Universidade Federal de Juiz de Fora
PGCC
Pós-graduação em Ciência da Computação

EducAAr: Integrating Ontologies, Machine Learning, and XAI for Academic Dropout Analysis in Higher Education Institutions

Wallyce Fellipe Oscar Azy

JUIZ DE FORA
MAIO, 2026

EducAAr: Integrating Ontologies, Machine Learning, and XAI for Academic Dropout Analysis in Higher Education Institutions

Wallyce Fellipe Oscar Azy

Universidade Federal de Juiz de Fora

PGCC

PGCC - Pós-Graduação em Ciência da Computação

Mestrado em Ciência da Computação

Orientador: Regina Maria Maciel Braga

Coorientador: Victor Stroële de Andrade Menezes

JUIZ DE FORA

MAIO, 2026

Ficha catalográfica elaborada através do programa de geração
automática da Biblioteca Universitária da UFJF,
com os dados fornecidos pelo(a) autor(a)

OSCAR AZY, WALLYCE FELLIPE.
EducAAr: Integrating Ontologies, Machine Learning, and XAI for Academic Dropout Analysis in Higher Education Institutions / WALLYCE FELLIPE OSCAR AZY. -- 2026.
123 p. : il.

Orientador: Regina Maria Maciel Braga
Coorientador: Victor Stroële de Andrade Menezes
Dissertação (mestrado acadêmico) - Universidade Federal de Juiz de Fora, Instituto de Ciências Exatas. Programa de Pós-Graduação em Ciência da Computação, 2026.

1. student dropout. 2. higher education. 3. machine learning. 4. explainability . 5. ontology. I. Maciel Braga, Regina Maria, orient. II. de Andrade Menezes, Victor Stroële, coorient. III. Título.

Wallyce Fellipe Oscar Azy

EducAAr: Integrating Ontologies, Machine Learning, and XAI for Academic Dropout Analysis in Higher Education Institutions

Dissertação apresentada ao Programa de Pós-graduação em Ciência da Computação da Universidade Federal de Juiz de Fora como requisito parcial à obtenção do título de Mestre em Ciência da Computação. Área de concentração: Ciência da Computação.

Aprovada em 20 de maio de 2026.

BANCA EXAMINADORA

Profª. Dra. Regina Maria Maciel Braga Villela - Orientadora

Universidade Federal de Juiz de Fora

Prof. Dr. Victor Ströele de Andrade Menezes - Coorientador

Universidade Federal de Juiz de Fora

Prof. Dr. José Maria Nazar David

Universidade Federal de Juiz de Fora

Profª. Dra. Cláudia Maria Lima Werner

Universidade Federal do Rio de Janeiro

Profª. Dra. Patricia Augustin Jaques Maillard

Universidade Federal do Paraná

Juiz de Fora, 25/05/2026.

Documento assinado eletronicamente por **Patricia Augustin Jaques Maillard**, **Usuário Externo**, em 25/05/2026, às 14:51, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Claudia Maria Lima Werner**, **Usuário Externo**, em 25/05/2026, às 15:36, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Regina Maria Maciel Braga Villela**, **Professor(a)**, em 26/05/2026, às 12:00, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Jose Maria Nazar David**, **Professor(a)**, em 26/05/2026, às 12:19, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

Documento assinado eletronicamente por **Victor Stroele de Andrade Menezes**, **Professor(a)**, em 26/05/2026, às 15:21, conforme horário oficial de Brasília, com fundamento no § 3º do art. 4º do Decreto nº 10.543, de 13 de novembro de 2020.

A autenticidade deste documento pode ser conferida no Portal do SEI-Ufjf (www2.ufjf.br/SEI) através do ícone Conferência de Documentos, informando o código verificador **3000401** e o código CRC **EBF648B7**.

EducAAr: Integrating Ontologies, Machine Learning, and XAI for Academic Dropout Analysis in Higher Education Institutions

Wallyce Fellipe Oscar Azy

DISSERTAÇÃO SUBMETIDA AO CORPO DOCENTE DO PGCC DA UNIVERSIDADE FEDERAL DE JUIZ DE FORA, COMO PARTE INTEGRANTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIA DA COMPUTAÇÃO.

Aprovada por:

Regina Maria Maciel Braga
Doutora em Engenharia de Sistemas e Computação

Victor Stroële de Andrade Menezes
Doutor em Engenharia de Sistemas e Computação

José Maria Nazar David
Doutor em Engenharia de Sistemas e Computação

Claudia Maria Lima Werner
Doutora em Engenharia de Sistemas e Computação

Patricia Augustin Jaques Maillard
Doutora em Ciência da Computação

JUIZ DE FORA

22 DE MAIO, 2026

Dedico este trabalho, primeiramente, a Deus, por me permitir chegar até aqui.

À minha família, pelo apoio durante toda essa caminhada.

E a todos que contribuíram, direta ou indiretamente, para que este momento fosse possível.

Resumo

A evasão estudantil no ensino superior é um problema complexo, com impactos acadêmicos, institucionais e sociais, o que torna relevante o desenvolvimento de abordagens capazes de apoiar sua análise. Neste contexto, esta dissertação propõe a arquitetura EducAAr (*Educational Analysis Architecture*), que articula um modelo canônico baseado em ontologia, modelos de aprendizado de máquina e técnicas de explicabilidade para produzir análises interpretáveis do risco de evasão estudantil. A pesquisa foi conduzida com base na abordagem *Design Science Research*, em dois ciclos de desenvolvimento. No primeiro ciclo, foi construída uma ontologia para integrar dados acadêmicos, sociais e institucionais, permitindo a organização das informações, a verificação de consistência e a realização de consultas sobre a trajetória estudantil. No segundo ciclo, a arquitetura foi expandida com modelos de aprendizado de máquina, calibração probabilística, explicabilidade baseada em SHAP e um painel de visualização dos resultados. A avaliação técnica foi realizada com dados anonimizados de graduação da Universidade Federal de Juiz de Fora, envolvendo 16 ofertas de cursos e 7.731 estudantes. Os resultados indicam que a arquitetura permite organizar dados educacionais heterogêneos, gerar estimativas de risco de evasão e apresentar os principais fatores associados a essas estimativas de forma interpretável. As explicações produzidas indicaram predominância de variáveis relacionadas ao desempenho acadêmico nos períodos iniciais, embora fatores de ingresso, assistência estudantil, bolsas, cotas e perfil discente também tenham aparecido em contextos específicos. Assim, a contribuição da dissertação está na proposição e avaliação técnica de uma arquitetura integrada para análise da evasão, sem substituir a interpretação institucional e sem tratar as associações identificadas como relações causais.

Palavras-chave: evasão estudantil, ensino superior, ontologia, aprendizado de máquina, explicabilidade.

Abstract

Student dropout in higher education is a complex problem with academic, institutional, and social impacts, which makes the development of approaches capable of supporting its analysis relevant. In this context, this dissertation proposes EducAAr (*Educational Analysis Architecture*), an architecture that articulates an ontology-based canonical model, machine learning models, and explainability techniques to produce interpretable analyses of student dropout risk. The research was conducted using the Design Science Research approach across two development cycles. In the first cycle, an ontology was developed to integrate academic, social, and institutional data, enabling information organization, consistency verification, and queries about student trajectories. In the second cycle, the architecture was extended with machine learning models, probabilistic calibration, SHAP-based explainability, and a visualization panel for the results. The technical evaluation was conducted using anonymized undergraduate data from the Federal University of Juiz de Fora, encompassing 16 course offerings and 7,731 students. The results indicate that the architecture can organize heterogeneous educational data, generate dropout risk estimates, and present the main factors associated with these estimates in an interpretable manner. The generated explanations indicated a predominance of variables related to academic performance in the initial periods, although admission factors, student assistance, scholarships, quotas, and student profile also appeared in specific contexts. Thus, the contribution of this dissertation lies in the proposal and technical evaluation of an integrated architecture for dropout analysis, without replacing institutional interpretation and without treating the identified associations as causal relationships.

Keywords: student dropout, higher education, ontology, machine learning, explainability.

Agradecimentos

A todos os meus parentes, pelo encorajamento e apoio.

À professora Regina Maria Maciel Braga e ao professor Victor Stroële de Andrade Menezes, pela orientação, amizade e, principalmente, pela paciência, sem a qual este trabalho não se realizaria.

Aos professores do Departamento de Ciência da Computação, pelos seus ensinamentos, e aos funcionários do curso, que, durante esses anos, contribuíram de algum modo para o meu enriquecimento pessoal e profissional.

"A minha graça te basta, porque o meu poder se aperfeiçoa na fraqueza".

2 Coríntios 12:9

Contents

List of Figures

List of Tables

List of Abbreviations

DCC	Department of Computer Science
DSR	Design Science Research
ML	Machine Learning
PGCC	Postgraduate Program in Computer Science
SHAP	SHapley Additive exPlanations
UFJF	Federal University of Juiz de Fora
XAI	Explainable Artificial Intelligence

1 Introduction

Student dropout remains a relevant problem in Brazilian higher education, with academic, institutional, and social impacts. In 2023, the country recorded 9,977,217 enrollments in higher education, which highlights the scale of the system and the strategic importance of understanding the factors that affect student retention. Despite this volume, the net schooling rate of the population aged 18 to 24, that is, the proportion of young people in this age group effectively enrolled in higher education, was 19.9%, indicating that access to higher education remains limited compared to national expansion goals (Instituto Semesp, 2025).

