Explainability in Intelligent Analysis of Students considering Dropout Factors: some insights

Wallyce Azy[1], Regina Braga[1], Victor Stroele[1], José Maria David[1], Fernanda Campos[1]

[1]Programa de Pós-Graduação em Ciência da Computação (PGCC/UFJF)
CEP 36036-900 – Juiz de Fora – MG – Brazil

wallyce.azy@ufjf.br, regina.braga@ufjf.br, victor.stroele@ufjf.br

jose.david@ufjf.br, fernanda.campos@ufjf.br

Abstract. *Context: In public institutions, there is great concern about the average number of students graduating at the end of undergraduate studies, which substantially impacts public higher education policies and public investments in Brazilian universities. Moreover, dropout imposes a financial and human burden, preventing students from learning. Problem: Brazil witnessed a university dropout rate of almost 55%. This problem affects society in general, which needs more suitably qualified professionals to face the challenges of the job market. Solution: This work aims to analyze, through AI explainability algorithms, the specific factors that lead to student dropout, considering specific courses from the great areas of science. We strive to explore the profile of students who have dropped out in recent years, stratified by course. Explainability algorithms allow the formal inspection of each factor that led to the dropout. Method: We used the Design Science Research methodology to conduct our study. An analysis with data from a specific university, considering the GDPR, was conducted to verify the proposal's feasibility. Results: Our results show that the solution can help identify key factors that lead to dropping out, stratified by areas, helping to provide specific actions to deal with this problem in universities.*

1. Introduction

Predicting student dropout rates in higher education has been one of the most discussed topics in academic literature in recent decades, especially with the advancement of machine learning (ML) techniques [Azy et al. 2024], [Silva et al. 2024]. The ability to identify students at risk of dropping out in advance allows higher education institutions to adopt more effective strategies for student retention.

In public institutions, there is great concern about the average number of students graduating at the end of undergraduate courses, which substantially impacts public higher education policies and public investments in Brazilian universities. Analyzing "when" and "why" students drop out is crucial for validating these policies. Many studies use artificial intelligence techniques to detect students at risk of dropping out [Silva et al. 2024, Assis and Marcolino 2024, da Silva 2021]. However, when machines completely make the decisions and humans are always at the receiving end, there is an urgent need to understand how those machines reach their conclusions. In this work, we are particularly interested in investigating the explainability of each factor's (variable in AI models) influence on dropouts, considering specific courses of a public institution.

DOI: 10.5753/sbie.2025.12698

We aim to analyze, through explainability algorithms, the variables that lead to student dropout from these courses. The study focuses on aspects not frequently analyzed, such as how students get into universities, including quotas and scholarships.

Explainable Artificial Intelligence (XAI) is crucial as AI models increasingly drive business decisions affecting many users. XAI provides methods to make AI outputs understandable, revealing why a model made a specific prediction, how its algorithms function, and identifying potential biases. This understanding empowers users to make informed recommendations and support decision-making. This is the case in this work, explaining why a certain student was classified as having a high probability of dropping out. [Mishra 2021] presents and implements a variety of XAI techniques — including SHAP (SHapley Additive exPlanations), LIME (Local Interpretable Model-agnostic Explanations), and Permutation Feature Importance — using Python-based libraries, offering practical guidance and code examples for interpreting both global and local behavior of machine learning models.

Similar to [Silva et al. 2024, Silva et al. 2025], this study builds on research into using XAI to identify factors that impact dropout. As a differential, our work uses semantic analysis for data preprocessing, which helps uncover missing data and implicit relationships. The semantic strategy was previously outlined in Federal University of Juiz de Fora (UFJF). Moreover, in this article, we create a timeline spanning the periods to verify which factors influence dropout from the beginning to the middle of the course. We explore applying XAI algorithms to identify these factors, focusing on developing strategies to support university student retention and advocating for financial resources for public universities. Our work focuses specifically on the pedagogy, civil engineering, and pharmacy courses, including a comparative analysis across these fields. Different strategies are needed for each area, tailored to their unique student profiles. This article, therefore, aims to provide an in-depth understanding of student dropout drivers. We investigate the following research question: *"How can explainability algorithms help identify the key factors of dropout and formulate effective public policies to combat it?"*. We conducted a case study using over 10 years of real-world data to validate our solution.

The article has six sections, including this introduction. Section 2 discusses related work. Section 3 presents the methodology. Section 4 details a case study, Section 5 presents the results, and Section 6 discusses the conclusions and future work.

2. Related Works

Several studies highlight the issue of university dropout, a persistent and significant challenge in the Brazilian educational context [Assis and Marcolino 2024]. To mitigate this issue, many research efforts utilize AI and ML algorithms to predict student dropout [Silva et al. 2024]. To ensure confidence in these predictions, it is crucial to provide a mechanism for analyzing the decisions generated by AI processing, thereby making models more transparent and interpretable. Although some models, such as Logistic Regression, are inherently interpretable, many of the most powerful and widely used algorithms, such as XGBoost, Random Forest, and deep neural networks, are considered "black boxes", making it difficult to understand the factors influencing their predictions [Mishra 2022]. The need for transparent and reliable AI models is particularly critical in sensitive and high-impact domains, as the case of education. Recent studies emphasize XAI's crucial role in promoting transparency, building trust, and enabling targeted interventions for students at risk of dropout [Liu et al. 2025]. Clarity on "why" a model predicts dropout for a given student can support more effective and personalized interventions by educational institutions. Recent projects in Latin America, including initiatives in Brazil, also leverage XAI to predict university dropout and inform retention policies [Silva et al. 2025].

