computer Engineering, Federal Univesity of Rio Grande do Norte, Natal 59078-970, RN, Brazil
[2] Department of Computing Engineering and Automation (DCA), Federal Univesity of Rio Grande do Norte, Natal 59078-970, RN, Brazil
[3] INEGI, Faculty of Engineering, University of Porto, 4200-465 Porto, Portugal
[4] Federal Institute of Rio Grande do Norte (IFRN), Natal 59015-000, RN, Brazil
* Correspondence: ivanovitch.silva@ufrn.br

Abstract: The school dropout problem has been recurrent in different educational areas, which has reinforced important challenges when pursuing education objectives. In this scenario, technical schools have also suffered from considerable dropout levels, even when considering a still increasing need for professionals in areas associated to computing and engineering. Actually, the dropout phenomenon may be not uniform and thus it has become urgent the identification of the profile of those students, putting in evidence techniques such as eXplainable Artificial Intelligence (XAI) that can ensure more ethical, transparent, and auditable use of educational data. Therefore, this article applies and evaluates XAI methods to predict students in school dropout situation, considering a database of students from the Federal Institute of Rio Grande do Norte (IFRN), a Brazilian technical school. For that, a checklist was created comprising explanatory evaluation metrics according to a broad literature review, resulting in the proposal of a new explainability index to evaluate XAI frameworks. Doing so, we expect to support the adoption of XAI models to better understand school-related data, supporting important research efforts in this area.

Keywords: eXplainable Artificial Intelligence; educational data science; school dropout; SHAP explanation method; learning analytics; artificial intelligence

Citation: Melo, E.; Silva, I.; Costa, D.G.; Viegas, C.M.D.; Barros, T.M. On the Use of eXplainable Artificial Intelligence to Evaluate School Dropout. *Educ. Sci.* **2022**, *12*, 845. https://doi.org/10.3390/educsci12120845

Academic Editors: Mohammed Saqr and Sonsoles López-Pernas

Received: 27 October 2022
Accepted: 18 November 2022
Published: 22 November 2022

1. Introduction

The technological revolution in last decades has brought new challenges to educational institutions in different levels, with schools striving to motivate students that are already being raised in a new age of information. Since the educational methods are not evolving in the same pace, due to inherent complexities of the teaching and learning process, students' dropout has become more common, rising different concerns that can be ultimately mapped to how our society evolves [1]. Hence, understanding how such dropouts are happening has become a very important issue for the entire schooling system [2].

The use of mathematical and computational tools to better understand the different dynamics in schools is not a novelty, with many works addressing it lately. Among the possibilities, Artificial Intelligence (AI) has gained momentum [3], bringing opportunities and challenges for the education system at the macro level. In fact, the formulation of policies and planning, the optimization of the provision and management of education, teacher training, the improvement of learning results, the expansion of access to schools, and the improvement of educational quality, are some of the expected results when school data are processed using proper AI solutions [4]. Since artificial intelligence algorithms may be exploited to highlight issues related to admission to public institutions and identification of school dropouts, new discussions and decisions can be better supported when artificial intelligence is properly leveraged.

Although AI solutions have a high potential to provide opportunities for teachers and educational managers to better exploit school data, artificial intelligence has also brought ethical concerns that should be considered, demanding new ways to process such data [5]. Actually, ethical violations when using AI may cause a digital division between social networks and economic groups, bringing relevant discussions from the point of view of humanity, ethics and justice [6]. As a result, the use of AI methods has to be as transparent as possible, also being easily accessible and understandable by all educational entities. Hence, when considering the school scenario, it has been recognized that the adopted AI models and results have to be more easily known, ensuring more efficient uses of the results while assuring a more ethical use of the data.

The term "school dropout" refers to the act of not attending classes, in which a student drops out school by any reason, with no subsequent return [7]. The adoption of AI solutions to understand school dropout levels has then to consider not only the available data, but also whether such solutions will be perceived by the educational entities, and the same is true for the identified patterns. In short, the intended "identification" patterns is of paramount important, which has been pursued following two different approaches: gray-box or black-box [3]. The construction of the former discloses the structure of the latter, explaining the elements associated to the processed data, while the construction of the second is inexplicable. Since we want to make the models more transparent, eXplainable Artificial Intelligence (XAI) comes as a way to transform AI from a black-box model to a gray-box. The objective of XAI is to create a set of techniques that produce more explainable models, while still maintaining high performance levels [8], which could be better adopted by schools when making decisions.

At this point, even though AI techniques have enabled the development of different data-centric approaches, a problem that has recently received more attention is that training datasets can reinforce or reflect prejudices [9]. This way, the construction of XAI models comes as a facilitator for educational professionals to use AI solutions based on school data, mitigating possible influences that the data may have to affect the model (bias and prejudice). Thus, together with ethical issues considerations, its is expected that XAI models can be more adequate to a school context [10].

Therefore, this article aims to apply and evaluate methods of explainability on a black-box model from a database of the Federal Institute of Rio Grande do Norte (IFRN) for predicting students in school dropout situations. For that, a comprehensive literature review will be performed, focusing on XAI approaches that can be applied to different educational scenarios. Then, an explainability index will be created to support the selection of the available XAI methods. Finally, promising methods will be identified and applied on the defined dataset, with the achieved results being compared in order to support the selection of the most adequate solution. With the proper method, we believe that XAI may become a valuable tool to understating different dynamics in schools.

The remainder of this article is organized as follows. Section 2 presents the state of the art of XAI, highlighting different practical issues. Section 3 presents XAI frameworks for local and global explanations. The school dropout problem and the defined case study are described in Section 4. Section 5 applies the adopted methodology and presents the achieved results. Finally, Conclusions and References are presented.

2. A Review of eXplainable Artificial Intelligence (XAI) Approaches

When talking about explanations in AI, it is initially required to define some key concepts. According to the XAI literature, researchers often use the terms "interpretability" and "explainability" as synonyms [11]. In this article, we consider that explainability is related to interpretability, in such a way that explainable models are interpretable whether their operations can be understood by humans. Thus, the explainability of a model can generate interpretability of its decisions.

Recent advances in the Machine Learning (ML) area have led to a growing interest in XAI to enable humans to derive decision-making information from machine learning models [12]. According to the authors in [12], XAI is concerned with understanding and interpreting the behavior of AI systems. The rapid progress in complexity and sophistication accelerated, eventually creating black boxes. However, on many occasions, it is challenging to understand the decision and bias of controlling and relying on unexpected or unpredictable outputs from systems. Therefore, the loss of control over the interpretability of decisions making becomes a critical issue for many data-driven automated applications [13]. Actually, according to [14], much of the literature related to XAI comes from the ML and data mining communities. The first focuses mainly on describing how black-box models work, while the second is interested in explaining decisions, even without understanding the details of how gray-box systems work in general. We understand XAI as a field of study that is concerned with making explainable the AI black-box models.

The following subsections further discuss the state of the art when adopting XAI methods.

2.1. Fundamental Concepts

Some works have discussed concepts that have influenced the way XAI approaches have been created and applied in practical scenarios. In [15], the authors sought to explain the need for XAI. Among the possibilities, that work argued that XAI can be exploited for control, improvement, and discovery. In fact, the concept of explainability is not only important for justifying decisions, but it also helps to prevent things from going wrong and to identify possible discrimination in more sensitive decision models. It may help to clarify model decisions that involve the use of sensitive data, in accordance with data usage regulation laws.

Among the most explainable models in the literature, there are Decision Trees, Decision Rules, and Linear Models [16]. These models are considered easily understandable and interpretable by humans. In [15], the authors consider Decision Trees as an understandable global predictor, assigning it the most explainable model status within the ML models.

Bringing explainability to ML models is a very challenging technical issue. A more explainable model has a linear shape, according to [13,14]. Most XAI frameworks try to linearize local and global predictions through adverse techniques [15].

In [9], it is often useful to visualize the characteristics of the parameters learned from a model and the explanation of its interactions with a dataset. The importance of the resource and attribution scores can be most useful to the insights when analyzed in a visual way, exposing patterns that would be difficult to discern. This is possible when a model has a linear shape, hence the importance of XAI methods that seek the linearization of black-box models, facilitating the visualization and interpretation of ML models. The authors also bring discussions about the importance of interactive visualizations, as they allow the real-time exploration of the parameters learned from the model.

For the authors in [9], XAI works directly with the concept of explainability, in parallel to interpretability, as a result of working with agnostic models. In this sense, explicability implies interpretability, but not the other way around. Explanation refers to the understanding, in simple terms, of how exactly a model works from the inside, while interpretability refers to the ability to observe the effect that changes on the input parameters will have on the expected outputs. The authors seek to divide the concepts of explainability and interpretability in their definition. We understand that these concepts are related to the identification of how the internal mathematics of the black-box model works, as well as the way of manipulating the model parameters so that it may be interpretable.