Beyond the challenge of expanding access, retention remains a critical issue. Considering the 2019 to 2023 cycle, the cumulative dropout rate in Brazilian higher education reached 58.3%. Even in public institutions, the indicator reached 41.9%, indicating that the problem is not confined to a single institutional segment. These numbers indicate that enrollment expansion alone does not determine students' academic trajectory, since admission does not guarantee program continuity or completion (Instituto Semesp, 2025).

In recent years, different studies have sought to anticipate dropout risk with the support of machine learning techniques (PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023). In general, this literature primarily uses academic and institutional data, such as grades, course failures, enrollment history, course progression, and attendance, sometimes combined with demographic and socioeconomic variables. These results show that dropout prediction already constitutes a consolidated line of investigation in the field of educational analysis.

However, in an institutional context, prediction alone is not sufficient. Knowing that a student presents a high probability of dropout is only part of the problem. It is also necessary to understand which factors support this estimate and how these factors can be presented in an interpretable manner to support academic and institutional analyses. Although recent literature already presents advances in this direction, works that

articulate, within a single framework, data organization, predictive modeling, explainability, and presentation of results for institutional use are still less frequent (BETTAHI; BELOUADHA; HARROUD, 2025; LUNDBERG; LEE, 2017).

Another aspect that hinders this process is the organization of educational data itself. Academic, registration, social, and institutional information is usually distributed across different databases, systems, and formats, which compromises its integration, standardization, and analytical reuse. As a consequence, even when data are available, they are not always structured in an adequate manner to support consistent analyses of student trajectories and dropout risk.

This dissertation is situated in this context. The work proposes EducAAr (*Educational Analysis Architecture*), an architecture developed to integrate educational data related to student dropout, predictive analysis, explainability, and visualization of results within a single structured workflow. The proposal is based on the idea that an architecture grounded in a canonical data model can support not only the integration of heterogeneous information, but also the production of interpretable analyses of dropout risk. In this dissertation, this support is understood as technical and analytical, since the architecture was not evaluated as an intervention tool with managers or academic teams.

The research was conducted using the Design Science Research (DSR) approach, since it is a methodology oriented toward the construction and evaluation of artifacts aimed at solving real problems (HEVNER, 2007; PIMENTEL et al., 2020). In the first cycle, EducAAr was developed with a focus on the organization and integration of data through an ontology, functioning as a canonical model for representing educational information. In the second cycle, the architecture was extended to incorporate machine learning models and explainability mechanisms, with the objective of producing dropout risk estimates accompanied by interpretive elements that can support institutional analyses.

Although the proposal was conceived for the higher education context more broadly, its evaluation was conducted using anonymized institutional data from the Federal University of Juiz de Fora (UFJF), a federal public university. In the second DSR cycle, the analysis involved 16 undergraduate course offerings, totaling 7,731 students in

the considered sample. This context allowed evaluating the architecture in a concrete institutional situation, without restricting its application to this single institution.

1.1 Research Question and Objectives

The research question that guides this dissertation is the following:

> *How can an architecture based on a canonical model for educational data integration combine machine learning, explainability, and visualization mechanisms to produce interpretable analyses of student dropout risk in higher education?*

Based on this Research Question, the general objective of this work is to develop and technically evaluate the EducAAr architecture, which integrates an ontology-based canonical model, machine learning models, explainability techniques, and a visualization panel to support interpretable analyses of student dropout risk in higher education.

More specifically, the work seeks to:

1. Build an ontology as a canonical model to integrate educational data from different institutional sources, allowing the representation, querying, and consistency verification of the information used in the analyses;

2. Transform the integrated educational data into structured datasets suitable for predictive modeling, preserving relevant information about academic trajectory, admission, student profile, quotas, scholarships, and student assistance;

3. Incorporate and technically evaluate machine learning models to estimate student dropout risk based on academic and institutional data;

4. Apply explainability techniques to identify the factors most associated with the risk estimates produced by the models;

5. Organize predictive metrics, explanatory results, and simulated student profiles in a visualization panel, making the outputs of the architecture more accessible for analysis;

6. Analyze how the integration between ontology, predictive modeling, explainability, and visualization contributes to a structured architectural workflow for dropout risk analysis in higher education.

1.2 Contribution

The main contribution of this dissertation is the proposal and technical evaluation of EducAAr as an integrated architecture for dropout risk analysis in higher education. The contribution is not restricted to the construction of predictive models, but lies in the articulation of four complementary components: i) the organization of heterogeneous educational data through an ontology used as a canonical model; ii) the transformation of the integrated data into structured datasets suitable for predictive modeling; iii) the estimation of dropout risk through calibrated machine learning models; and iv) the interpretation and presentation of the results through SHAP-based explainability and a visualization panel.

From a scientific perspective, the dissertation contributes by showing how these components can be combined in a coherent architectural workflow for the analysis of student dropout risk. The work also contributes by making explicit the relationship between data integration, prediction, explanation, and result visualization, which are often treated separately in the literature.

In this dissertation, EducAAr is evaluated as a technical artifact. Therefore, the architecture should not be understood as an automatic decision-making mechanism or as a tool whose institutional adoption was validated with managers or academic teams. Its contribution is to provide a structured and technically evaluated basis for producing interpretable analyses of dropout risk, while preserving the need for institutional interpretation and avoiding causal readings of the associations identified by the models.

1.3 Organization

In addition to this introduction, this dissertation is organized as follows. Chapter 2 presents a systematic literature review structured around two axes: techniques and models

applied to student dropout prediction, and the use of explainability techniques in predictive models. Chapter 3 describes the adopted methodology, the foundations of DSR, the architecture requirements, and the two development cycles of EducAAr. The evaluation chapter presents the results from the two cycles, including validation of the ontological modeling, predictive analysis, interpretation of the explanatory results, and evaluation of the visualization panel as part of the architecture. Finally, the closing chapter discusses the work's main contributions, its limitations, and future work.

2 Systematic Literature Review

Student dropout is a recurring problem in higher education and brings academic, social, and financial impacts. The 15th edition of the *Higher Education Map in Brazil* indicates that, in the 2019 to 2023 cycle, the cumulative dropout rate in higher education was 58.3% (Instituto Semesp, 2025). Therefore, predicting dropout in advance can contribute to the implementation of more targeted actions, with potential to reduce abandonment.

To understand how the literature has approached this topic, a systematic review was conducted, structured around two axes. The first axis addresses the techniques and models used in the prediction of student dropout. The second axis analyzes the use of *Explainable Artificial Intelligence* (XAI) techniques associated with predictive models, focusing on the interpretation of results and on supporting institutional analyses.

An evidence-based systematic review adopts a planned and transparent process to locate, select, evaluate, and synthesize studies on a specific topic (KITCHENHAM; BRERETON, 2013). In this dissertation, the hybrid approach proposed by (MOURÃO; OLIVEIRA; FIGUEIREDO, 2020) was adopted, combining searches in digital databases with *Backward* and *Forward Snowballing* strategies.

In this approach, the search in a digital database provides the initial set of studies selected from predefined strings and explicit inclusion and exclusion criteria. Snowballing is then used to expand this set by examining both the references cited by the selected studies and the later works that cite them. Thus, the review does not depend exclusively on the search string or on the ranking returned by the database.

This strategy was adopted because the topic of this dissertation involves two related but not identical dimensions: dropout prediction and explainability in machine learning models. By combining database search and snowballing, the review sought to reduce the risk of overlooking relevant studies and to obtain a broader view of models, data sources, metrics, explainability techniques, and methodological gaps.

The review was organized into three stages, which will be described in the following sections:

1. Planning;

2. Conduction;

3. Results and discussion.

Before presenting the planning of the review, it is important to clarify the theoretical basis that guides the reading of the selected studies. In this dissertation, student dropout is understood as a multifactorial phenomenon, associated with academic, institutional, social, and contextual dimensions. Therefore, the analysis of dropout risk cannot be reduced to the isolated observation of a single variable or to the predictive result alone.

The first theoretical element considered in this work is the organization of educational data. Institutional information related to students is commonly distributed across different systems, files, and formats, such as academic history, admission data, scholarships, student assistance, quotas, and student profile information. For this reason, the use of a canonical model becomes relevant, since it allows heterogeneous data to be represented in a common structure before the predictive stage.

The second element is predictive modeling. In the context of this dissertation, machine learning models are used to estimate dropout risk from structured educational data. These models are not treated as autonomous decision-making mechanisms, but as computational resources capable of identifying patterns associated with dropout and completion. Since dropout datasets may present class imbalance and different behavior across programs and periods, their evaluation requires complementary metrics, including measures focused on the positive class, discrimination capacity, precision-recall behavior, and probabilistic calibration.