We conducted a systematic literature mapping in May 2025, followed by snowballing, to understand how XAI enhances model explainability in machine learning, covering both naturally interpretable models and post-hoc methods for complex algorithms.

We aimed to identify key approaches, analyze their applications and limitations, and theoretically support XAI's use in higher education dropout prediction. Analysing the collected articles, some key points stood out: i) Models analyzed and techniques used: The studies addressed a variety of models, with an emphasis on deep neural networks (CNNs), XGBoost, Random Forest, and regression. Although more accurate, complex models reinforce the need for clear explanations, especially in sensitive contexts such as health and education. The most widely used techniques were SHAP, LIME, Grad-CAM, and Permutation Importance, with SHAP being the most common, notable for its balance between explanatory precision and generalization. ii) Application focus: Some studies focus on biomedical or safety applications, but all emphasize the critical role of user confidence in predictions, a factor equally vital in education. Understanding why a model predicts student dropout enables more targeted and effective interventions.iii) Contribution to this work: By proposing a predictive model for student dropout, our work integrates an intelligent process to predict students prone to dropout and an explanation of the variables (factors) that led to this conclusion. As discussed in the reviewed articles, techniques such as SHAP and LIME can be applied to the model trained with students' data to generate individualized explanations. This improvement enhances the model's reliability for institutional use, allowing decisions to be made based on transparent evidence. iv) Summary of gaps and potential contribution: Despite existing research, there's a need for more in-depth XAI applications in educational dropout prediction. Some studies detect dropout without explainability, while others limit XAI to a narrow range of indicators [Silva et al. 2025, Silva et al. 2024]. Our work aims to bridge this gap by practically applying XAI to university dropout, integrating institutional, demographic, and academic data with combined semantic [Azy et al. 2024] and machine learning techniques, which offer greater gains than using them in isolation [Silva et al. 2024]. This enables a more comprehensive analysis of dropout factors.

We propose a predictive model that identifies at-risk students and provides individualized explanations for their dropout likelihood. Additionally, we conduct a temporal analysis of dropout risk across semesters for students in three distinct undergraduate courses (humanities, exact sciences, and biological sciences) to identify common versus course-specific factors. Thus, our work stands out by i) applying explainability techniques in an unexplored domain (Brazilian Undergraduate education); ii) integrating institutional, demographic, and academic data, processed by semantic and ML algorithms; iii) utilizing individualized explanations to support institutional decisions on retention interventions and public policies regarding universities.

3. Methodology

This work extends previous research by our group [Azy et al. 2024], employing the Design Science Research (DSR) methodology [Peffers et al. 2007]. DSR involves iteratively developing, evaluating, and refining artifacts to solve practical problems. This new cycle extended the architecture to incorporate ML techniques and XAI algorithms. This allows us to interpret and analyze the key factors contributing to student dropout rates at a public university.

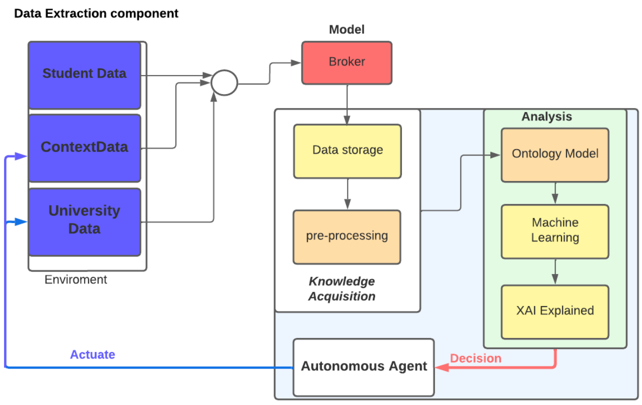

Figure 1. Architecture Main Components

The extended architecture (Figure 1) comprises three components: **Data Extraction**, which integrates different data types from educational sources; **Ontological model**, using the OWL 2.0 language, the ontological model represents the concepts involved in the student's academic trajectory, inferring implicit relationships and information to enrich the data semantically; and **Analysis**, responsible for processing the data previously enriched and performing analysis, predictions, and explainability processing.

In the previous work, the analysis performed to identify dropout factors was based on semantic rules, which, based on inference processing, identified some factors that led to dropout. Although the approach allowed for the discovery of factors, we did not conduct a more detailed investigation of the results, making it impossible to identify the paths that led to dropout, which is an important requirement to ensure the reliability of the results and the definition of strategies to mitigate dropout. Therefore, in this new DSR cycle, the analysis component has been remodeled to use ML algorithms and explainability techniques (XAI), allowing us to understand why the predictive model made its decisions. The architecture evaluation enables a more complete and transparent analysis of the factors contributing to dropout.

4. Case Study

Evaluation is fundamental to the Design Science Research (DSR) methodology. This evaluation aims to identify factors influencing undergraduate dropout from a public university. To verify the solution's feasibility, we conducted a study with data from students of public university courses, respecting the General Data Protection Regulation (GDPR). The analysis encompasses data from more than 10 years.