The authors in [17] wondered about the role of XAI. For them, the interpretation of black-box models must take into account the explanation of how the model learns. They discussed the main benefits of XAI, which are: the verification, improvement and learning of the model, in addition to making clear issues of data transparency and increased reliability of AI models. Thus, understanding how the model learns is essential for explanation, and consequently for the interpretation of the model.

For the authors in [18], XAI aims to improve the performance of a model, showing how the method is better than others by considering a common metric. These methods have been tested with humans. As a way of testing the model, three aspects were pointed out by the authors regarding the evaluation of explanation with humans: explanations are used to identify what is missing in the model, the time it takes to be explained and the expertise of the user. In these explanations, aspects such as composition, monotonicity, information uncertainty, graphs and numbers as units of explanation must be highlighted. The researchers sought to discuss issues related to the explainability and interpretability of the black-box models taking into account ethical issues with the participation of the user to validate their explanations. Human participation in the process of designing the XAI model is indispensable.

The explanations of XAI are treated as justification for the process of interpreting the AI models for [19]. An explanation must show the importance of the features and their weight in each decision. In addition, explanations can be divided into three types: through randomly chosen items, choices by similar users, and based on features, with more than one type of explanation being able to be combined. For them, a work of visualization of the explanations is essential for the model to be interpretable, in addition to provide to the user of the model a greater interactivity with the resources it contains. In addition, the complementing of explanations by different XAI models may ensure greater interpretation of the model.

The work of [18] also has a contribution in relation to the choice of a better presentation of the results of a AI model. According to the researchers, the simulation of using the model with the knowledge about the desired inputs and outputs, as well as asking what can be changed in the model so that an output is the desired one, are important steps for the model to become explainable.

Regarding the transparency of a XAI model, it can be characterized into three levels, according to [20]: it can be simulable, it can be partitioned, or the output value is the explanation itself. For that work, these explanations can be accomplished through texts, graphics, formulas, and examples.

Among the discussed views on XAI, in general we can divide them into two aspects: the first in relation to the explanation of the forecast, and the second through the design of intrinsically interpretable models that can be explained through the interaction with the model. These factors will be related to the discussions in next subsections.

2.2. Applying XAI Methods

The function of an explanation is to facilitate learning. Only in this way is it possible to use a AI model for the benefit of knowledge generation by stakeholders interested in education. According to [21], people employ cognitive biases and social expectations in processes of explanation. In this sense, XAI serves for users to trust the decisions of the AI. This type of interaction raises ethical issues that require transparency in the construction and the use of these models in several sectors of society, such as in education, for example.

The search for explanations is intrinsically related to human curiosity and the desire to find meaning in the world. Usually, unexpected events are the ones that call our attention and there is a search for an explanation of the facts [22].

The work in [21] divides the explanation process into actors, according to the scheme presented in Figure 1.

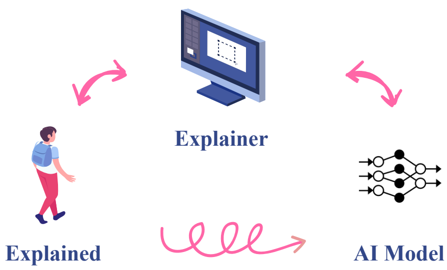

Figure 1. The explanation process, according to [21].

In Figure 1, there are three actors in the explanation process:

- Explained: The human who accesses the explanation;
- Explainer: The XAI interface between the explained and the AI Model;
- AI Model: The black-box model that needs an explainer to make its decisions explainable to humans.

In general, trust is lost when users fail to understand traits of behavior or decisions. This happens when the respondent tries to answer her/his doubts about the explanation by directly accessing the AI model. For [21], interpretability in XAI is related to how well a human can understand the decisions of a model.

In this sense, the explanation is used as the manager of a social process. It is a transfer of knowledge between the beliefs of the *explainer* and the *explained* with the objective that the latter has sufficient information to understand the causes of the event through a social interaction. Therefore, the explanations are contextual [21].

According to [23], prior knowledge is altered when someone asks for an explanation to be provided. These explanations are requested aiming to find meaning and managing social interaction with the result of the model [24]. Thus, the explanations have the objective of transmitting knowledge, persuasion, promoting learning and attributing blame, the latter influenced by beliefs and biases. This way, it is understood that the *explainer* and *explained* may have different goals.

In the explanation process, [21] maps four types of questions that the subject can ask when seeking interpretation of the model through interaction with the *explainer*. They are presented in Table 1.

Table 1. Questions asked in the explanation process, according to [21].

Question	Reasoning	Description
What?	Associative	Search for the reason why unobserved events could have occurred, due to observed events.
How?	Interventionist	Simulate a change in the situation to see if the event still happens.
Why?	Counterfactual	Simulate alternative causes to check if the event still happens.
How?	Abductive	Try to infer causes that explain events, making assumptions about hypotheses and testing them.

For certain types of questions, there will necessarily be some reasoning of what the *explainer* must perform in order to answer each one of them. Explainable models can be used to exemplify robust models, such as decision trees, decision lists, or local and global black-box approaches using a linear model, for example [16].

For the evaluation of explanations, the main criteria used is the consistency with previous personal beliefs, probabilities, simplicity of presentation, and generalization of the AI model. According to [25], simpler and more general explanations are better when compared to more elaborated ones. Better elaborated explanations have a greater number of causes and, therefore, become less generalist.

So far, several works have brought interesting metrics for analysis that were considered to evaluate the explanations made in the case study with the presented XAI frameworks. It was taken into account the argument that the greater the variety of explanations and objectivity considered in their views, the greater the contribution to the construction of an interpretation, hence the generation of knowledge. Thus, the more items an explanation provides, the better the XAI model will be.

2.3. Recent Developments in XAI for Education

Within the scope of education, there is an area that is concerned with the application of AI and data solutions for the optimization of educational processes, decision making, and learning support. This large area of study is called Educational Data Science (EDS) [26], encompassing works from Learning Analytics, Educational Data Mining (EDM), and Artificial Intelligence in Education [27]. The EDS can be explained as the intercession between Learning Analytics, EDM and also AI in Education, as shown in Figure 2.

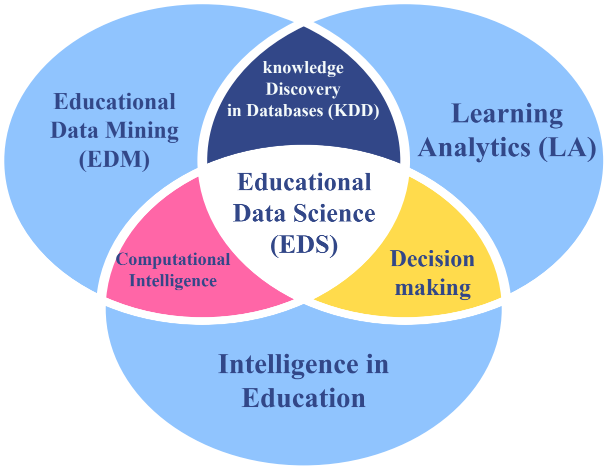

Figure 2. Defining the Educational Data Science research field.

In general, the EDS brings together areas of computer science, education, statistics and other social sciences to examine and understand the phenomenon of Education. Therefore, it can be defined as a data-driven, systemic, transdisciplinary and dynamic field that combines technical and social skills with a deep understanding of educational practice in different learning environments.

EDS uses procedures and techniques to gather, organize, process and interpret sources of large and diverse educational data, ensuring the consistency of these sets and creating visualizations to assist in understanding complex data. In addition, mathematical models are built to communicate insights to educators, managers, instructional designers, students and other stakeholders [28].

The EDS area has contributed to the common objective of improving the quality of the analysis of educational data on a large scale, to support both basic research and pedagogical practice. Among the most used techniques, machine learning and data clustering algorithms are being applied, mostly when analyzing student performance [29].

Learning Analytics has been providing the development of tools that optimize the student's teaching and learning process and the management of the education process as a whole [30]. Learning Analytics processes support teaching and learning, such as student assessment, curricula and activities, student performance feedback, help with self-regulation of learning, personification and increased quality of engagement in activities, in addition to performance prediction, such as school dropout. The use of AI in these processes is already present in some initiatives, but the teacher's adherence to the use of these tools requires an explanation of how it works so that confidence is gained [31].