The third element is explainability. A risk estimate alone is insufficient for institutional analysis if it is not accompanied by information about the factors that contributed to that estimate. Explainable Artificial Intelligence is therefore considered in this work as a necessary component for interpreting the behavior of predictive models, especially when the analysis involves academic, admission, assistance, quota, gender, and ethnicity variables.

Thus, the systematic review presented in this chapter is not limited to identifying

predictive techniques. It also seeks to understand how the literature deals with data, models, metrics, explanations, and limitations, providing the theoretical and methodological basis for the construction of EducAAr.

2.1 Planning

The Scopus database was chosen because it gathers a significant volume of peer-reviewed publications and offers search, filtering, and citation analysis features. Since this dissertation deals with two related topics with distinct objectives, the protocol was applied separately to each axis, with its own research questions, search *strings*, and selection criteria.

Axis 1 – Techniques and models applied to student dropout prediction

Table 2.1: Mapping of research questions for Axis 1 and their objectives

Mapping Questions	Objectives
RQ1.1: Which ML techniques and statistical models are most used in the prediction of dropout in higher education?	Identify which models are most recurrent in the literature, such as logistic regression, decision trees, boosting methods, neural networks, and classifier ensembles, observing usage trends and the justifications presented in the studies.
RQ1.2: Which types and sources of data are employed in these models?	Map the types of data used, such as academic, administrative, social, and behavioral information, observing their origin and availability.
RQ1.3: Which validation strategies and metrics are applied to evaluate predictive performance?	Analyze the validation strategies used in the experiments and the most recurrent metrics, and analyze their adequacy for the problem of student dropout.

Continued on next page

Table 2.1 – continued

Mapping Questions	Objectives
RQ1.4: Which limitations and gaps remain in existing approaches?	Identify the main difficulties, including generalization, reproducibility, comparison across institutions, and data standardization.

Axis 2 – Use of *Explainable Artificial Intelligence* (XAI) techniques associated with predictive models

Table 2.2: Mapping of research questions for Axis 2 and their objectives

Mapping Questions	Objectives
RQ2.1: What are the main explainability techniques applied to ML models?	Identify the most cited XAI methods, such as SHAP, LIME, PFI, and PDP/ICE, and in which contexts they are used.
RQ2.2: What are the advantages and limitations of these techniques?	Analyze the strengths and weaknesses of the main approaches, considering factors such as clarity, computational cost, stability, and practical usefulness.
RQ2.3: How can explainability support institutional analysis in the context of student dropout?	Investigate how studies use explanations to support institutional analyses.

Search and Selection Protocol

The protocol was applied separately for each axis, with its own search *strings* and selection criteria.

Axis 1 – Student Dropout Prediction

In Axis 1, the objective was to identify the models used to predict student dropout, the types of data used, the most frequent used evaluation metrics, and the main methodological limitations. The search *string* was built based on four central dimensions: educational

context, dropout, prediction, and analytical technique.

1. Educational context: ("higher education" OR "college" OR "university");
2. Student dropout: ("dropout" OR "retention" OR "attrition");
3. Prediction: ("prediction" OR "forecasting" OR "modeling");
4. Analytical technique: ("machine learning" OR "statistical models").

The final *string* was:

```
("higher education" OR "college" OR "university")
AND ("dropout" OR "retention" OR "attrition")
AND ("prediction" OR "forecasting" OR "modeling")
AND ("machine learning" OR "statistical models")
```

Inclusion Criteria (Axis 1):

- Articles published between 2015 and 2025;
- Studies related to dropout prediction in higher education;
- Use of *Machine Learning* techniques or predictive statistical models;
- Methodological description sufficient for analysis.

Exclusion Criteria (Axis 1):

- Studies focused on other educational levels;
- Works without detailing of the predictive process;
- Duplicate studies or non-systematic reviews.

Axis 2 – Explainability (XAI)

In Axis 2, the aim was to identify the most used XAI techniques, their advantages and limitations, and how they have been used to support interpretation and analysis in applied contexts. The adopted *string* was:

```
("machine learning" OR "artificial intelligence")
AND ("explainability" OR "explainable artificial intelligence"
OR "XAI" OR "interpretability" OR "model explanation")
```

Inclusion Criteria (Axis 2):

- Publications between 2015 and 2025;
- Studies with central focus on the application or evaluation of XAI techniques;
- Application to *Machine Learning* models.

Exclusion Criteria (Axis 2):

- Works without adequate description of the technique used;
- Superficial mentions of explainability, without practical application;
- Duplicate studies or non-systematic reviews.

2.2 Conduction

The selection process followed a sequential flow. First, the search strings were applied in Scopus according to each axis. Then, the retrieved studies were screened by title and abstract, and the inclusion and exclusion criteria were applied. Since the review records were consolidated from the set effectively selected after screening and criteria application, this dissertation reports the selection flow from this stage onward.

After this process, 25 studies retrieved from Scopus were selected for Axis 1 and 20 studies for Axis 2. Subsequently, *Backward* and *Forward Snowballing* were applied. This step added 10 studies to Axis 1 and did not add new studies to Axis 2. Therefore, the final set analyzed in the review consisted of 35 studies in Axis 1 and 20 studies in Axis 2.

Table 2.3: Summary of the selection flow adopted in the systematic review

Stage	**Axis 1**	**Axis 2**
Studies selected from Scopus after screening and criteria application	25	20
Additional studies included through *Backward* and *Forward Snowballing*	10	0
Final set of studies analyzed	35	20

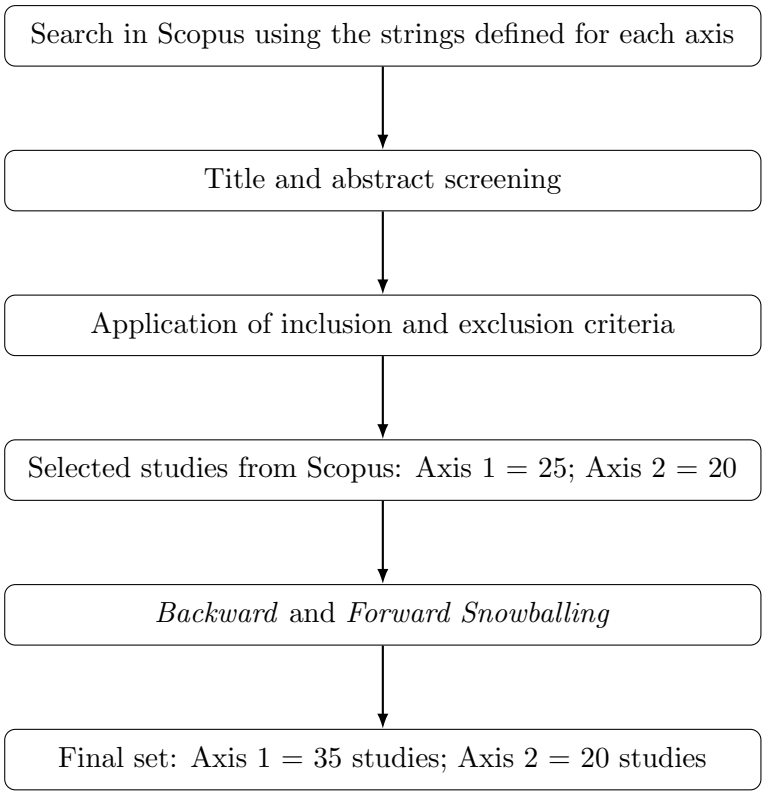

Figure 2.1: Selection flow of the systematic literature review

Complementary verification in Learning Analytics and EDM venues. In addition to the main selection flow, a complementary verification was conducted in publication venues strongly associated with Learning Analytics and Educational Data Mining, especially the Learning Analytics and Knowledge Conference (LAK) and the International Conference on Educational Data Mining (EDM). This verification was not treated as a new systematic search and did not change the final number of studies included in each axis. Its purpose was to reduce the risk of overlooking relevant studies published in venues directly related to the dissertation topic.

This complementary verification identified three studies used to contextualize and reinforce the discussion. In the context of dropout prediction, (MANRIQUE et al., 2019)

analyzed student representation, representative features, and classification algorithms for degree dropout prediction. In the context of explainability, (GUNASEKARA; SAARELA, 2024) synthesized studies on explainability in Educational Data Mining and Learning Analytics. Finally, (MURESAN; CARDEI; CARDEI, 2025) explored heterogeneous graph deep learning models for student success prediction, indicating that graph-based representations are an emerging direction in educational prediction tasks.

2.2.1 Axis 1 – Student dropout prediction

Source and date of the search. The search was conducted on Scopus on March 31, 2025. The results were initially filtered based on title and abstract. After applying the inclusion and exclusion criteria, 25 studies retrieved directly from the database were selected to compose the analysis of this axis.

Subsequently, *Backward* and *Forward Snowballing* strategies were applied, without circular return to the seed studies, with the objective of broadening coverage and reducing dependence on the highest-ranked results in the database. This process resulted in the inclusion of 10 additional studies. Thus, Axis 1 gathered 35 studies in total: 25 selected directly from Scopus and 10 added through snowballing. Table 2.4 presents the 25 studies selected directly from Scopus.