Scope Definition: We defined the scope of the evaluation using the GQM (Goal, Questions, Metrics) framework [Basili and Weiss 1984]. The **Goal** is to identify the factors influencing dropout, considering data from students from specific courses, in a public Brazilian university context. The Research **Question** (RQ) to be investigated in this evaluation is: *"How can explainability algorithms help identify the key factors of dropout and formulate effective public policies to combat it?"*. The **Metrics** used are accuracy, F1-score, area under the ROC curve for evaluation of ML models, and global explainability to understand the factors' impact on models' decisions.

Conduction: The data used in this study were obtained through the Fala.BR government platform. The data encompass three courses from the Federal University of Juiz de Fora (UFJF): Civil Engineering, Pharmacy and Pedagogy, encompassing exact sciences, biological sciences, and humanities areas, allowing a comparative analysis of dropout rates from the perspective of students from different areas and educational contexts. The data were provided anonymized at the source, containing only a random and untraceable identifier, to relate student data to their respective academic records. The time frame was defined between the years 2000 to 2024. No sensitive or personal information was included in the study, ensuring compliance with the General Data Protection Law (LGPD). Table 1 shows how the data were organized, and Table 2 shows the data distribution for each course, totaling approximately 150,000 records.

Table 1. Detailed Data of the Analyzed Students

Student	Academic	Scholarship and Academic Project
Anonymized identifier, Type of admission, Type of quota, Ethnicity, Gender, Municipality of the student, Status (completed or dropped out), Year of admission, Semester of admission, among others.	Anonymized identifier, Subject, Class semester, Class year, Subject code, Subject department, Grade, Status (Approved, Retake Grade, Retake Attendance), Class start date, Class end date.	Anonymized identifier, Scholarship (Name), Project (Name), Paid Indicator, Modality (If it refers to an academic project or student assistance), Start date of validity, End date of validity.

Table 2. Data distribution among the courses analyzed

Course	Student Data	Academic Data	Scholarship and Academic Project Data
Civil Engineering	876	59820	1904
Pharmacy	740	55019	3197
Pedagogy	594	25531	1435

Data Processing: The data underwent processing by Data Extraction and Ontological components for semantic enrichment. Predictive modeling was then performed to evaluate the ability of ML models to predict student dropout risk, using real historical data from students who either completed or dropped out of their courses. The primary objective was to present dropout scenarios and, crucially, to analyze the explainability of the decisions made by these models. We utilized Random Forest, Gradient Boosting, and XGBoost models due to their robustness with tabular data, compatibility with the explainability techniques employed, and their prevalence in the systematic review's findings. The target variable of the study is the student's status, which was categorized into two classes: 0 - completed and 1 - dropped out. The modeling considered data related to specific semesters (periods) to predict whether the student will drop out, using only the data up to the specific semester. The dataset was subdivided into 70% for model training and 30% for analysis of the results. 100 independent runs were performed with different partitionings of the data. The data was balanced in each run using the SMOTE (Synthetic Minority Oversampling Technique) technique, since the amount between completed and evaded students is unbalanced. The evaluation of the models considered the following metrics: Accuracy, macro F1-score, F1-score of the evasion class (F1_Pos), and area under the ROC curve (AUC-ROC). For each run, an optimal threshold was adjusted based on the precision-recall curve to maximize the F1-score of the evasion class. The model with the best performance in each repetition was stored with its metrics, predictions, and confusion matrix. Although the main objective of this step was to evaluate the predictive performance, the main contribution of this work is in the explainability of the models' decisions, performed through the SHAP and LIME techniques.

4.1. Explainability Techniques

The SHAP analysis was conducted using TreeExplainer, an optimized variation of the SHAP technique. This approach provides consistent and interpretable explanations and is suitable for educational scenarios where model transparency is essential [LUNDBERG; LEE, 2017]. TreeExplainer was applied to the models with the best results obtained in the 100 executions of the machine learning techniques, generating SHAP values from the balanced training data. The results were represented in graphs showing the attributes and the impacts on the model's decisions, generating a global interpretation of the factors influencing student dropout.

The LIME technique was also used to provide local explanations for specific instances. LIME simulates small variations in the input data attributes to analyze the individual impact on the prediction. With this, it is possible to infer which factors contributed to each instance classification, both as a tendency to drop out and a tendency to complete. This technique allows the identification of specific factors, providing insights for more personalized interventions, validating the coherence of decisions, and revealing recurring dropout patterns, which can support more effective retention policies.

We segmented the data by course and analyzed time frames of 2, 3, and 4 semesters to predict dropout risk using only the information available up to that point. We then trained and evaluated ML models for predictive performance for each semester. Next, we applied SHAP and LIME explainability techniques to the best-performing model from each semester to interpret the most relevant factors contributing to dropout prediction. SHAP provided global explanations, showing the average importance of variables across all students, while LIME offered local explanations, allowing us to analyze specific predictions for individual students.

We analyzed the results by semester to understand how the explainability factors for dropout change over time in each course. The generated graphs and files show the consistency of the most relevant variables and potential pattern shifts at different points in a student's academic journey. We used the positive F1-score (F1-Pos) as the primary metric to evaluate model effectiveness. This choice was deliberate because we focus on accurately identifying high-risk students (i.e., positive dropout cases). F1-Pos balances precision and sensitivity for the dropout class, preventing the model from achieving high accuracy while failing to effectively detect actual dropouts—a common issue with unbalanced datasets.