Among the possibilities of using AI in the context of education with Learning Analytics, we were able to prospect activities in which students can receive recommendations on resources according to their performance, objectives and motivations, to be able to graphically analyze the results of their learning process, compare them with those of the rest of the class, and observe performance and contributions related to collaborative activities. Directors can use the information to design a better allocation of human and material resources to improve the overall quality of their academic offering. Finally, teachers and researchers can test and adapt their theories based on educational data [30].

Education Data Mining is interested in employing a data-based approach to make better decisions, as it is already common in business intelligence. Thus, in the context of EDM, there are statistical, ML and data mining methods and techniques for researching patterns and building predictive models or decision rules that can be adapted to educational data [32]. Using business analysis tools, which include predictive modeling, educators and administrators can make better decisions based on data.

Learning management systems, which provide students with course content and interactive tools, can be a resource for collecting student data. This tool is very useful for school managers in making important decisions, such as where to invest resources to prevent dropouts, for example. It is important that the models that can support these activities are explainable, thus making XAI essential for the processes in AI in education [28].

In general, we identified that XAI has an important role in providing the explained information so that confidence is increased when using ML models to support educational activities. Technologies and enabling methods for XAI potentially belong to the context of Education 4.0, with the aim of producing artificial explanations, which can lead to innovative explainable models and identification of previously unperceived patterns. Putting all these together, a robust way to support teaching and learning processes within the EDS study fields is achieved.

3. XAI Frameworks for Local and Global Explanations

There are several techniques for interpreting black-box models that try to deal with the limitations and challenges of traditional model interpretation techniques and to face issues of exchange of accuracy for interpretability of ML models.

In general, there are XAI models applied to black boxes such as Multilayer Perceptron (MLP), for example [15]. They are XAI models specific to a model type. However, Refs. [22,33] emphasize that XAI approaches can potentially be adopted to explain any type of model. It means that they are post hoc agnostic models, where those approaches may be adopted to explain any type of model.

For [22], it is easier to automate interpretability when it is separated from the ML model. According to the author, the XAI agnostic methods are better scaled and will have great adherence to the investigation by the XAI community. The advantage of agnostic model interpretability lies on its modularity, and the black-box model can be replaced by another model.

Thus, agnostic models are not linked to a specific type of ML model. Agnostic interpretations of models are generally post hoc, and can be local or global interpretable models [33]. A global agnostic model is an explainable model that is trained to approximate predictions to a black-box model that can be essentially any model, regardless of its complexity or training algorithm. A local agnostic model, on the other hand, seeks explanations for a specific model decision. There are works that develop XAI models that use Bayesian network models, causality algorithms, among others [15]. These models tell us which features are most important. Hence, Ref. [15] emphasizes that these approaches can potentially be adopted to explain any type of model. For these applications, the models generally need to be accurate and explainable, where interpretability means that we can understand how the model uses input resources to make predictions [33].

There are also XAI models applied to black boxes such as LIME [34], SHAP [35], Shapley values [36], among others [15]. LIME (Local Interpretable Model-agnostic Explanations) is a local explanation framework: it reports the decision path, uses a heuristic approach that assigns credit to each input resource and applies agnostic approaches to the model that require execution of the model for each explanation [34].

In LIME, an explanation is denoted as $g \in G$, where G is a set of models widely held to be interpretable. LIME generates a new dataset consisting of perturbed samples and the corresponding predictions of the black-box model [34]. LIME then trains an interpretable model, which is weighted by the proximity of the sampled instances to the instance of interest. The learned model should be an approximation of the black-box model predictions locally, but it does not have to be a good global approximation. LIME specifies the explanation as:

$$explanation(x) = \arg_{\min_{g \in G}} L(f, g, \pi_x) + \Omega(g) \tag{1}$$

The explanation model, for instance x, is the model g that minimizes loss L, which measures how close the explanation is to the prediction of the original model f, while the model complexity $\Omega(g)$ is kept low. G is the family of possible explanations. The proximity measure π_x defines how large the neighborhood around instance x is that we consider for the explanation. In practice, LIME only optimizes the loss part.

Among the ways in which it works, several agnostic approaches are applied, which require repeated execution for each explanation. The values constituted by the structure of the SHAP framework are called Shapley values, also being an XAI framework. These values provide a measure of the importance of each feature by approximating the predictions in a linear problem. The Shapley value is calculated for each variable trying to find the correct weight so that the sum of all Shapley values is the difference between the predictions and the average value of the model, corresponding to the contribution of each resource to approximate the forecast to the expected value [36].

In Shapley values, we are interested in how each feature affects the prediction of a data point. When calculating how much each feature contributed to the prediction, we have the sum of all feature contributions for one instance.

$$\sum_{j=1}^{p} \phi_j(\hat{f}) = \sum_{j=1}^{p} (\beta_j x_j - E(\beta_j X_j)) \tag{2}$$

$$\sum_{j=1}^{p} \phi_j(\hat{f}) = (\beta_0 + \sum_{j=1}^{p} \beta_j x_j) - (\beta_0 + \sum_{j=1}^{p} E(\beta_j X_j)) \tag{3}$$

$$\sum_{j=1}^{p} \phi_j(\hat{f}) = \hat{f}(x) - E(\hat{f}(X)) \tag{4}$$

when $\hat{f}$ is the linear model, x_j is the instance for which we want to compute the contributions, β_j is the weight corresponding to feature j, the contribution ϕ_j of the j feature on the prediction. $E(\beta_j X_j)$ is the mean effect estimate for feature j. The contribution is the difference between the feature effect minus the average effect. This is the predicted value for the data point x minus the average predicted value. Feature contributions can be negative.

The Shapley value is defined by a function value val of players in S. The Shapley value of a feature value is its contribution to the payout, weighted and summed over all possible feature value combinations.

$$\phi_j(val) = \sum_{S \subseteq \{x_1, x_2, \ldots, x_p\} / \{x_j\}} \frac{|S|!(p - |S| - 1)!}{p!} (val(S \cup \{x_j\}) - val(S)) \tag{5}$$

where S is a subset of the features used in the model, x is the vector of feature values of the instance to be explained and p the number of features. $val_x(S)$ is the prediction for feature values in set S that are marginalized over features that are not included in set S:

$$val_x(S) = \int \hat{f}(x_1, x_2, \ldots, x_p) d\mathbb{P}_{x \notin S} - E_x(\hat{f}(X)) \tag{6}$$

SHAP (SHapley Additive exPlanations), on the other hand, assigns to each resource a value of importance for a specific forecast, being a framework for global explanation. It identifies a new class of measures of importance of additive characteristics and ensures that the solution is unique in the SHAP class [37]. It is also identified the class of methods of importance to the additive resource, showing that there is a unique solution in that class that adheres to desirable properties. Despite this, as the agnostic methods of models depend on the post hoc modeling of an arbitrary function, they can be extremely slow and suffer sampling variability [15].

The SHAP explanation method computes Shapley values from coalitional game theory. The feature values of a data instance act as players in a coalition. Shapley values tell us how to fairly distribute the prediction among the features. A player can be an individual feature value, as well as a group of feature values. SHAP specifies the explanation as:

$$g(z') = \phi_0 + \sum_{j=1}^{M} \phi_j z'_j \tag{7}$$

where g is the explanation model, $z' \in \{0,1\}^M$ is the coalition vector, M is the maximum coalition size, and $\phi_j \in \mathbb{R}$ is the feature attribution for a feature j, the Shapley values. The representation as a linear model of coalitions is a trick for the computation of the all ϕ. For x, the instance of interest, the coalition vector x' is a vector of all 1's. The formula is simplified to:

$$g(x') = \phi_0 + \sum_{j=1}^{M} \phi_j \tag{8}$$

It is worth mentioning that in the literature there are XAI explainers for specific models and for global or local explanations. An example of an XAI model for local explanations is ELI5 [38]. It is an XAI framework developed in Python that helps to debug ML classifiers and explain predictions in an intuitive manner. It does not support agnostic explanations of the model, and support for models is limited to tree-based models and other parametric or linear models.

According to [9], the main appeal behind linear models is the simplicity of the linear relationship between inputs, weight parameters learned, and outputs. These models are often interpretable implicitly, since they can naturally weigh the influence of each resource and disturb the learned inputs or parameters with a predetermined effect on the outputs. However, different correlations among various resources can make it difficult to analyze the independent attribution of each resource in the resulting forecasts. There are many different forms of resource importance, but the general concept is to make the parameters more easily interpretable, in such a way that they are presented in an explainable way to users.