Table 2.4: Studies selected directly from Scopus for Axis 1 (25) and main ML techniques employed

Study Title	Year	Techniques Used
Predicting student dropouts with machine learning: An empirical study in Finnish higher education (VAARMA; LI, 2024)	2024	LR, RF, SVM (comparative)
Temporal and Between-Group Variability in College Dropout Prediction (GLANDORF et al., 2024)	2024	Supervised models; between-group/time analyses
Predicting Student Retention in Higher Education Using Machine Learning (SALLOUM et al., 2024)	2024	Random Forest (OvR), grid search

Continued on next page

Table 2.4 – continued

Study Title	Year	Techniques Used
Forecasting Student Attrition Using Machine Learning (PRAJWAL; SAHANA; KANCHANA, 2024)	2024	RF, SVM, KNN
Dropout Prediction of University Students in Bangladesh using Machine Learning (AKTER et al., 2024)	2024	RF, SVM, KNN (institutional set)
Predictive Modeling and Explainability for Academic Dropout Risk Detection Using Machine Learning (CASTRO; GARCIA; PELAEZ, 2025)	2025	RF/GBM + XAI (SHAP/LIME)
Crossing individual university boundaries: a comprehensive approach to predicting dropouts... (BERENS et al., 2025)	2025	Multi-institutional modeling; LR/trees
A Modular and Explainable Machine Learning Pipeline for Student Dropout Prediction... (BETTAHI; BELOUADHA; HARROUD, 2025)	2025	Modular pipeline; ensembles + XAI (SHAP)
Analyzing College Student Dropout Risk Prediction in Real Data Using Walk-Forward Validation (SANTOS; PONTI; RODRIGUES, 2023)	2023	Walk-forward; RF/XGB
Prediction of student attrition risk using machine learning (BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022)	2022	LR, RF, SVM
Towards a Students' Dropout Prediction Model in Higher Education Institutions Using ML (OQAIDI; AOUHASSI; MANSOURI, 2022)	2022	RF, SVM, KNN
Knowledge discovery for higher education student retention... (PALACIOS et al., 2021)	2021	C4.5, RF, SVM

Continued on next page

Table 2.4 – continued

Study Title	Year	Techniques Used
Predicting First-Year Computer Science Students Drop-Out with ML Methods (MAKSIMOVA; PENTEL; DUNAJEVA, 2021)	2021	LR, RF, SVM
Using ML methods to identify significant variables for... first-year Informatics Engineering dropout (BELLO et al., 2020)	2020	Variable selection + DT/RF
Interpretable Deep Learning for University Dropout Prediction (BARANYI; NAGY; MOLONTAY, 2020)	2020	FCNN, XGBoost; SHAP
Forecasting Learner Attrition for Student Success at a South African University (AJOODHA; JADHAV; DUKHAN, 2020)	2020	LR, RF, SVM
Model for the Prediction of Dropout in Higher Education in Peru... (JIMÉNEZ; JESÚS; WONG, 2023)	2023	RF, DT, NN, SVM
Predictive Model to Identify College Students with High Dropout Rates (OSORIO; SANTACOLOMA, 2023)	2023	LR/RF/SVM
Explaining Factors of Student Attrition at Higher Education (ALCAUTER; MARTÍNEZ-VILLASEÑOR; PONCE, 2023)	2023	LR/trees
Machine Learning in Higher Education: Predicting and Mitigating Student Dropout (ABOUELNOUR et al., 2024)	2024	Supervised classifiers
Human vs Machine Learning: Best Approach to Early Detect University Dropout Rates (AGUAYO-MAURI et al., 2025)	2025	Human vs. ML comparison

Continued on next page

Table 2.4 – continued

Study Title	Year	Techniques Used
Predicting Student Performance and Academic Success... Hybrid XGBoost-LSTM (VEMULAPALLI et al., 2025)	2025	XGBoost+LSTM
Supporting minority student success by using machine learning to identify at-risk students (JAYARAMAN; GERBER; GARCIA, 2019)	2019	Supervised classifiers
Early dropout prediction in distance higher education using active learning (KOSTOPOULOS et al., 2017)	2017	Active learning + classifiers
Early prediction of college attrition using data mining (MARTINS et al., 2017)	2017	DT, RF, SVM

Snowballing and circularity control. *Snowballing* was applied based on seed studies with methodological relevance or recent prominence, such as (BARANYI; NAGY; MOLONTAY, 2020; PALACIOS et al., 2021; VAARMA; LI, 2024). The procedure allowed adding classic and recent studies, studies without a circular return to the original works themselves, using DOI and title verification with the aid of the Connected Papers tool (Connected Papers, 2026).

2.2.2 Axis 2 – Explainability (XAI)

The Axis 2 search was conducted on Scopus on November 6, 2025. After title and abstract screening and the application of the defined criteria, 20 studies obtained directly from the database composed the set analyzed in this axis (Table 2.5).

Table 2.5: Seed studies of Axis 2 (20) and focus on XAI techniques

Study Title (citation)	Year	XAI Techniques (emphasis)
Explainable Artificial Intelligence (XAI) Approaches in Predictive Maintenance: A Review (SHARMA et al., 2024)	2024	Review; SHAP, LIME; PDP/PI
Interpretability and Transparency of ML in File Fragment Analysis with XAI (JINAD; ISLAM; SHASHIDHAR, 2024)	2024	SHAP and LIME
XAI: Conception, Visualization and Assessment Approaches Towards Amenable XAI (NIZAM; ZAFAR, 2023)	2023	Taxonomy; visualization; evaluation
XAI in the Veterinary and Animal Sciences Field (AQIB et al., 2023)	2023	SHAP, LIME; counterfactuals
Exploring the Landscape of XAI: A SLR of Techniques and Applications (HAMIDA et al., 2024)	2024	SHAP, LIME, Grad-CAM, counterfactuals
Can surgeons trust AI? Perspectives on ML in surgery and the importance of XAI (BRANDENBURG et al., 2025)	2025	Integrated Gradients; saliency
Effectiveness of XAI Techniques for Improving Human Trust... (WIRATSIN; RAGKHITWETSAGUL, 2025)	2025	SHAP, LIME, Grad-CAM
XAI: A Systematic Literature Review on Taxonomies and Applications in Finance (MARTINS et al., 2024)	2024	SHAP, LIME, PDP/ICE, PFI
Explainable Artificial Intelligence (XAI) in auditing (ZHANG; CHO; VASARHELYI, 2022)	2022	SHAP, LIME
SLR: Integration of XAI in IDS (cybersecurity) (MOHALE; OBAGBUWA, 2025)	2025	SHAP, LIME, rules, counterfactuals
Utilizing XAI to Address Deep Learning in the Biomedical Domain (SHARMA, 2023)	2023	Grad-CAM, LIME, counterfactuals

Continued on next page

Table 2.5 – continued

Study Title (citation)	Year	XAI Techniques (emphasis)
Open and Extensible Benchmark for XAI Methods (MOISEEV; BALABAEVA; KOVALCHUK, 2025)	2025	Benchmark of XAI methods
An Efficient XAI-Based Framework for a Robust and Explainable IDS (NUGRAHA; JNANASHREE; BAUSCHERT, 2024)	2024	SHAP + LIME
Revisiting the Performance–Explainability Trade-Off in XAI (CROOK; SCHLUTER; SPEITH, 2023)	2023	Discussion on trade-off
XAI Model for Cancer Image Classification (SINGHAL et al., 2024)	2024	Grad-CAM
Methods, Techniques, and Application of XAI (DUMKA et al., 2024)	2024	Attribution, rules, counterfactuals
The Enlightening Role of XAI in Medical & Healthcare Domains (ALI et al., 2023)	2023	SHAP, LIME, CAM/Grad-CAM
XAI – From Theory to Methods and Applications (ORTIGOSSA; GONÇALVES; NONATO, 2024)	2024	Theory, methods, and applications
XAI Techniques for Image Classification Models in Diverse Domains (JOSHI; BAGADE, 2023)	2023	Comparisons in computer vision
XAI for Trustworthy AI in 6G Networks (SHAFIK, 2025)	2025	Applications in networks

Snowballing and thematic saturation. *Backward* and *Forward Snowballing* strategies were also applied, with duplicate verification by DOI and title with the aid of the Connected Papers tool (Connected Papers, 2026). In this axis, *snowballing* showed strong overlap with the already selected set and did not result in the inclusion of new studies. Even so, it helped to recover foundational references and recurring applications, used as conceptual support throughout the dissertation.

2.3 Results and Discussion

2.3.1 Axis 1 – Student dropout prediction

The analysis of the 35 studies in Axis 1 revealed recurring patterns in the literature on dropout prediction in higher education.

The complementary verification in Learning Analytics venues also reinforced this axis. In particular, (MANRIQUE et al., 2019) analyzed student representation, representative features, and classification algorithms for degree dropout prediction. This study is relevant to this dissertation because it shows that dropout prediction depends not only on the selected algorithm, but also on how student trajectories are represented before modeling.