5. Results

Figure 2 displays the confusion matrices for the top-performing XGBoost, Random Forest, and Gradient Boosting models applied to the **Civil Engineering** course's **second-semester** data. These models were selected based on their F1_Pos scores. XGBoost emerged as the best performer with an F1_Pos of 0.7477, demonstrating superior sensitivity for the dropout class (fewer false negatives). Random Forest achieved a comparable F1_Pos of 0.7156, though with a slight decrease in its ability to correctly identify dropouts. Gradient Boosting recorded an F1_Pos of 0.7018, showing a balanced distribution but more classification errors for the positive class.

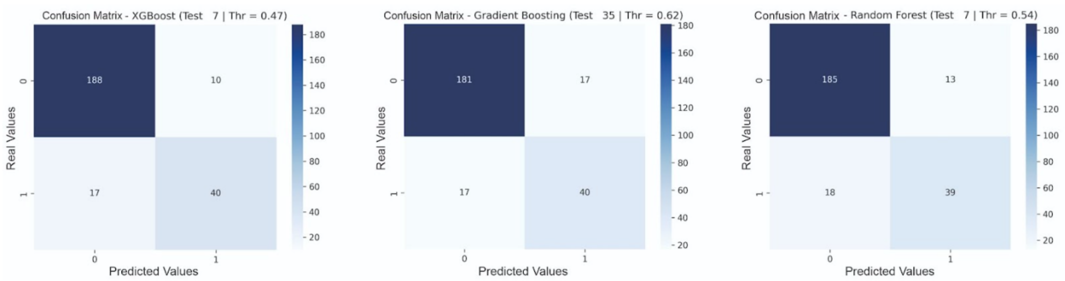

Figure 2. Confusion Matrices for the best XGBoost, Random Forest, and Gradient Boosting Models

Figures 3 and 4 present excerpts of the global explainability graphs generated by the SHAP technique, highlighting the attributes with the most significant impact on the models' decisions. All models consistently identify the number of subjects approved in the 2nd period as the most crucial factor in predicting dropout. Additionally, attributes like admission grade, admission via the serial evaluation system, general competition quota, and history of failures were among the primary predictors across all cases. These results indicate that the three models share a core set of variables influencing their decisions, emphasizing the student's academic performance and institutional history. This consistency across models reinforces the reliability of their predictions, despite their differing approaches. Beyond SHAP's global explainability, the LIME technique was also applied to provide local interpretations of these predictions. Figure 5 illustrates an example of a student classified as a dropout with a 99% probability. The primary factors driving this classification were a low number of approved subjects in the second semester and a lack of participation in student assistance programs. Conversely, the student's admission through the ENEM broad competition quota and their mixed-race ethnicity acted as mitigating factors, though insufficient to alter the dropout prediction.

With data up to the **third semester**, we analyzed how predictive capacity and influential factors evolved. The models maintained consistent performance, with minimal variation in F1_Pos values. Random Forest performed best (0.7045), followed by XGBoost (0.6897) and Gradient Boosting (0.6809). Despite increased data complexity, all models showed good ability to identify at-risk students. SHAP analysis reinforced the number of approved subjects in the 3rd period as the main predictive factor. Attributes previously identified as relevant, such as admission grade, admission type (serial), failures, and scholarship grants, also remained significant. The LIME technique demonstrated consistency with the global results, highlighting that key local factors are linked to accumulated academic performance and the presence or absence of student aid. This reinforces LIME's potential to guide targeted, individualized interventions.