4. Applying the Proposed Method for School Dropout Evaluation

The problem of school dropout is directly related to the issue of the right to education. In fact, school success should be related to broader measures of individual development and to proficiency in curriculum content. However, there are several factors that promote school dropout that are not mapped by databases [28,39,40]. External factors, beyond the control of the school, also influence school engagement and success.

As the participation in any activity, the engagement of young people in school activities has an extensive component (whether or not they have any engagement), and an intensive component (the level of engagement) [39]. There are AI models that seek to explain school dropout in the dimension of evaluating an engagement in extensive components, which can be measured and placed in a database to be used by the algorithms, generating predictions of students in dropout situations. Intensive components, on the other hand, are difficult to map, and school management based on student data and indicators should promote actions to verify student engagement in not mapped situations.

There is a difference between school abandonment and dropout, with the latter being more expensive to the State. Dropout occurs when a student who went to school in a given year stops enrolling at the beginning of the next academic year. Abandonment, on the other hand, occurs when a student who enrolled at the beginning of the year stops attending school at any given time during the school year. Therefore, a student can drop out without ever leaving school [39]. The disengagement of young people in school activities has, in principle, consequences both for that young person's life, as well as for society in general.

In certain cases, the engagement of these students can be detected using AI models, if the evidence is part of the explanation of the model.

There are several factors that determine the lack of student engagement in school activities, as shown in Figure 3.

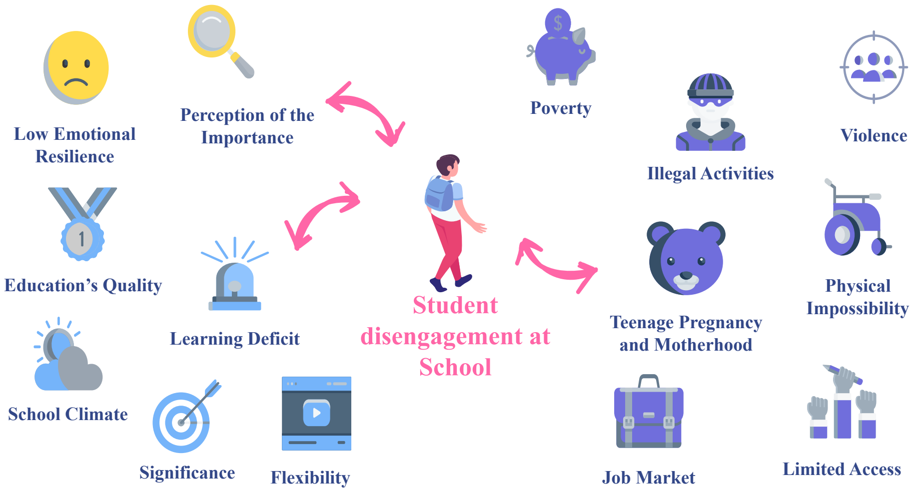

Figure 3. Factors that promote student disengagement at school, according to [39].

In general, the student disengagement is not the result of a single factor, but of a multiplicity of them. Each requires differentiated actions in order to mitigate its consequences. Thus, for an engagement promotion policy to be effective, it needs to have a wide range of actions capable of acting on this whole set of factors. In studies on features that most influence school dropout situation, when comparing people with the same socio-demographic characteristics, such as age, sex, race, and education, the salary of people with graduation is 544% higher when compared to illiterate people and the chances of occupation is 422% higher [7].

In a clustering study of students at risk of dropping out, the work in [40] suggested a three-group solution: (1) students with moderate to high cognitive abilities, but deficient learning strategies; (2) students with moderate to high cognitive and learning abilities; (3) students with low cognitive functions and moderate learning ability. According to the authors in [40], the literature on the causes and the dropout persistence suggests individual, institutional, and socioeconomic factors.

In an investigative study, [41] analyzed the main features in relation and concluded that data related to the student's ethnicity, study shift, and whether the student lives with a guardian, contributed more greatly to the classification in the dropout problem. In the research on the identification of school dropout, the importance of features that contribute to the identification of dropout can be organized into three major working groups, according to [39]:

- Assessments based on the opinion of key actors: Usually the teachers blame the students' family conditions and lack of interest, while the students blame the school's lack of structure;
- Inferences about the relative importance of a wide variety of factors from the analysis of the behavior of young people: Some variables are not mapped and data are collected for other purposes;
- Inference about the importance of specific factors also from the analysis of the behavior of young people: In addition to variables being collected for other purposes, the studies seek to predict more general aspects, but they answer specific questions that influence students' lack of interest in continuing their studies.

These evidences are present in educational databases, however, they are not always integrated or were collected for other purposes, such as school registration, mapping the profiles of students who underwent a certain evaluation, among others. In the first case, it is important to make families, parents and teachers aware of the real causes of school dropout, related to the motivation and engagement in school activities. In the second and third cases, it is important that the researchers, when carrying out their inferential research on the defined subjects, take into account the context and the integrity of the information contained in the databases that will be used for the construction of the XAI model.

4.1. Considered Dataset

The database of the defined case study comes from the Federal Institute of Rio Grande do Norte (IFRN), which includes data information of students from all 19 Campi spread throughout the mesoregions of Rio Grande do Norte in Brazil. Information on the family, school performance, financial field, ethnicity, demographics, and work situation, are present in the dataset.

An overview of the case study pipeline is depicted in Figure 4. For reproducibility, all source codes used in the implementation of the methodology and to obtain results were available in a public repository (https://github.com/elvismelo/paper_school_dropout accessed on 10 November 2022).

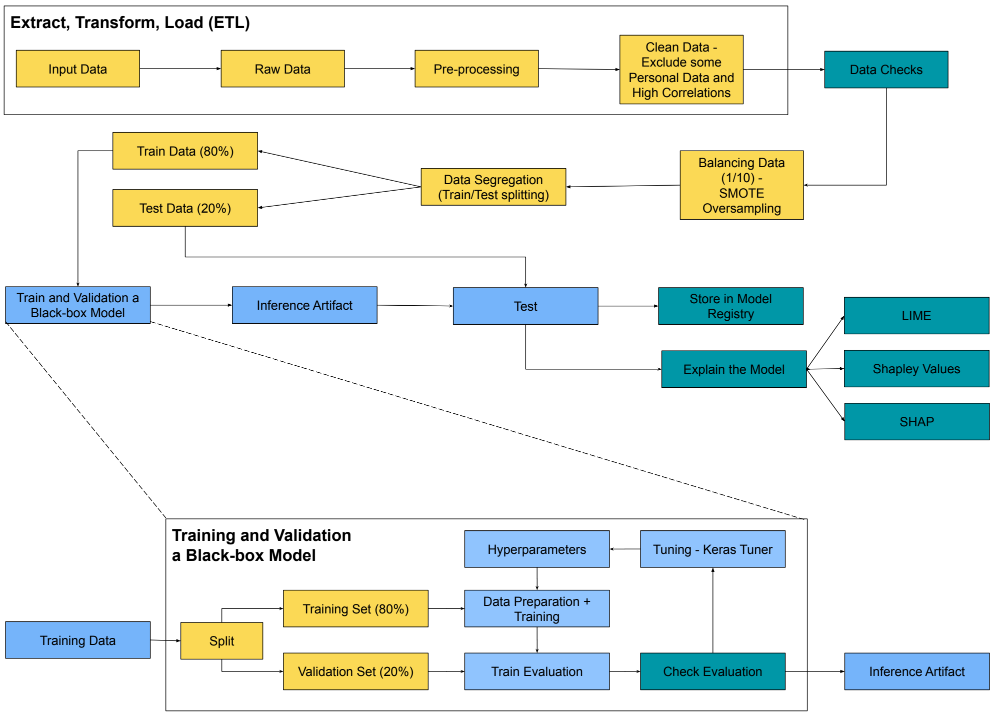

Figure 4. Pipeline of the IFRN dropout case study. Here, yellow steps are related to data processing, green steps are related to decision making and XAI, and blue steps are associated to the black-box model.

When observing the nature of the variables, 116 are dichotomous (false-0 or true-1 values) and 14 are numeric. The features related to the Campi, as well as the courses that the students come from have been removed. They corresponded to 41 features and were Boolean values that demarcated the regionality of the data, since the data were from all Campi in the state of Rio Grande do Norte. Student enrollment numbers, highly correlated values, as well as identifier in the system were also removed. As a result, 75 features were considered in this IFRN school dropout case study.

4.2. Applying XAI

For a more adequate analysis of the problem, it was interesting to see how the data in this area are arranged. Based on a previous study using the problem database [2], an unbalanced data problem is considered in a 1/10 proportion, with the minority class being the number of students in dropout situations (See Figure 5).