RQ1.1 – Techniques used

Traditional models, such as logistic regression, decision trees, and Random Forest, continue to be frequently used as a baseline for comparison (PALACIOS et al., 2021; CARDONA et al., 2023). These models appear frequently because they allow evaluating the performance of newer approaches against methods already consolidated in the literature. There is, however, an increase in the use of boosting methods and classifier ensembles, especially in more recent studies (BETTAHI; BELOUADHA; HARROUD, 2025; VAARMA; LI, 2024). These models tend to perform well on tabular data and yield more stable results than those obtained with models used in isolation.

This point is relevant to this dissertation because the educational data used are also organized in tabular format, with academic, institutional, and student profile variables. Furthermore, dropout is a multifactorial phenomenon, which makes the use of models capable of capturing different data patterns appropriate.

When there is a greater volume of temporal data or more detailed historical records, studies appear that combine tree-based models with neural networks (VEMULAPALLI et al., 2025; BARANYI; NAGY; MOLONTAY, 2020). Although these arrangements may yield gains, they do not appear to be the dominant standard in the analyzed literature.

RQ1.2 – Types and sources of data

Most works use academic and institutional data, such as grades, course failures, enrollment history, course progression, and attendance (PALACIOS et al., 2021; BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022). These data are important because they directly record the student's trajectory within the institution and allow for the observation of signals associated with retention or dropout.

Demographic and socioeconomic variables are also used, employed in a complementary manner to the analysis of the student's profile (VAARMA; LI, 2024; OQAIDI; AOUHASSI; MANSOURI, 2022; GUTIERREZ-PACHAS et al., 2023). In the context of this dissertation, this aspect underscores the need to integrate diverse types of information, as dropout analysis does not depend solely on academic performance but also on institutional and social characteristics that help contextualize the student's trajectory.

Fewer studies incorporate behavioral cues, especially data on the use of virtual learning environments, platform interactions, and temporal records of activities. When these data are available, they tend to improve predictive capacity. Nevertheless, these data are not always standardized in institutional systems, which reinforces the importance of working with information that is more readily available in academic databases.

RQ1.3 – Validation and metrics

Cross-validation remains widely used, but some studies have adopted strategies that better reflect how a dropout model would be used in an institution. In a practical application, the model tends to be trained with data from students from previous years or classes and, afterward, applied to more recent students still being monitored. For this reason, strategies such as cohort evaluation and temporal validation become important (SANTOS; PONTI; RODRIGUES, 2023). In these approaches, students are analyzed considering admission groups or the temporal order of the data, which reduces the risk of an artificially optimistic evaluation.

Regarding metrics, the most frequent are accuracy, F1-Score, and AUC-ROC. In imbalanced contexts, as is common in the dropout problem, F1-Score and AUC-ROC are more informative than accuracy alone (CARDONA et al., 2023; PALACIOS et al.,

2021). This occurs because accuracy may indicate good overall performance even when the model struggles to correctly identify dropout students, who correspond to the class of greatest interest.

Beyond classification metrics, more recent studies also highlight the importance of evaluating the quality of the probabilities produced by the models. This point is relevant because, in the context of dropout, the model's output is often interpreted as a risk estimate. Thus, it is not enough to indicate which students appear to be at higher risk; it is necessary that the assigned probability be consistent with the observed risk, thereby making its interpretation more useful for supporting institutional analyses.

Table 2.6: Evaluation metrics identified in the studies analyzed in Axis 1

Metric	**Purpose in dropout prediction**	**Examples of studies**
Accuracy	Provides a general measure of correct classifications, but may be insufficient in imbalanced datasets.	(PALACIOS et al., 2021; BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022; MAKSIMOVA; PENTEL; DUNAJEVA, 2021)
Precision	Indicates the proportion of students predicted as dropout who actually belong to the dropout class.	(VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023)
Recall	Indicates the proportion of dropout students correctly identified by the model.	(VAARMA; LI, 2024; PALACIOS et al., 2021; BETTAHI; BELOUADHA; HARROUD, 2025)
F1-Score	Balances precision and recall, being useful when the dropout class is of special interest.	(PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023)
ROC-AUC	Evaluates the model's ability to separate dropout and non-dropout students across thresholds.	(BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025)
Average Precision / PR-AUC	Evaluates performance in precision-recall space, especially relevant in imbalanced datasets.	(BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023)
Brier Score	Assesses the quality of predicted probabilities, which is relevant when the model output is interpreted as risk.	(BETTAHI; BELOUADHA; HARROUD, 2025)

The metrics identified in the literature support the evaluation strategy adopted in

this dissertation. Since the objective is not only to classify students, but also to estimate dropout risk, the evaluation combines metrics focused on the positive class, such as $F1_{pos}$, metrics of discrimination, such as ROC-AUC, metrics appropriate to imbalanced data, such as Average Precision, and a probabilistic calibration metric, represented by the Brier Score.

RQ1.4 – Limitations and gaps

A recurring limitation is the difficulty of comparing results across institutions and courses, due to the lack of standardized data and contextual differences (BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022). This limitation is important for this dissertation because it shows that the problem lies not only in the choice of the predictive model, but also in the way data are organized prior to analysis. When institutions record and structure their data differently, it becomes more difficult to reuse models, compare results, and interpret dropout patterns consistently.

Another frequent point is the difficulty of generalization: models that perform well in one institution do not always perform well in other scenarios (BERENS et al., 2025). Issues related to reproducibility, the absence of code or data availability, and the need for continuous model monitoring also appear. In this sense, works that adopt organized, modular, and reproducible pipelines point to a more promising path (BETTAHI; BELOUADHA; HARROUD, 2025).

Synthesis of Axis 1

In general terms, the literature indicates five main points:

(i) Tree-based models and boosting methods appear frequently and present good performance on tabular data;

(ii) Ensembles tend to produce more stable results than isolated models;

(iii) Academic and institutional data continue to be the main basis of the studies;

(iv) Temporal validations and cohort analyses bring the evaluation closer to the way the model would be used in an institution;

(v) Difficulties of comparison, generalization, and reproducibility still persist.

These results support the adoption of an ensemble-based approach, focused on institutional data and the careful evaluation of performance across different observation windows. In addition, they reinforce the need for a prior stage of data organization and standardization, a factor that is addressed in this dissertation using ontology as a canonical data model.

2.3.2 Axis 2 – Explainability (XAI)

The analysis of studies in Axis 2 shows that explainability has been used across different domains to make the results of machine learning models more understandable. The complementary verification in Educational Data Mining and Learning Analytics also reinforced the relevance of this axis. (GUNASEKARA; SAARELA, 2024) synthesized studies on explainability in these domains and highlighted that data types, models, and evaluation metrics influence how transparent and interpretable educational models can become. This reinforces the position adopted in this dissertation that explainability should be treated as part of the analytical process, rather than as a merely visual complement to prediction. In the context of this dissertation, this point is important because dropout prediction alone is not sufficient to support an institutional analysis. In addition to estimating risk, it is necessary to understand which factors contributed to that estimate.

In general terms, explainability techniques seek to answer why a model produced a given result. This explanation can occur at two levels. Local explanation seeks to interpret a specific prediction, that is, which variables contributed to a given instance being classified as such. Global explanation, in turn, seeks to understand the model's general behavior, indicating which variables are most influential in the analyzed dataset.

RQ2.1 – Most used techniques

Among the studies analyzed, SHAP and LIME are the most used techniques for post-hoc explanation. This type of explanation is applied after the model has already been trained and produced a prediction. Thus, the technique does not replace the predictive model, but helps to interpret the factors that influenced the presented result.

In tabular data, SHAP stands out for allowing both local and global readings of the model. The technique is based on Shapley values, originally proposed in game theory, and seeks to estimate how much each variable contributes to shifting the prediction relative to a model reference value (LUNDBERG; LEE, 2017). In other words, SHAP seeks to indicate how much each attribute contributed to increasing or decreasing the produced estimate.

In local explanation, SHAP assigns each variable a contribution value for a specific prediction. In the context of dropout, this allows, for example, observing whether course failures, course approvals, admission grade, or participation in institutional programs increased or decreased a student's risk estimate. In a global explanation, individual values can be aggregated, allowing the identification of which variables had the greatest influence on the model's overall behavior (LUNDBERG; LEE, 2017).

LIME is also widely cited, mainly for local explanations and for inspecting specific cases (RIBEIRO; SINGH; GUESTRIN, 2016; NAGY; MOLONTAY, 2024; ZANELLATI; GORI; FURLANELLO, 2024). Unlike SHAP, its logic consists of perturbing the data around a specific prediction and approximating, in that region, the behavior of a complex model, often treated as a black box, by means of a simpler and interpretable model. In this way, it allows observing which variables were most relevant for a given classification, although its explanation is limited to the case analyzed.

In addition to SHAP and LIME, techniques such as PFI and PDP/ICE also appear in the literature, generally as support for global interpretation and effect visualization (MARTINS et al., 2024; HAMIDA et al., 2024). PFI allows observing the importance of a variable by measuring the model's performance drop when its values are permuted or shuffled. PDP and ICE help visualize how changes in a variable can influence the model's output, whether in the dataset's average behavior or in individual cases.