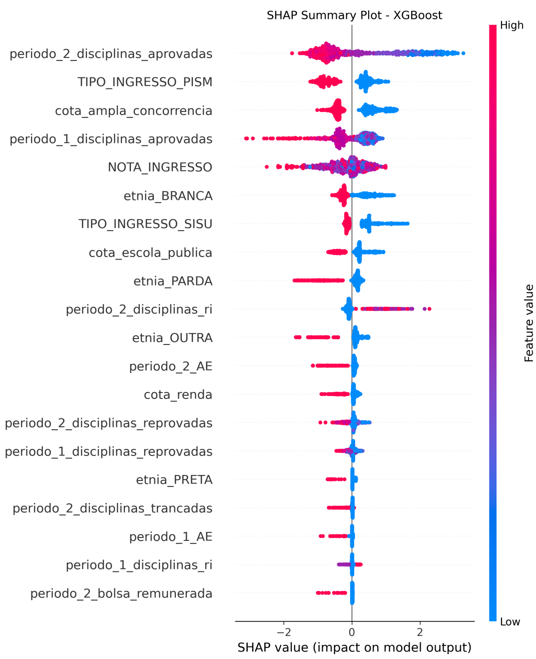

Figure 3. SHAP Summary Plot – XGBoost

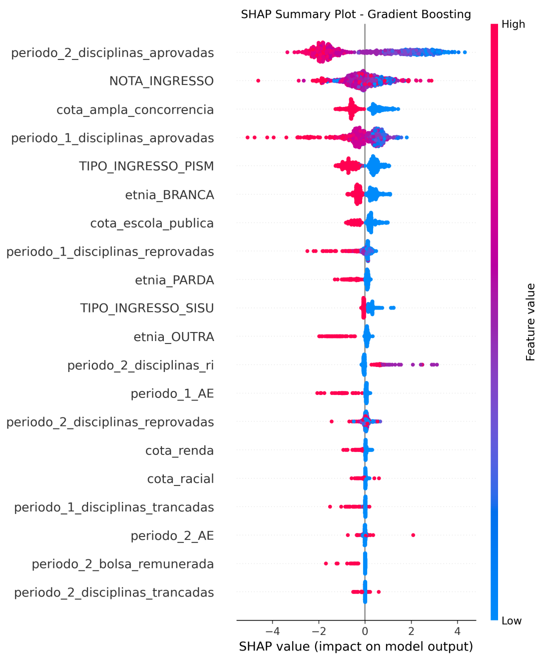

Figure 4. SHAP Summary Plot – Gradient Boosting

Figure 5. Example of LIME explanation – Civil Engineering (XGBoost), period 2

Next, we analyzed data accumulated up to the **fourth semester** to see if model accuracy continued to improve and if relevant attributes remained consistent. Random Forest again performed best (F1_Pos = 0.7778), showing superior dropout identification capabilities. XGBoost followed with an F1_Pos of 0.7218, and Gradient Boosting with

0.6744. All models maintained high overall accuracy (above 90%). The SHAP analysis revealed that the number of subjects approved in the 4th period became the most impactful variable. Models continued to assign relevance to cumulative performance variables (approvals and failures in the first three periods), admission grade, admission type (serial), and institutional quotas. These results indicate that as data becomes more robust, models increasingly emphasize performance histories as primary risk indicators, leading to greater consistency across models and increased confidence in predictions. Local explainability using LIME showed that, even for a student with good past approvals, a limited number of subjects approved in the fourth period, coupled with the absence of affirmative actions (like a general competition quota) and lack of scholarships, strongly influenced their classification as a dropout. This highlights that even with a strong academic history, a recent decline in performance can signal imminent dropout risk, underscoring the value of LIME's individualized analysis.

Similar analyses were conducted for the Pharmacy and Pedagogy courses, using data from periods 2, 3, and 4. Although the methodology applied was the same as in Civil Engineering, the results revealed distinct dropout patterns, with variations in the most relevant factors and predictive performance over time. Random Forest model produced the best results throughout all periods of the Pharmacy course, whereas XGBoost excelled in the Pedagogy course. This difference in performance between the models relates to the data features of each course, including the complexity of interactions among variables and the structure of dropout patterns over time.

Our analysis of the **Pharmacy course** reveals key factors influencing student retention, with academic performance consistently being paramount. In the **second semester**, subject approvals were the main factor for retention, showing a strong positive correlation in all SHAP analyses. Admission via serial exam evaluation and ENEM broad competition also significantly impacted retention. The admission grade showed a positive impact, even surpassing serial exam admission in the Gradient Boosting model. Students identifying as White were more likely to persist across all models, holding more weight than other ethnic or racial groups. Approvals from the first period also had a positive, though less significant, impact. Failures due to attendance issues and subject withdrawals were linked to dropout risk, but with less weight. Student assistance had a moderate effect across all models, with notable significance in the Random Forest model, where it outweighed current-period failures.

In the **third semester**, existing patterns persisted, with academic performance dominating as the primary factor. Current period approvals had the greatest impact across all models, followed by approvals from previous periods, underscoring the need for a consistent academic trajectory. Entrance grade (Serial exam or ENEM) was also a pertinent factor, ranking second in XGBoost and immediately after performance variables in Gradient Boosting. Admission via serial evaluation exam continued to positively impact all models, while ENEM broad competition also contributed, albeit less significantly. Failures due to lack of attendance and accumulated failures from previous periods remained correlated with dropout risk, indicating a pattern of academic deficiency over time. Student assistance gained increased relevance at this stage, particularly in the Gradient Boosting model, suggesting its protective effect intensifies as the course progresses.

By the **fourth semester**, admission by serial evaluation exam became more prominent among retention factors, appearing as the most impactful variable in the Random Forest model and second in XGBoost. Nevertheless, academic performance remained decisive: current period approvals were the main factor in Gradient Boosting and XGBoost, while cumulative approvals from previous periods also ranked highly across all three models. Though with lower weight, admission grade, self-declaration as White, and ENEM broad competition maintained a positive impact. Failures and withdrawals continued to be associated with dropout, but with reduced influence compared to retention factors. Student assistance also stood out, particularly in the Random Forest model, where receiving aid in the third and fourth periods had a significant impact. In the other models, only third-period assistance was among the main variables. Notably, none of the models listed variables related to scholarships (paid or unpaid) among the relevant factors, suggesting a limited impact of these specific actions at this stage of the academic journey.

Our analysis of the **Pedagogy course** reveals varying factors influencing student retention across different semesters, with academic performance consistently playing a central role. In the **second semester**, the number of approved subjects was the primary retention factor across all models. Admission methods, through serial evaluation exam and ENEM open competition, also showed significant influence. Interestingly, self-declaration as Black showed a stronger association with retention in the XGBoost model, while self-declaration as White held greater influence in the Gradient Boosting and Random Forest models. This divergence suggests more complex relationships between variables in Pedagogy, potentially explaining XGBoost's better performance with non-linear and heterogeneous data. Student assistance also emerged as a retention factor in the Random Forest and Gradient Boosting models, highlighting the importance of early institutional support in this course. Additionally, Gradient Boosting pointed to failures in the current period as the main dropout indicator, while Random Forest and XGBoost emphasized the number of subjects failed as a risk sign.