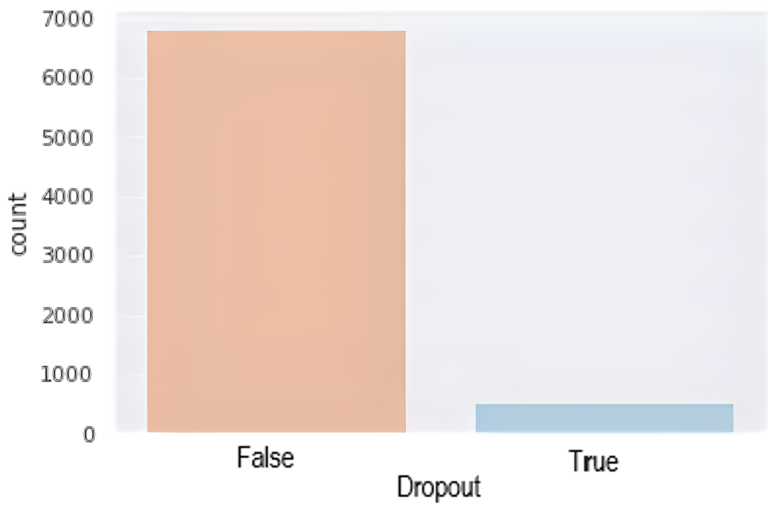

Figure 5. Class unbalance dropout.

Thus, in the study, an ML architecture was found through several tests. The best way to balance the database was using a technique called SMOTE [42,43]. The SMOTE technique for balancing the dataset was chosen because it is an oversampling technique, which would decrease the difference between the majority class (which are the non-dropping out students) and the minority class (which are the dropping out students). Thus, the model received the same proportion for training.

For the use of the balanced database in a black-box classification model, we used MLP as it is an universal approximator [33]. Regarding the architecture, using an optimized grid search Keras Tuner tool [44] to find the best combination of the number of neurons in the hidden layer, the number of hidden layers and the learning rate. In this process, 50 attempts were combined for each value placed in the architecture, with 1 execution each. The values in the hidden layers ranged from 25 to 200, with the addition of 25 neurons to each attempt in each layer.

When testing the different possible architectures, the values found were: 200 neurons in two hidden layer, and a learning rate of 0.001. The model was trained 500 epochs, using 20% of the training set to assess accuracy at each training season, using the Batch Normalization technique [45]. Weights were initialized using He's technique [46]. The activation function in the hidden layers was ReLU [47] and in the Softmax [48] output layer with two neurons, corresponding to the classes of the problem: dropout or not dropout. The Softmax function was used for returning the probability for each of the classes in the dropout model. A summary of the MLP architecture can be seen in the Figure 6.

As can be seen in Figure 6, the optimizer chosen was Adam [49] and the loss function was Sparse Categorical Cross Entropy [50], due to the nature of the data, with a large amount of dichotomous data. The work in [43] referred to these data as being mostly categorical. The metric for assessing the loss function was accuracy. As a result, in the training process the model returned 99% accuracy and in the test 97%. An 80/20 proportion of the data was used, in view of its order of magnitude. In the classification evaluation process, the confusion matrix was used (See Figure 7).

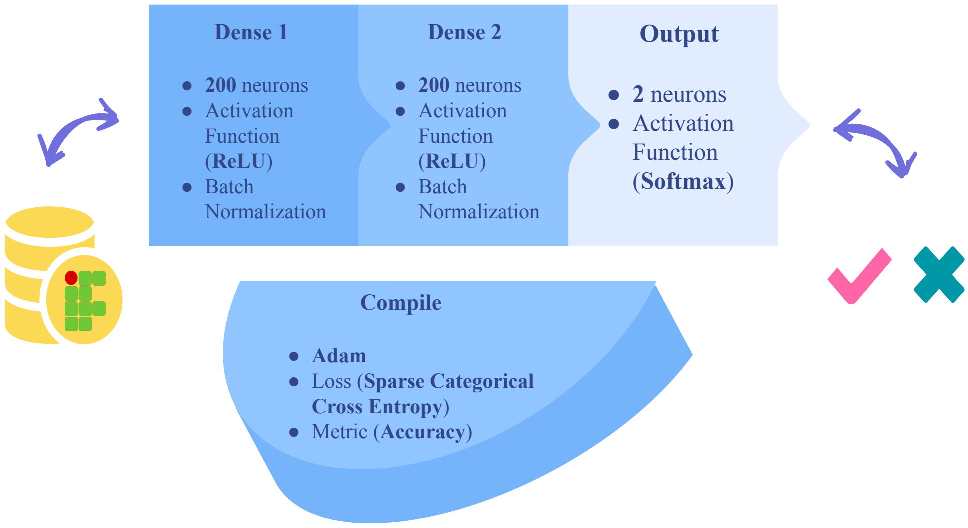

Figure 6. Defined MLP architecture.

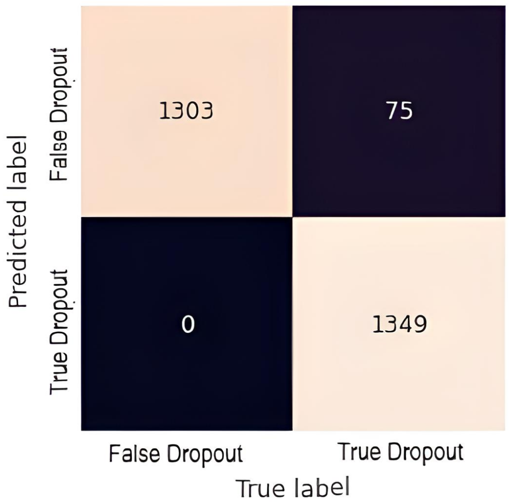

Figure 7. Confusion matrix for the MLP.

In the confusion matrix in Figure 7, a relatively small amount of false negatives was observed. A student in this case is in a dropout situation, but the model cannot identify her/him. According to our problem, this situation is one of the most complicated to deal with, considering that the model is built precisely to identify these students. These cases will be discussed in the results section, as well as students correctly classified as dropout and not dropout.

Using the implementation of the XAI frameworks for local and global explanations, the LIME [34], SHAP [35] and Shapley values [36] libraries were imported. It is worth mentioning that the objective of the work was not to build a black-box model, but based on its results, to apply the XAI frameworks presented in the reviewed works. We created Table 2 to make the pertinent inferences about the models based on existing literature about XAI for the education scenario.

Based on definitions in Table 2, an index was defined in order to make comparisons between the *explainers*, which we call the XAI explainability index:

$$IE_{XAI} = \frac{\sum_{i=1}^{14} M_{i_{satisfied}}}{14} \tag{9}$$

where each $M_{i_{satisfied}}$ is the metric satisfied divided by the total number of metrics (14). The IE_{XAI} was used to compare the best XAI model for the defined case study.

Table 2. Metrics by XAI framework for the school dropout problem.

Metric	Description	LIME	Shapley Values	SHAP
M1	Understand how the inputs are mathematically mapped to the outputs [17,18,21,33]		X	X
M2	View the characteristics of the parameters [9,19]	X	X	X
M3	Visualize interactions with the data set [9,19]			X
M4	Interactive views [9,19,20]			X
M5	Understand why one method is better than another using a common metric [18]			
M6	Understand what can be changed in a model so that the output is the desired one [18,20,21]			
M7	Show the importance of features [9,19]	X	X	X
M8	Show the weight of the features in each decision [9,19]	X	X	X
M9	Generates local post hoc explanations of black-box models [22]	X	X	X
M10	Generates global post hoc explanations of black-box models [22]			X
M11	Understand why unobserved events could have occurred [21]	X		
M12	The presentation of the explanation is simple [20,21,25]	X		X
M13	Show probabilities [20,21,25]	X		X
M14	Generalization of the AI model [20,21,25]	X		X
	IE_{XAI}	0.57	0.35	0.78

For organizational purposes, explanations given to the model were displayed in local and global explanations of the black box for each framework, according to their limitations, considering that some explainers were built for certain types of explanations.

5. Results ans Discussions

The explanations were processed and the defined explainability index (IE_{XAI}) was calculated for each explainer. For that, the explainers were analyzed according to the categories presented in Table 2.

The first-employed explainer was LIME [34]. Due to its nature, it was built for local explanations. Then, based on the analysis categories defined in Table 2, we first analyzed a situation where the model has guessed that a student is in a dropout situation, as depicted in Figure 8.