This use of explainability techniques appears in studies across education, finance, auditing, security, and other areas (BARANYI; MOLONTAY, 2020; NAGY; MOLONTAY, 2024; ZANELLATI; GORI; FURLANELLO, 2024; MARTINS et al., 2024; ZHANG; CHO; VASARHELYI, 2022; JINAD; ISLAM; SHASHIDHAR, 2024). This discussion is important because it allows understanding how explanations can be used not only to

analyze individual classifications, but also to identify general patterns associated with dropout.

RQ2.2 – Advantages and limitations

SHAP emerges as a consistent approach to explain tree-based models and ensembles, allowing an articulated reading of individual cases and the overall behavior of the model (LUNDBERG; LEE, 2017). This characteristic is relevant, as it allows analyzing both the estimated risk for a specific student and the factors that recur across the set of evaluated students.

LIME is pointed out as a useful alternative for local reading, particularly because of its simplicity of use and interpretation (RIBEIRO; SINGH; GUESTRIN, 2016). However, because it is more suited to explaining specific cases, its use may be less appropriate when the objective is also to produce a consolidated view of the model's most influential factors.

On the other hand, limitations are also recurrent. The computational cost can be high in some scenarios, especially for large data volumes. In addition, the stability of the explanations and the effects of correlated variables remain points of attention (HAMIDA et al., 2024; CROOK; SCHLUTER; SPEITH, 2023). In other words, the explanation should not be taken as an automatic or definitive answer, but as support for analysis. In the case of a dropout, this means that a variable highlighted by the model should be interpreted alongside the academic and institutional context.

RQ2.3 – Institutional analysis support

The studies indicate that explainability can contribute at three levels. First, in the analysis of individual cases, when it is necessary to understand why a given prediction was produced. Second, in the global reading of the model, by highlighting the most influential factors in the analyzed set. Third, in the transparency of the analytical process, a relevant aspect in institutional and regulated contexts (ZHANG; CHO; VASARHELYI, 2022; WIRATSIN; RAGKHITWETSAGUL, 2025; BRANDENBURG et al., 2025).

In the educational context, the analyzed studies show that the use of explanations

can help interpret the estimated risk, identify recurring factors, and support more careful analyses on retention and dropout (BARANYI; MOLONTAY, 2020; NAGY; MOLONTAY, 2024; ZANELLATI; GORI; FURLANELLO, 2024). For this dissertation, this point is central, since the objective is not only to classify students at risk, but also to present elements that help understand which factors are associated with this classification.

The use of XAI is especially relevant because dropout prediction may involve academic, institutional, and profile variables, including admission data, scholarships, student assistance, quotas, gender, and ethnicity. Without explainability, the model could produce a risk estimate without indicating whether this estimate was mainly influenced by academic performance, institutional support, admission characteristics, or student profile variables. In this context, explainability helps make the model behavior more transparent and auditable, reducing the risk of treating the prediction as an opaque score detached from the educational context.

Thus, explainability can support institutional analysis by transforming the model's output into more interpretable information. Instead of presenting only a dropout probability, the explanatory analysis allows observing whether the estimated risk is more strongly associated with, for example, course failures, course withdrawals, low performance, admission method, or participation in institutional programs. This reading does not replace human interpretation, but can guide better-targeted institutional analyses.

Synthesis of Axis 2

The Axis 2 literature indicates that:

(i) SHAP and LIME are the most recurrent techniques in post-hoc explanations;

(ii) SHAP stands out in applications with tabular data and tree-based models;

(iii) Techniques such as PFI and PDP/ICE appear as support for global interpretation and effect visualization;

(iv) The usefulness of the explanation depends not only on the method, but also on clarity, stability, and cost of use;

(v) In the dropout context, explainability helps transform the prediction into a more comprehensible analysis for institutional support.

These results show that explainability should not be treated merely as a visual complement to the model, but as an important part of the process of interpreting the produced estimates. In the case of student dropout, this interpretation is necessary because identifying students at risk must be accompanied by elements that help understand the factors associated with that risk.

2.4 Discussion

Analyzing the two axes reveals an important convergence. In Axis 1, the studies indicated the use of tree-based models, especially ensembles and boosting methods, as robust approaches for dropout prediction (PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025). In Axis 2, SHAP appears as one of the main explanation techniques for this type of model, with emphasis on applications in tabular data (LUNDBERG; LEE, 2017; BARANYI; MOLONTAY, 2020).

Therefore, the choice of ensembles for prediction and SHAP for explanation is not random. These two decisions are present in recent literature and relate directly to the type of problem addressed in this dissertation, that is, estimating dropout risk based on institutional data and, at the same time, explaining the factors associated with this estimate.

The choice of XGBoost, LightGBM, and CatBoost is also aligned with the characteristics of the data used in this work. The educational data were organized mainly as tabular attributes, composed of numerical and binary variables representing academic trajectory, admission, quotas, scholarships, student assistance, and student profile. Tree-based boosting models are appropriate for this type of structure because they can capture non-linear patterns and interactions among variables without requiring the same assumptions as linear models. In addition, these algorithms are compatible with SHAP-based explanations, which is important for the interpretability objective of EducAAr.

The complementary verification also showed that other forms of representation

have been explored in recent educational prediction studies. For example, (MURESAN; CARDEI; CARDEI, 2025) used heterogeneous graph deep learning models to predict student success, incorporating dynamic features and relationships among different educational entities. This type of approach is relevant because it shows that educational trajectories can be represented not only as tabular attributes, but also as relational structures. In this dissertation, however, the tabular representation was maintained because the objective was to evaluate the integration between ontology, tree-based predictive models, SHAP explanations, and visualization within the EducAAr architecture.

Another convergence is the concern with stability. In prediction studies, this appears in the defense of temporal validations and cohort analyses (SANTOS; PONTI; RODRIGUES, 2023; GLANDORF et al., 2024). In explainability, it arises in the discussion about the robustness of explanations and caution against simplified readings (HAMIDA et al., 2024). This reinforces that obtaining good performance is not enough. It is also necessary that the behavior of the model and of the explanations be sufficiently consistent to support institutional analyses.

The review also reveals a persistent concern. More sophisticated models can improve predictive performance, but may demand greater interpretive effort (CROOK; SCHLUTER; SPEITH, 2023). In a problem such as student dropout, this concern is fundamental, since a high accuracy rate is not enough for institutional analysis. It is also necessary that the results be accompanied by understandable explanations.

In this context, the combination of an ensemble, probabilistic calibration, and SHAP presents itself as a viable option to balance performance and interpretation. Probabilistic calibration consists of adjusting the probabilities produced by the model so that they are better aligned with the observed outcomes. Thus, when the model indicates a given dropout risk, this value is expected to be more consistent with the empirical distribution observed in the data. This makes the model output more suitable for institutional analysis, since the probability is no longer treated only as a classification score, but as a calibrated risk estimate.

Finally, the literature reveals that many studies stop at the experimental stage. Even when they present good results, they do not always discuss how these predictions

and explanations could be organized into analytical workflows for institutional use.

2.5 Gaps and Research Opportunities

The review allowed identifying two main gaps that guide this dissertation.

1. **Difficulty in data integration and standardization.** The Axis 1 studies show that it is still difficult to compare results across courses and institutions, due to differences in data organization and the absence of standards (BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022; BERENS et al., 2025). In the analyzed set, no recurring use of canonical models as a common basis prior to the modeling stage was identified. This factor opens space for a proposal that better addresses data organization before the application of predictive models.

2. **Little integration between prediction and explainability in the same architecture.** The analyzed studies show advances both in prediction and in explainability, but works that use these two parts in a single structure are still few, with focus also on reproducibility, monitoring, and presentation of results for institutional use (BETTAHI; BELOUADHA; HARROUD, 2025). This gap is relevant because, in the context of dropout, the usefulness of the model depends not only on predicting risk, but also on presenting explanations that help interpret that risk.

In addition to these gaps, the review also justifies the selection of the techniques used in this dissertation. Although LIME appears frequently in the literature, SHAP is more aligned with the type of model and type of data used in this work. This is because the adopted models are tree-based and the data are organized in tabular format. In addition, SHAP allows analyzing both individual explanations and global patterns, which more closely matches the objective of this research.

Based on these gaps, this dissertation proposes the EducAAr architecture. The work distinguishes itself by integrating data organization, predictive modeling, explainability, and results presentation into a single structure designed to support interpretable institutional analyses of dropout risk.

In this sense, the contribution of EducAAr is not to replace institutional interpretation or to prescribe interventions automatically. Instead, the architecture provides an organized workflow that connects heterogeneous data, predictive estimates, explanatory factors, and visualization resources. This workflow offers a technical basis for analyzing dropout risk while preserving the need for contextual interpretation by institutional actors.

In the next chapter, the procedures adopted for developing the architecture and evaluating it are presented.

3 Methodology

This chapter presents the methodology adopted in the development of the EducAAr (Educational Analysis Architecture). This research was conducted based on *Design Science Research* (DSR) (PIMENTEL et al., 2020; HEVNER, 2007), since it is a methodology aimed at the construction and evaluation of artifacts with practical purpose and scientific grounding.