The pattern of academic performance continued into the **third semester**, with approvals in the current and previous periods remaining the dominant retention factors across all models. Admission by serial evaluation exam was relevant in Random Forest and XGBoost. In the Gradient Boosting model, self-declaration as White and admission grade had the most impact. While higher admission grades generally correlate with retention, it's not the sole determinant. Student assistance variables were present in all three models, showing a relevant positive impact, especially in Gradient Boosting and Random Forest for both current and previous periods. In XGBoost, only third-period assistance was a main factor. Paid scholarships also showed an association with retention, with moderate weight in Random Forest, but lower weight in XGBoost.

By the **fourth semester**, academic performance continued as the primary retention factor. Approvals in the third semester had the greatest impact in Gradient Boosting and Random Forest, while approvals in the fourth semester were most impactful in XGBoost, reinforcing the importance of recent performance. Entry by serial evaluation exam remained relevant across all models, with the highest weight in Random Forest. Student assistance also played a significant role, appearing among the top influential variables in Gradient Boosting and Random Forest, and present with less prominently in XGBoost. This suggests that financial support has a more pronounced positive effect from the intermediate stages of this course onwards. Notably, self-declaration as White re-emerged as a retention factor, with greater weight compared to other ethnic-racial groups, particularly in Gradient Boosting and XGBoost. This recurrence underscores the need to address racial disparities in dropout, even in courses like Pedagogy, which are typically associated with greater diversity. All graphs derived from these analyses are available at [1].

5.1. Discussion

The analysis of the three courses revealed that while academic performance is a crucial overarching factor, dropout profiles vary significantly across different fields. Across all models and periods, the number of subject approvals in the current period consistently emerged as the primary factor associated with retention. This highlights the centrality of immediate academic progress in a student's trajectory and indicates that consistent progression is key to preventing higher education dropout, regardless of the course or model. The SHAP algorithms were essential for accurately understanding the individual weight of each variable in these diverse contexts. Our use of tree-based models like XGBoost, Random Forest, and Gradient Boosting showed that, despite sharing a core of relevant variables, each model displayed different sensitivities to certain factors. For instance, in Pedagogy, XGBoost was more sensitive to variables related to ethnic and social profiles, while Gradient Boosting emphasized the impact of failures more strongly. This divergence in sensitivity was particularly pronounced in the Pedagogy course, which exhibited greater dispersion in relevant factors among models compared to the high convergence seen in Civil Engineering and Pharmacy.

Entry via a serial evaluation exam positively impacted all three courses, proving most significant in Civil Engineering and Pharmacy. This suggests that this admission method may aid student adaptation, fostering more stable academic journeys. However, retention policy variables showed varied impacts. Student assistance, for example, had a limited effect in Civil Engineering, appeared sporadically in Pharmacy, but gained consistent importance in Pedagogy from the second period onward, remaining relevant through the fourth semester. Similarly, paid scholarships only impacted the Pedagogy course, particularly in later periods, and were irrelevant in the other courses. Unpaid scholarships held no significant weight across any analyzed contexts. A key finding is that no variable related to racial, social, or disability quotas appeared among the main factors in our models. While ENEM's open competition showed some positive weight, specific quota types (e.g., racial, income, public school, and disabled) were not selected by the models as directly impacting dropout or retention. This doesn't invalidate the importance of affirmative action, but suggests that other variables, such as academic performance and institutional support, may mediate its effects. These differences highlight that student profiles and course dynamics influence retention. In Civil Engineering, models indicated that even academically strong students are at risk if their current performance drops, underscoring the need for continuous monitoring and intervention. In Pharmacy, retention was also strongly linked to performance, suggesting a need for enhanced academic monitoring in early periods. For Pedagogy, beyond performance, strengthening student assistance and paid scholarship policies is crucial, given these students' greater sensitivity to institutional support.

By analyzing these results, we can answer our research question: *"How can explainability algorithms help identify the key factors of dropout and formulate effective*

[1] [Azy et al. 2024]

public policies to combat it?". Our findings demonstrate that while academic performance is a universal and critical factor for student retention across all courses and models, dropout profiles significantly differ between major areas of science. This highlights that universities should invest in specific, stratified policies tailored to individual courses and broader scientific areas. A detailed analysis, supported by explainability algorithms like SHAP, reveals that targeted actions can be taken by examining these stratified data. Acknowledging that grades and admission via serial evaluation exams consistently appear as important factors for general student retention policies in all analyzed courses is crucial. However, the need for course-specific policies is equally vital. It's important to emphasize that these data pertain to specific courses at a Brazilian public university. More comprehensive analyses are needed, encompassing the broader scenario of Brazilian public universities and the full diversity of courses across all major research areas. Nevertheless, this study underscores the immense value of individualized analyses in designing more effective public policies for student retention, recognizing and addressing the diverse profiles of graduates within each scientific field.