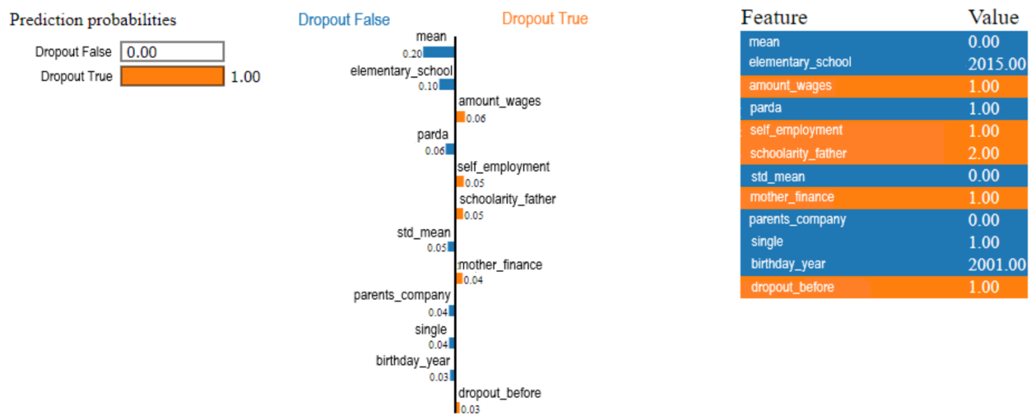

Figure 8. Dropout evaluation by LIME.

When considering the results in Figure 8, it was observed that the mean values of the year of elementary school completion, in addition to the student being of parda race (a person with different ethnic ancestry who is based on a mixture of skin colors between whites, blacks and indigenous people in Brazil, according to the Brazilian Institute of Geography and Statistics), had greater importance in the classification. It means that such a student is not in a dropout situation. Moreover, it was interpreted that some variables, such as the amount of minimum wages and the person responsible for the student being self-employed, contributed to the student being considered as in a dropout situation, as described by LIME in its explanation.

We were able to interpret that a student having a mean score of zero and being of parda race are characteristics that make the student not be in a group of dropouts, just as the amount of minimum wages of the student's family being equal to one is inclines the student towards the dropout group. These features together correspond to 47% of the importance value in the model and they were highlighted by the LIME explainer. Similarly, comments were made about the other classes of predictions in the MLP model with the LIME framework.

One fact to be noted is that the value of the parda race is not contributing to the student's prediction to be a potential dropout. This fact has already been pointed out in the PISA report [51], in which students who are black or parda race are more likely to perform poorly than white students, for example.

Observing the analysis metrics, we made some additional comparisons. The metrics satisfied with the LIME explainer were M2, M7, M8, M9, M11, M12, M13 and M14, as defined in Table 2.

In general, LIME was able to satisfy 8 of the 14 metrics established in the literature review, that is, it had an explainability index (IE_{XAI}) of 0.57.

Regarding the visualization of the characteristics of the parameters, it is possible to identify the value of each feature, such as the averages that influenced the prediction of students in situations of dropout and non-dropout. According to [9], this is a good parameter when choosing an AI model explainer.

The metrics M7 and M8 were selected due to the importance of each feature, as well as the weight of their values in each decision, as pointed out in [9,19]. For example, the average value was 20% important in the classification of the model, and the value corresponding to the average was equal to zero.

For M9, [22] argues about generating local explanations. These are intrinsic to the LIME explainer as to their nature of linearizing the model for local explanations based on sampling [34].

In a more interpretative perspective, M11 has the potential to be solved with the LIME explainer, considering that the importance weights of the features are presented, as well as their values. An argument in the explanation process that uses associative reasoning seeks to make unidentified associations due to the observed events, according to [21]. These associations were made through LIME, when we prospected the values returned by the model after it classified a student.

In relation to M12, M13 and M14 metrics, the values in the visualization given by the LIME explainer brought simplicity in the presentation and generalization of the model [20], particularly when it displayed lists of rules and a decision tree, pointed out by [16] as universal explainable models present in the literature.

The second applied explanation was Shapley values [36]. It is worth noting that these values are the basis of the SHAP library for local and global explanations [35]. Due to their nature, they are considered as local explanations based on global values, thus being an XAI framework. Similarly to the previous analysis, we first verified a situation when the model has guessed that a student is a dropout.

Due to the additive nature proposed by [36], the degree of importance of each feature is shown in the construction of the Shapley value that corresponds to the true class of the situation. In this case, a value equal to 1 means that the model was correct in the classification of the student in a dropout situation.

According to the visualization given by the explainer, Shapley values in Figure 9, it was observed that when the mean value equals to, the main responsibility of the student working autonomously and being female are among the main features that contribute so that the student can be classified as a dropout. Thus, the father having a level of education equal to 2 (elementary school incomplete) is a feature that subtracts from the corresponding Shapley Value.

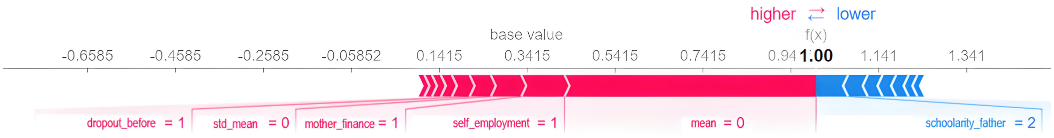

Figure 9. Case of dropout from Shapley Values.

From this information, we can see that students who have a mean score of zero are more likely to be considered dropouts by the model. In addition, the features considered important for the classification are similar to those used by the LIME explainer.

Considering the metrics for the evaluation of the explainer Shapley values, the satisfied metrics were M1, M2, M7, M8 and M9. Thus, the Shapley values applied to the problem are able to satisfy 5 of the 14 metrics established in the literature review, that is, an explainability index (IE_{XAI}) of 0.35.

Due to the very additive nature of the Shapley values [36], the model's input values are mapped to the output in such a way as to linearize the local forecast, providing an explanation of how the inputs are mapped mathematically to the outputs. In this case, the Shapley value is the explanation itself, satisfying M1 [17,18,21,33]. The Shapley values are post hoc local explanations, satisfying M9 [22].

In relation to M2, it is possible to visualize the values of each class of the features involved in the problem, such as the average values, if the student's financially responsible parent is the mother, among other characteristics [9]. Similarly, M7 and M8 are satisfied for showing the importance of features and their weight in each decision [9,19]. In this case, the weight is displayed by means of the size of the bar, not necessarily the additive or subtractive value in the corresponding Shapley value. Likewise, the bar emphasizes the importance of the feature, as well as whether its meaning is additive or subtractive for the explanation.

Finally, the third evaluated explainer is SHAP [35]. It uses the Shapley values [36] as its base and it seeks global explanations of the model. Then, in order to present a general explanation for the classification of the black box, some of the possible graphs were plotted using the available library, in order to identify the main contributions of the features for the classification of the school dropout prediction model.

With the plot of the contribution of each feature of the problem in Figure 10, we can pay attention to the importance of the average for the classification of both classes of the variable that indicates whether the student is in a situation of dropout or not. In addition, the year in which a student finished elementary school also deserves mention, the amount of wages the family earns per month, if the financial person in charge of the family is the mother, the percentage of school attendance during the year and if the student is of the parda race. These features were identified in the LIME and Shapley values explainers, and were confirmed with the SHAP explainer. It is worth mentioning that they are socioeconomic features, which are associated to students' predisposition to disengagement.

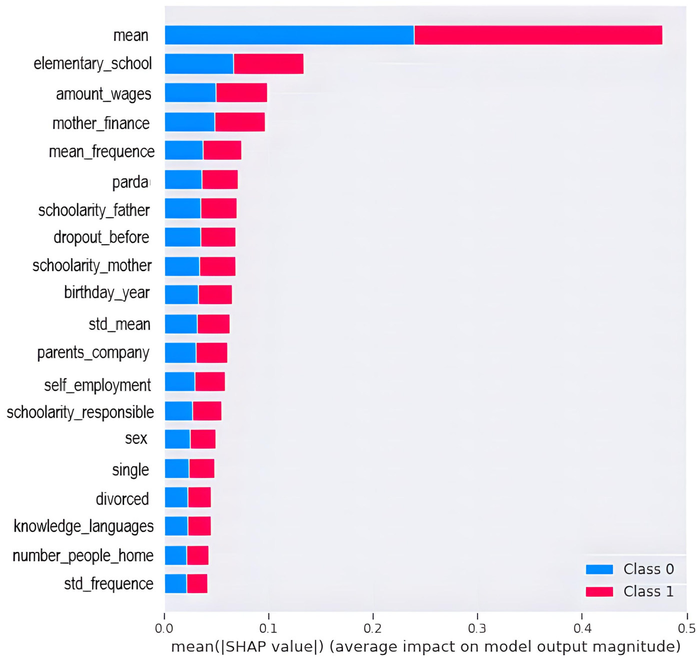

Figure 10. Contribution of the main features to the classification of school dropout in the black-box model—SHAP.