The Systematic Literature Review, presented in Chapter 2, identified two areas needing further research: standardizing the organization and integration of educational data, and the few studies that combine dropout prediction with explainability in a single framework (BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022; BERENS et al., 2025; BETTAHI; BELOUADHA; HARROUD, 2025). These research directions significantly structured the methodological decisions of this dissertation.

This chapter details how the architecture was planned, built, and evaluated, showing the relationships among the investigated problem, the defined requirements, and the decisions adopted throughout the research. Section 3.1 presents the characterization of the research. Section 3.2 describes the DSR methodology. Section 3.3 presents the general planning of the investigation. Section 3.4 defines the requirements of the artifact. Subsequently, the two development cycles of EducAAr and the corresponding validation procedures are detailed.

3.1 Research Characterization

This research is applied, as it addresses a specific problem in the educational domain through the development of a computational artifact. It also has a constructive and evaluative character, since it involves the design, implementation, and technical evaluation of the EducAAr architecture.

From the methodological point of view, the study follows the logic of Design Science Research. Therefore, it was not structured as a qualitative study in the strict

sense, nor was it organized around formal statistical hypotheses. Instead, the research was guided by a research question, by artifact requirements, and by technical evaluation criteria associated with the architecture.

The evaluative aspect appears in different stages of the work. In the first cycle, the ontology was evaluated through consistency verification and queries over the integrated data. In the second cycle, the evaluation involved the construction and assessment of predictive models, the analysis of performance metrics, the calibration of probabilities, the interpretation of SHAP values, and the organization of the results in the visualization panel.

The quantitative component is present in the use of machine learning techniques, performance metrics, calibrated probabilities, and SHAP values. The interpretation of these results is used to discuss the behavior of the artifact and the factors associated with the model estimates. However, this interpretive discussion does not characterize the research as qualitative, since no qualitative data collection protocol, such as interviews, focus groups, or thematic coding of textual data, was conducted.

In this sense, EducAAr was conceived as an artifact capable of integrating educational data, producing dropout risk estimates, generating explanations, and organizing the results for analysis. The purpose of the research is not only to observe the dropout phenomenon, but to propose and technically evaluate an architecture that supports a more structured analysis of this phenomenon.

The EducAAr proposal is not limited to a closed system or to a single predictive model. It is an architecture designed to support different analyses on educational data, including information organization, pattern identification, dropout risk estimation, and interpretation of predictive results. This characteristic is aligned with DSR, which values artifacts that can be reused, adapted, and evolved in response to new problems and contexts.

3.2 Design Science Research Methodology

Design Science Research (DSR) is a methodology widely used in research on Computing and Information Systems when the focus is on the construction of artifacts to solve real

problems. Instead of restricting itself to the observation or description of phenomena, DSR proposes a process oriented to the design, construction, and evaluation of solutions (HEVNER, 2007).

In this dissertation, DSR was adopted because it provides an adequate basis for the development of EducAAr, enabling the articulation of practical problems, the establishment of theoretical foundations, the construction of the artifact, and the evaluation of results. According to (PIMENTEL et al., 2020), DSR seeks both the production of an artifact and the generation of knowledge from its construction and use, which is directly related to the objectives of this work.

In general terms, DSR starts with the identification of a problem, proceeds to the definition of requirements, the design of the artifact, and its evaluation. This process can occur across multiple iterations, so the solution is refined throughout the research. Figure 3.1 presents the general DSR cycle considered in this dissertation.

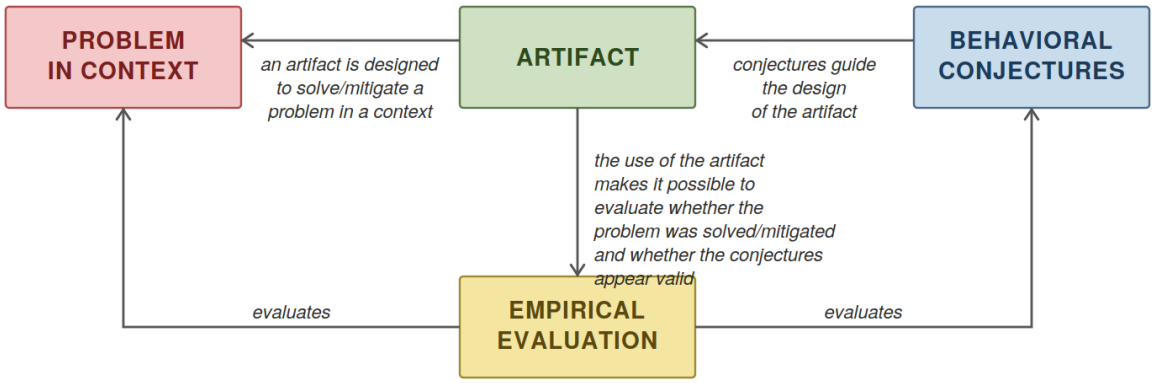

Figure 3.1: Iterative cycle of Design Science Research: problem identification, design, construction, and evaluation. Source: adapted from (PIMENTEL et al., 2020)

As proposed by (HEVNER, 2007), DSR can be understood through three interconnected cycles: relevance, rigor, and design. The relevance cycle connects the research to the demands of the context in which the problem is situated. The rigor cycle ensures that the constructed solution is supported by already consolidated knowledge. The design cycle corresponds to the actual construction and evaluation of the artifact.

In practice, these principles were materialized in two development cycles of EducAAr. The first cycle focused on the organization and integration of data through an

ontology. The second cycle expanded the architecture to incorporate machine learning, explainability, and visualization mechanisms, giving continuity to the previously built foundation.

The relevance cycle was represented by the institutional problem of dropout analysis and by the need to organize heterogeneous educational data for analytical use. The rigor cycle was supported by the systematic literature review, by the theoretical foundations of ontologies, machine learning, probabilistic calibration, and explainability, and by previous work related to the first version of EducAAr. The design cycle was materialized in the construction, refinement, and technical evaluation of the architecture through the two development cycles described in this chapter.

3.3 Research Planning

The research planning was elaborated based on the relationship between the investigated problem, the results of the systematic review, and the objectives defined for the EducAAr architecture. The main Research Question (RQ) was the following:

> *How can an architecture based on a canonical model for educational data integration combine machine learning, explainability, and visualization mechanisms to produce interpretable analyses of student dropout risk in higher education?*

Because the study follows a Design Science Research approach, it was not structured around formal statistical hypotheses, but around a research question, artifact requirements, and technical evaluation criteria.

Based on this RQ, two central ideas were formulated to guide the development of the work. The first is that the integration of data into a canonical model can facilitate its representation, querying, and consistency verification. The second is that the combination of predictive models, explainability techniques, and visualization mechanisms within a single architecture can support the production of interpretable analyses of student dropout risk (BETTAHI; BELOUADHA; HARROUD, 2025).

Based on this, the DSR cycles, the architecture requirements, and the general

evaluation criteria were defined. These elements function as a connecting axis between the problem, the proposed solution, and the results.

3.4 Functional and Non-Functional Requirements

The requirements definition was used to guide both the architecture's construction and its evaluation. These requirements were derived from the gaps identified in the literature review and from the needs observed in the institutional context of higher education (BARRAMUÑO; MEZA-NARVÁEZ; GÁLVEZ-GARCÍA, 2022; BERENS et al., 2025; BETTAHI; BELOUADHA; HARROUD, 2025).

3.4.1 Functional Requirements

The functional requirements define what the EducAAr architecture must be able to do:

- **FR01 – Data Integration**: integrate academic, administrative, and contextual data from different sources;
- **FR02 – Canonical Model**: use an ontology to represent concepts, attributes, and relationships of the educational domain;
- **FR03 – Inference**: allow obtaining derived information from the organized data;
- **FR04 – Machine Learning Analysis**: incorporate machine learning models for the analysis of structured data;
- **FR05 – Explainability**: provide mechanisms that allow interpreting the models' results;
- **FR06 – Visualization and Analysis Support**: present metrics, explanations, and simulations in an accessible manner to support the interpretation of dropout risk analyses.

3.4.2 Non-Functional Requirements

The non-functional requirements concern the quality and the constraints of the architecture:

- **NFR01 – Reliability**: ensure data consistency and correct execution of the analyses;
- **NFR02 – Extensibility**: allow inclusion of new data, variables, and courses;
- **NFR03 – Flexibility**: adapt to different scenarios and analytical objectives;
- **NFR04 – Sustainability**: enable maintenance and evolution over time;
- **NFR05 – Privacy and Ethics**: respect the LGPD (Brazilian General Data Protection Law) and use anonymized data (BRASIL, 2018);
- **NFR06 – Transparency**: present understandable results and explanations;
- **NFR07 – Computational Feasibility**: allow the execution of the modeling, explanation, and visualization procedures using the computational resources available in the research environment.

These requirements were revisited throughout the two DSR cycles and served as a reference for evaluating the artifact.