6. Conclusions

This paper explored an explainable ML approach to analyze higher education dropout, using data from Civil Engineering, Pharmacy, and Pedagogy courses at a Brazilian public university. By combining XGBoost, Random Forest, and Gradient Boosting with SHAP and LIME, we transparently identified key factors influencing student retention or dropout during the initial semesters. Explainability was crucial to move beyond the "black box" of AI in education: SHAP offered a global view of the most influential factors, while LIME provided local interpretations for individualized interventions. This enabled both accurate predictions and the extraction of reliable knowledge to support pedagogical and institutional decisions. The main contribution lies in showing that predictive models coupled with explainability generate accurate classifications and actionable insights, guiding more effective retention policies tailored to each course.

We plan to expand the analysis to other courses and universities and include new data attributes. A future perspective is the "Observatório da Retenção no Ensino Superior Brasileiro," a platform providing free analyses with public data and personalized ones with private datasets. These findings may lead to tools for academic management that proactively alert about dropout risks and propose tailored actions. A promising direction is applying local explainability (LIME) to assess individual students, empowering institutions with targeted interventions and strengthening retention through transparent analyses.

Acknowledgments

Profs. Victor Ströele and Regina Braga thank the support from the Brazilian National Council for Scientific and Technological Development (CNPq) for the research productivity grants 313568/2023-5 (PQ-2) and 307194/2022-1 (PQ-2), respectively.

References

Adachi, A. (2009). Evasion and dropouts from ufmg undergraduate courses. In Portuguese. Pós em Educação - UFMG, Minas Gerais.

Aguirre, C. E. and Pérez, J. C. (2020). Predictive data analysis techniques applied to dropping out of university studies. In *2020 XLVI Latin American Computing Conference (CLEI)*, pages 512–521. IEEE.

Ajoodha, R., Dukhan, S., and Jadhav, A. (2020). Data-driven student support for academic success by developing student skill profiles. In *2020 2nd International Multidisciplinary Information Technology and Engineering Conference (IMITEC)*, pages 1–8. IEEE.

Aquines Gutiérrez, O., Hernández Taylor, D. M., Santos-Guevara, A., Chavarría-Garza, W. X., Martínez-Huerta, H., and Galloway, R. K. (2022). How the entry profiles and early study habits are related to first-year academic performance in engineering programs. *Sustainability*, 14(22):15400.

Assis, M. V. O. and Marcolino, A. S. (2024). A predictive model for dropout risk in a computer science education program. In *Simpósio Brasileiro de Informática na Educação (SBIE), 35*, pages 1560–1573, Rio de Janeiro/RJ. Sociedade Brasileira de Computação.

Aulck, L. et al. (2019). Mining university registrar records to predict first-year undergraduate attrition. pages 9–18. No publisher listed.

Azy, W., Braga, R., Stroele, V., David, J. M. N., Campos, F., Chaves, L. J., and Campos, L. (2024). Intelligent analysis of students profile about dropout factors: A study in information system course context. In *Anais do XXXV Simpósio Brasileiro de Informática na Educação (SBIE)*, Rio de Janeiro, RJ. Sociedade Brasileira de Computação.

Baranyi, M., Nagy, M., and Molontay, R. (2020). Interpretable deep learning for university dropout prediction. pages 13–19. Cited by: 57.

Bardagi, M. and Hutz, C. (2014). University dropout and student support services: a brief review of brazilian literature. *Psicologia Revista*. In Portuguese. Accessed: Aug. 2023.

Basili, V. R. and Weiss, D. M. (1984). A methodology for collecting valid software engineering data. *IEEE Transactions on Software Engineering*, (6):728–738.

Boulic, R. and Renault, O. (1991). 3d hierarchies for animation. In Magnenat-Thalmann, N. and Thalmann, D., editors, *New Trends in Animation and Visualization*. John Wiley & Sons ltd.

Cunha, A., Tunes, E., and Da Silva, R. (2001). Evasion from the chemistry course at the university of brasília: The interpretation of the evoked student. *Química Nova*. In Portuguese. Accessed: Aug. 2023.

da Cruz, R. C., Juliano, R. C., Monteiro Souza, F. C., and Correa Souza, A. C. (2023). A score approach to identify the risk of students dropout: an experiment with information systems course. In *Proceedings of the XIX Brazilian Symposium on Information Systems*, pages 120–127.

da Silva, C. (2021). A holistic profile ontology for undergraduate students. In Portuguese. Accessed: Aug. 2023.

da Silva, L. M., Dias, L. P. S., Rigo, S., Barbosa, J. L. V., Leithardt, D. R., and Leithardt, V. R. Q. (2021). A literature review on intelligent services applied to distance learning. *Education Sciences*, 11(11):666.

de Oliveira, P., da Silva, G., Dourado, R., and Rodrigues, R. L. (2021). Linking engagement profiles to academic performance through sna and cluster analysis on discussion forum data. In *LALA*, pages 39–47.

Educação, D. (2022). Evasion breaks records in higher education. In Portuguese. Accessed: Aug. 2023.

El-Rady, A. A. (2020). An ontological model to predict dropout students using machine learning techniques.

Federal, G. General data protection law (lgpd) - brazil. effective date. Available at: https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm.

Fernández-López, M., Gómez-Pérez, A., and Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. In *Spring Symposium Series*. Facultad de Informática (UPM).

Gonzalez-Nucamendi, A., Noguez, J., Neri, L., Robledo-Rella, V., García-Castelán, R. M. G., and Escobar-Castillejos, D. (2022). Learning analytics to determine profile dimensions of students associated with their academic performance. *Applied Sciences*, 12(20):10560.