By observing the density of student data in relation to the main features that the black-box model took into account in the classification, according to the SHAP explainer, we were able to identify some characteristics about this data. Figure 11 depicts the achieved results.

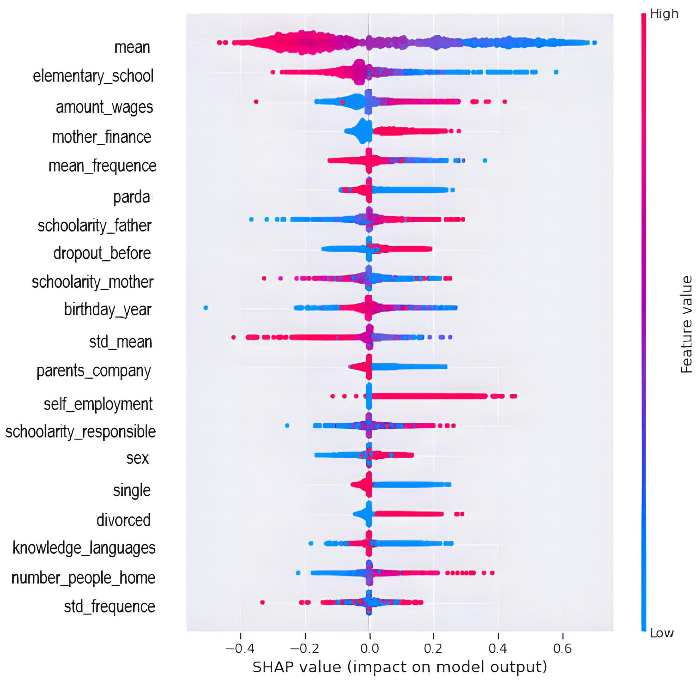

Figure 11. Density of student data dropout in relation to Shapley Values.

For dropout students, the highest mean values have positive contributions to the classification. Here, those with low means values are classified as dropout, with their subtractive contribution to the associated Shapley value. The features that identify the year of completion of elementary school, school attendance and race also have the same behavior.

In [39], the authors pointed out that although it is very difficult to isolate the impact of the various dimensions of poverty on student disengagement, insufficient income and parents' education tend to be the most important factors to influence school dropout prediction models.

In the case of the SHAP explainer, consulting the metrics and interpretations performed through the visualizations, we observed that the satisfied metrics were M1, M2, M3, M4, M7, M8, M9, M10, M12, M13 and M14.

In general, the SHAP explainer applied to the problem is able to satisfy 11 of the 14 metrics, with an explainability index (IE_{XAI}) of 0.78.

Regarding the metrics not shared with the Shapley values, the M3 is possible due to the interactive plotting of the SHAP explanation in relation to the Shapley values [9,19]. In this case, we were able to verify aspects regarding the distribution of the average values in relation to the education of the responsible students, identifying possible patterns according to the output Shapley value and the corresponding class. Due to interactivity, it also satisfies the M4 metric [9,19,20]. The fact that the instructor generates a global post hoc explanation of the black-box [22] model provided us with a more general view of how variables behave in certain situations.

Finally, the presentation of the explanation by the SHAP is simple, considering that the explanations are presented in the form of linear models [16]. The nature of the explanations is not exclusive, that is, each explainer brought contributions to the interpretation of the problem, with the values of school average, characteristics of education and work of the student's mother, and the amount of minimum wages earned by the family, being the main features that appeared in the explanations given by the explainers LIME [34], Shapley values [36] and SHAP [35].

Due to its large number of contributions to the implementation of the Shapley values, we found that the best explanation was the SHAP with IE_{XAI} of 0.78. It is worth noting that none of the evaluated works satisfied the M5 metric, in which an explanation should provide an understanding of why one method is better than another using a common metric [18], as well as the possibility of change in a model so that the output is the desired one, as M6 [18,20,21]. These are interesting metrics for the explanation to be complete, but they were not considered in those works.

6. Conclusions

This article presented valuable results and discussions when identifying features of students in school dropout situation, such as mean score, parental education and family income. The use of metrics and the proposed explainability index IE_{XAI} have shown to be adequate for the evaluation of black-box models regarding the explainability of each explainer. However, despite the achieved results and the identification of SHAP as the best option in the considered scenario, it had some limitations regarding the large computational expense, the calculation and plotting of global forecasts, sampling sensibility, and lack of scalability.

The performed analysis may help managers to identify important features for recognizing students in school dropout situation and to plan specific action strategies. In this sense, XAI approaches come as an important resource to implement solutions that have privacy issues and that require same level of model explicability. At this point, we believe that the achieved results can support in future research in this area.

As future works, it is intended the evaluation of new XAI libraries and new school datasets, in order to allow the generation of different local and global explanations. Overall, we believe that additional comparison results can be worth it, opening new research trends.

Author Contributions: Conceptualization, E.M. and I.S.; methodology, E.M. and I.S.; software, E.M. and I.S.; validation, E.M., T.M.B., D.G.C., C.M.D.V. and I.S.; formal analysis, E.M. and I.S.; investigation, E.M. and I.S.; resources, E.M. and I.S.; data curation, E.M. and T.M.B.; writing—original draft preparation, E.M., I.S. and D.G.C.; writing—review and editing, E.M., I.S., D.G.C. and C.M.D.V.; visualization, E.M. and I.S.; supervision, E.M. and I.S.; project administration, E.M. and I.S.; funding acquisition, E.M. and I.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Piscitello, J.; Kim, Y.K.; Orooji, M.; Robison, S. Sociodemographic risk, school engagement, and community characteristics: A mediated approach to understanding high school dropout. *Child. Youth Serv. Rev.* **2022**, *133*, 106347. [CrossRef]
2. Barros, T.M.; Souza Neto, P.A.; Silva, I.; Guedes, L.A. Predictive Models for Imbalanced Data: A School Dropout Perspective. *Educ. Sci.* **2019**, *9*, 275. [CrossRef]
3. Zawacki-Richter, O. Systematic review of research on artificial intelligence applications in higher education—Where are the educators? *Int. J. Educ. Technol. High Educ.* **2019**, *16*, 39. [CrossRef]
4. Chen, X.; Xie, H.; Zou, D.; Hwang, G.J. Application and theory gaps during the rise of artificial intelligence in education. *Comput. Educ. Artif. Intell.* **2020**, *1*, 100002. [CrossRef]
5. Cortiz, D. A narrative review of fairness and morality in neuroscience: Insights to artificial intelligence. *AI Ethics* **2022**, 1–12. [CrossRef]
6. Klockmann, V.; von Schenk, A.; Villeval, M.C. Artificial intelligence, ethics, and intergenerational responsibility. *J. Econ. Behav. Organ.* **2022**, *203*, 284–317. [CrossRef]
7. Polat, S. Reasons for school dropout in vocational high school. *Educ. Res. Rev.* **2014**, *9*, 711–718. [CrossRef]
8. Lent, M.V. An explainable artificial intelligence system for small-unit tactical behavior. In Proceedings of the National Conference on Artificial Intelligence, San Jose, CA, USA, 25–29 July 2004; pp. 900–907.
9. Raschka, S.; Patterson, J.; Nolet, C. Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information* **2020**, *11*, 193. [CrossRef]
10. Khosravi, H.; Shum, S.B.; Chen, G.; Conati, C.; Tsai, Y.S.; Kay, J.; Knight, S.; Martinez-Maldonado, R.; Sadiq, S.; Gašević, D. Explainable Artificial Intelligence in education. *Comput. Educ. Artif. Intell.* **2022**, *3*, 100074. [CrossRef]
11. Koh, P.W.; Liang, P. Understanding black-box predictions via influence functions. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 1885–1894.
12. Paleja, R.; Ghuy, M.; Ranawaka Arachchige, N.; Jensen, R.; Gombolay, M. The Utility of Explainable AI in Ad Hoc Human-Machine Teaming. In *Proceedings of the Advances in Neural Information Processing Systems*; Ranzato, M., Beygelzimer, A., Dauphin, Y., Liang, P., Vaughan, J.W., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2021; Volume 34, pp. 610–623.
13. Kabir, M.H.; Hasan, K.F.; Hasan, M.K.; Ansari, K. Explainable Artificial Intelligence for Smart City Application: A Secure and Trusted Platform. In *Explainable Artificial Intelligence for Cyber Security: Next Generation Artificial Intelligence*; Ahmed, M., Islam, S.R., Anwar, A., Moustafa, N., Pathan, A.S.K., Eds.; Springer International Publishing: Cham, Switzerland, 2022; pp. 241–263. [CrossRef]
14. Guidotti, R. A Survey of Methods for Explaining Black Box Models. *ACM Comput. Surv.* **2018**, *51*, 1–42. [CrossRef]
15. Adadi, A. Peeking inside the black-box: A survey on Explainable Artificial Intelligence (XAI). *IEEE Access* **2018**, *6*, 52138–52160 [CrossRef]
16. Kindermans, P.J.; Schütt, K.T.; Alber, M.; Müller, K.R.; Erhan, D.; Kim, B.; Dähne, S. Learning how to explain neural networks: PatternNet and patternattribution. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018; pp. 1–16. [CrossRef]
17. Montavon, G.; Samek, W.; Müller, K.R. Methods for interpreting and understanding deep neural networks. *Digit. Signal Process.* **2018**, *73*, 1–15. [CrossRef]
18. Doshi-Velez, F.; Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv* **2017**, arXiv:1702.08608.
19. Biran, O.; Cotton, C. Explanation and justification in machine learning: A survey. In Proceedings of the IJCAI-17 Workshop on Explainable AI (XAI), Melbourne, Australia, August 2017; pp. 8–13. Available online: http://www.cs.columbia.edu/~orb/papers/xai_survey_paper_2017.pdf (accessed on 10 November 2022).
20. Xu, F.; Uszkoreit, H.; Du, Y.; Fan, W.; Zhao, D.; Zhu, J. Explainable AI: A brief survey on history, research areas, approaches and challenges. In Proceedings of the CCF International Conference on Natural Language Processing and Chinese Computing, Dunhuang, China, 9–14 October 2019; Springer: Berlin/Heidelberg, Germany, 2019; pp. 563–574.

21. Miller, T.; Howe, P.; Sonenberg, L. Explainable AI: Beware of Inmates Running the Asylum. *arXiv* **2017**, arXiv:1712.00547
22. Molnar, C. *Interpretable Machine Learning*; Lulu. com: Morrisville, NC, USA, 2020.
23. Lombrozo, T. The structure and function of explanations. *Trends Cogn. Sci.* **2006**, *10*, 464–470. [CrossRef]
24. Malle, B.F. *How the Mind Explains Behavior*; Folk explanation, Meaning and social interaction; MIT-Press: Cambridge, MA, USA, 2004.
25. Thagard, P. Explanatory coherence. *Behav. Brain Sci.* **1989**, *12*, 435–467. [CrossRef]
26. Baker, R.; Siemens, G. Educational data mining and learning analytics. In *The Oxford Handbook of Innovation*; Sawyer, K., Ed.; Cambridge University Press: Cambridge, UK, 2014; pp. 253–274.
27. Silva, L.A. Ciência de Dados Educacionais: Definições e Convergências entre as Áreas de Pesquisa. Anais dos Workshops do VI Congresso Brasileiro de Informática na Educação (WCBIE). 2017. Available online: http://ojs.sector3.com.br/index.php/wcbie/article/view/7462 (accessed on 10 November 2022).
28. Williamson, B. *Big Data in Education: The Digital Future of Learning, Policy and Practice*; Sage: New York, NY, USA, 2017.
29. Maschio, P.T. Um Panorama Acerca da Mineração de Dados Educacionais no Brasil. Anais do XXIX Simpósio Brasileiro de Informática na Educação (SBIE). 2018. Available online: http://ojs.sector3.com.br/index.php/sbie/article/view/8194 (accessed on 10 November 2022).
30. Sciarrone, F.; Temperini, M. Learning Analytics Models: A Brief Review. In Proceedings of the 2019 23rd International Conference Information Visualisation (IV), Paris, France, 2–5 July 2019; pp. 287–291. [CrossRef]
31. Harris, J.; Mishra, P.; Koehler, M. Teachers' Technological Pedagogical Content Knowledge and Learning Activity Types: Curriculum-based Technology Integration Reframed. *J. Res. Technol. Educ.* **2009**, *41*, 393–416. [CrossRef]
32. Liñán, L.C.; Pérez, A.A.J. Educational Data Mining and Learning Analytics: Differences, similarities, and time evolution. *RUSC Univ. Knowl. Soc. J.* **2015**, *12*, 98–112. [CrossRef]
33. Dosilovic, F.K. Explainable Artificial Intelligence: A Survey. In Proceedings of the MIPRO, Opatija, Croatia, 21–25 May 2018.
34. Ribeiro, M.T.; Singh, S.; Guestrin, C. "Why Should I Trust You?": Explaining the Predictions of Any Classifier. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, 13–17 August 2016; pp. 1135–1144.
35. Lundberg, S.M.; Lee, S.I. A Unified Approach to Interpreting Model Predictions. In *Advances in Neural Information Processing Systems 30*; Guyon, I., Luxburg, U.V., Bengio, S., Wallach, H., Fergus, R., Vishwanathan, S., Garnett, R., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2017; pp. 4765–4774.
36. Shapley, L. A value for n-person games. In *Contributions to the Theory of Games*; Princeton University Press: Princeton, NJ, USA, 1953; Volume II, pp. 307–317.
37. Lundberg, S.M.; Lee, S.I. A unified approach to interpreting model predictions. *Adv. Neural Inf. Process. Syst.* **2017**, 30. Available online: https://proceedings.neurips.cc/paper/2017/hash/8a20a8621978632d76c43dfd28b67767-Abstract.html (accessed on 10 November 2022).
38. Korobov, M.; Lopuhin, K. ELI5. 2017. Available online: https://pypi.org/project/eli5/ (accessed on 10 November 2022).
39. Wilson, S.J.; Tanner-Smith, E.E.; Lipsey, M.W.; Steinka-Fry, K.; Morrison, J. Dropout prevention and intervention programs: Effects on school completion and dropout among school-aged children and youth. *Campbell Syst. Rev.* **2011**, *7*, 1–61. [CrossRef]
40. Gallego, M.G.; Perez de los Cobos, A.P.; Gallego, J.C.G. Identifying students at risk to academic dropout in higher education. *Educ. Sci.* **2021**, *11*, 427. [CrossRef]
41. Márquez-Vera, C.; Cano, A.; Romero, C.; Noaman, A.Y.M.; Mousa Fardoun, H.; Ventura, S. Early dropout prediction using data mining: A case study with high school students. *Expert Syst.* **2016**, *33*, 107–124. [CrossRef]
42. Chawla, N.V.; Bowyer, K.W.; Hall, L.O.; Kegelmeyer, W.P. SMOTE: Synthetic Minority Over-sampling Technique. *J. Artif. Intell. Res.* **2002**, *16*, 321–357. [CrossRef]
43. Barros, T.M.; Silva, I.; Guedes, L.A. Determination of Dropout Student Profile Based on Correspondence Analysis Technique. *IEEE Lat. Am. Trans.* **2019**, *17*, 1517–1523. [CrossRef]
44. O'Malley, T.; Bursztein, E.; Long, J.; Chollet, F.; Jin, H.; Invernizzi, L. Keras Tuner. 2019. Available online: https://github.com/keras-team/keras-tuner (accessed on 10 November 2022).
45. Ioffe, S.; Szegedy, C. Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift. In Proceedings of the 32nd International Conference on International Conference on Machine Learning—Volume 37. JMLR.org, ICML'15, Lille, France, 7–9 July 2015; pp. 448–456.
46. He, K.; Zhang, X.; Ren, S.; Sun, J. Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification. In Proceedings of the Proceedings of the IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
47. Agarap, A.F. Deep Learning Using Rectified Linear Units (ReLU). *arXiv* **2018**, arXiv:1803.08375.
48. Dukhan, M.; Ablavatski, A. The Two-Pass Softmax Algorithm. In Proceedings of the 2020 IEEE International Parallel and Distributed Processing Symposium Workshops (IPDPSW), New Orleans, LA, USA, 18–22 May 2020; pp. 386–395. [CrossRef]
49. Kingma, D.P.; Ba, J. Adam: A Method for Stochastic Optimization (2014). In Proceedings of the 3rd International Conference for Learning Representations, San Diego, CA, USA, 7–9 May 2015.

50. Liong, V.E.; Lu, J.; Tan, Y. Multi-Label Deep Sparse Hashing. In Proceedings of the 2018 IEEE Visual Communications and Image Processing (VCIP), Taichung, Taiwan, 9–12 December 2018; pp. 1–4. [CrossRef]
51. OCDE. *Results PISA 2018 (Programme for International Student Assessment)*; OCDE: Paris, France, 2019; Volume I–III, pp. 1–11.