Gruber, T. R. (1993). A translation approach to portable ontology specifications.

INEP (2017). Methodology for calculating higher education flow indicators. In Portuguese.

Knuth, D. E. (1984). *The TEX Book*. Addison-Wesley, 15th edition.

Kochkach, A., Kacem, S. B., Elkosantini, S., Lee, S. M., and Suh, W. (2024). On the different concepts and taxonomies of explainable artificial intelligence. In Bennour, A., Bouridane, A., and Chaari, L., editors, *Intelligent Systems and Pattern Recognition. ISPR 2023*, volume 1941 of *Communications in Computer and Information Science*. Springer, Cham.

Lakkaraju, H., Aguiar, E., Shan, C., Miller, D., Bhanpuri, N., Ghani, R., and Addison, K. L. (2015). A machine learning framework to identify students at risk of adverse academic outcomes. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '15)*.

Liu, B., Li, C., and Wan, Z. (2025). Using explainable ai (xai) to identify and intervene with students in need: A review. In *Proceedings of the 2024 3rd International Conference on Artificial Intelligence and Education (ICAIE '24)*, pages 636–641, New York, NY, USA. Association for Computing Machinery.

Menolli, A., Horita, F., Dias, J. J. L., and Coelho, R. (2020). Bi–based methodology for analyzing higher education: A case study of dropout phenomenon in information systems courses. In *XVI Brazilian Symposium on Information Systems*, pages 1–8.

Mishra, P. (2021). *Practical explainable AI using Python: artificial intelligence model explanations using Python-based libraries, extensions, and frameworks*. Apress, Berkeley, CA.

Mishra, P. (2022). *Practical explainable AI using python: Artificial intelligence model explanations using python-based libraries, extensions, and frameworks*. Apress.

Mourão, E. et al. (2020). On the performance of hybrid search strategies for systematic literature reviews in software engineering. *Information and Software Technology*, 123:106294.

Ozkaya, I. (2023). Application of large language models to software engineering tasks: Opportunities, risks, and implications. *IEEE Software*, 40(3):4–8.

Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3):45–77.

Priyambada, S. A., Er, M., Yahya, B. N., and Usagawa, T. (2021). Profile-based cluster evolution analysis: Identification of migration patterns for understanding student learning behavior. *IEEE Access*, 9:101718–101728.

Ram, S. et al. (2015). Using big data for predicting freshmen retention. No publisher listed.

Runeson, P. and Höst, M. (2009). Guidelines for conducting and reporting case study research in software engineering. *Empirical Software Engineering*, 14:131–164.

Saito, T. and Rehmsmeier, M. (2015). The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLoS ONE*, 10(3):e0118432.

Saqr, M., López-Pernas, S., Helske, S., and Hrastinski, S. (2023). The longitudinal association between engagement and achievement varies by time, students' profiles, and achievement state: A full program study. *Computers & Education*, 199:104787.

SEMESP (2024a). Mapa do ensino superior, 14ª edição / 2024. Accessed on: 7 Dec. 2024.

SEMESP (2024b). Mapa do ensino superior no brasil. Accessed on: 5 Jun. 2024.

Senthil Kumaran, V. and Malar, B. (2023). Distributed ensemble based iterative classification for churn analysis and prediction of dropout ratio in e-learning. *Interactive Learning Environments*, 31(7):4235–4250.

Sharma, P. (2023). Utilizing explainable artificial intelligence to address deep learning in biomedical domain. In *Medical data analysis and processing using explainable artificial intelligence*, pages 19–38. CRC Press.

Shevskaya, N. V. (2021). Explainable artificial intelligence approaches: Challenges and perspectives. In *2021 International Conference on Quality Management, Transport and Information Security, Information Technologies (IT&QM&IS)*. IEEE.

Silva, F. d. C., Feitosa, R. M., Batista, L. A., and Santana, A. M. (2024). Análise comparativa de métodos de explicabilidade da inteligência artificial no cenário educacional: um estudo de caso sobre evasão. In *Simpósio Brasileiro de Informática na Educação (SBIE), 35*, pages 2968–2977, Rio de Janeiro/RJ. Sociedade Brasileira de Computação.

Silva, F. d. C., Santana, A. M., and Feitosa, R. M. (2025). An investigation into dropout indicators in secondary technical education using explainable artificial intelligence. *IEEE Revista Iberoamericana de Tecnologias del Aprendizaje*, 20:105–114.

Smith, A. and Jones, B. (1999). On the complexity of computing. In Smith-Jones, A. B., editor, *Advances in Computer Science*, pages 555–566. Publishing Press.

Vasanth, S., Keerthana, S., and Saravanan, G. (2024). Demystifying ai: A robust and comprehensive approach to explainable ai. In *2024 International Conference on Intelligent Computing and Emerging Communication Technologies (ICEC)*, pages 1–5. IEEE.

Veloso, T. and De Almeida, E. (2024). Evasion in ufmt undergraduate courses. 24ª Reunião Anual - Technical Report. In Portuguese.

Vinker, E. and Rubinstein, A. (2022). Mining code submissions to elucidate disengagement in a computer science mooc. In *LAK22: 12th International Learning Analytics and Knowledge Conference*, pages 142–151.

World Wide Web Consortium (2012). Owl 2 web ontology language document overview. W3C Recommendation.