Table 3.1: Relationship between requirements, architectural decisions, and evaluation evidence

Requirement	**Architectural decision**	**Evaluation evidence**
FR01 – Data Integration	Use of the ontology to integrate student records, academic history, admission data, quotas, scholarships, and student assistance information.	Instantiation of institutional data in the first DSR cycle and generation of integrated datasets in the second cycle.
FR02 – Canonical Model	Definition of an OWL ontology as the canonical representation of the educational domain.	Consistency verification with the HermiT reasoner and use of the ontology as the basis for data organization.
FR03 – Inference	Use of ontological relationships, restrictions, and inferred properties to derive information not directly represented in the original files.	Execution of reasoning procedures and SPARQL queries over the instantiated ontology in the first cycle.
FR04 – Machine Learning Analysis	Transformation of the integrated data into tabular datasets and use of tree-based models combined in a calibrated ensemble.	Evaluation of XGBoost, LightGBM, CatBoost, and the ensemble using $F1_{pos}$, ROC-AUC, Average Precision, and Brier Score.
FR05 – Explainability	Use of SHAP to explain global and local predictions produced by the models and by the ensemble.	Analysis of global SHAP rankings, local explanations, and triangulation of explainability results across the 16 programs.
FR06 – Visualization and Analysis Support	Development of a panel to organize metrics, explanations, simulations, technical details, and LLM-assisted textual summaries.	Evaluation of the panel using the Ciência da Computação – Noturno program as a reference case.
NFR01 – Reliability	Separation between training, validation, and test sets; use of stratified partitions; multiple repetitions; and consistency verification in the ontology.	Stability analysis across 30 repetitions and logical validation of the ontology.
NFR02 – Extensibility	Modular organization of the architecture, allowing the inclusion of new programs, variables, and datasets.	Application of the second cycle to 16 undergraduate course offerings.
NFR03 – Flexibility	Organization of the workflow into independent stages: ontology, vectorization, modeling, explainability, and visualization.	Reuse of the same architecture across different programs and observation windows.
NFR04 – Sustainability	Separation between data integration, predictive modeling, explanation generation, and panel visualization.	Possibility of updating models, variables, and visualizations without redefining the entire architecture.
NFR05 – Privacy and Ethics	Use of anonymized data and processing in a restricted research environment after approval by the Ethics Committee.	Description of the CAAE approval process and data access conditions.
NFR06 – Transparency	Presentation of performance metrics, SHAP explanations, simulation results, prompts, and cautionary notes against causal interpretation.	Explainability analysis, panel description, and threats to validity section.
NFR07 – Computational Fea-	Adoption of tree-based models and post-hoc explainability techniques	Successful execution of training, calibration, explanation, and panel

Table 3.1 makes explicit how the requirements guided the main architectural decisions of EducAAr and how each group of requirements was addressed during the evaluation. This relationship was important to avoid treating the requirements as isolated statements. Instead, they functioned as design criteria for the construction of the artifact and as reference points for the technical evaluation conducted in the two DSR cycles.

3.5 First DSR Cycle: Ontological Modeling and Exploratory Analysis

The first cycle focused mainly on meeting the requirements FR01, FR02, and FR03, related to data integration, canonical modeling, and inference. In this stage, the main objective was to organize educational data from different sources into a single structure that could represent academic, social, and institutional information in an articulated manner. The results of this cycle were presented in (AZY et al., 2024).

The motivation for this first cycle is that, in many institutional contexts, the data needed to analyze dropouts are dispersed across different files, systems, and formats. This hinders both the integrated interpretation of the information and the development of broader analyses of students' trajectories.

In this cycle, EducAAr was initially built as a foundation for the organization and querying of data. The objective was not yet to predict dropout individually, but to structure the information and enable more consistent exploratory analyses. For this purpose, an ontology was specified as a canonical model and as a means of querying the integrated information.

3.5.1 Ontology Development

The ontology was developed based on the six phases of Methontology (FERNÁNDEZ-LÓPEZ; GÓMEZ-PÉREZ; JURISTO, 1997). The specification stage was defined from the beginning of the research, as the purpose of the ontology was to organize academic information from different sources to support analyses of dropout in higher education

The phases considered were:

- **Specification** – definition of the purpose of the ontology;
- **Conceptualization** – definition of classes, subclasses, and relationships;
- **Formalization** – logical organization of the concepts;
- **Implementation** – coding in OWL;
- **Integration** – loading and instantiation of the data;
- **Maintenance/Evolution** – updating of the ontology as needed.

The central class of the model is the **Student** class, which represents each student. From it, other classes were associated to describe different dimensions of the academic trajectory and student profile.

Among the main classes defined, the following stand out:

- **Status**: academic situation of the student, with subclasses *Active*, *Completed*, and *Withdrawn*;
- **Nativity**: place of birth, distinguishing students from the same city as the institution and from other locations;
- **PeriodDropout**: period range in which dropout occurred, when applicable;
- **Ethnicity**: self-declared ethnicity;
- **Gender**: gender of the student;
- **AdmissionType**: admission method;
- **StudentAssistance**: participation in institutional assistance;
- **Grade**: grade obtained in a course subject;
- **CourseSubject**: course subject associated with the grade;
- **AcademicPerformance**: cumulative academic performance.

For the **AcademicPerformance** class, four levels were defined: *Insufficient*, *Regular*, *Good*, and *Excellent*. The lower range was defined based on the general approval criterion of most Brazilian universities[1], which sets 60 points as the minimum average. The ranges above this limit were used to distinguish different performance levels among approved students.

This structure enabled the integration of attributes originally distributed across distinct databases, facilitating integrated analyses of the academic trajectory.

3.5.2 Properties and Restrictions

The object properties, or relationships between the ontological classes, were defined to establish the relationships between *Student* and the other classes of the ontology. Among the main relationships are the associations between student and academic situation, admission method, gender, ethnicity, place of birth, institutional assistance, academic performance, and grades.

The relationship between *Grade* and *CourseSubject* was also defined, allowing each grade to be linked to the corresponding course subject.

Data properties were used to represent numerical values, such as grades and limits used in the classification of performance. In addition, restrictions were applied to ensure structural consistency. A student can be associated with several grades, while each grade is linked to a single course subject. Categories such as gender, ethnicity, academic situation, and performance, in turn, were defined with cardinality of exactly one instance per student.

The subclasses of these categories were also defined in a disjoint manner, ensuring that each student belongs to only one option within each group.

In the implementation phase, the Protégé tool was used, which is widely employed for constructing OWL ontologies (Stanford Center for Biomedical Informatics Research, 2025).

[1]This is a concept adopted in the work but which can be easily modified so that other types of approval can be incorporated, including private Brazilian universities and universities outside the country.

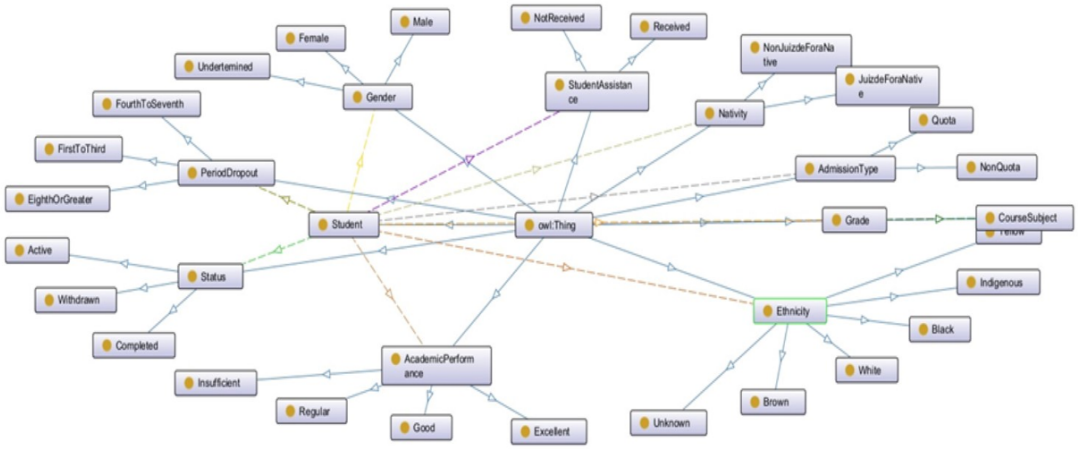

Figure 3.2: Ontology developed in the first cycle.

3.5.3 Instantiation and Integration

The ontology was instantiated in Python using the Owlready2 library (LAMY, 2024). The process began with the reading of the institutional files and with the automatic creation of the instances corresponding to the classes of the model.

For each student, instances linked to data of status, admission, gender, ethnicity, nativity, and assistance were created. The academic history was converted into instances of *Grade* and *CourseSubject*. Afterward, the cumulative academic performance was calculated based on the grades and represented by the *AcademicPerformance* class.

This process transformed previously dispersed records into a single, related data structure. From this, the HermiT *reasoner* (GLIMM et al., 2014) was used to verify the model's consistency and enable inference. It also became possible to perform SPARQL queries and apply SWRL rules over the loaded data.

3.5.4 Validation of the First Cycle

The validation of the first cycle was planned to verify whether the ontology met the requirements FR01, FR02, and FR03. The focus of this evaluation was to confirm whether the constructed structure could integrate diverse data sources, maintain logical consistency, and enable the retrieval of relevant information

For this purpose, the HermiT *reasoner* (GLIMM et al., 2014), integrated with Protégé, was used to identify possible conflicts between classes, restriction violations, or

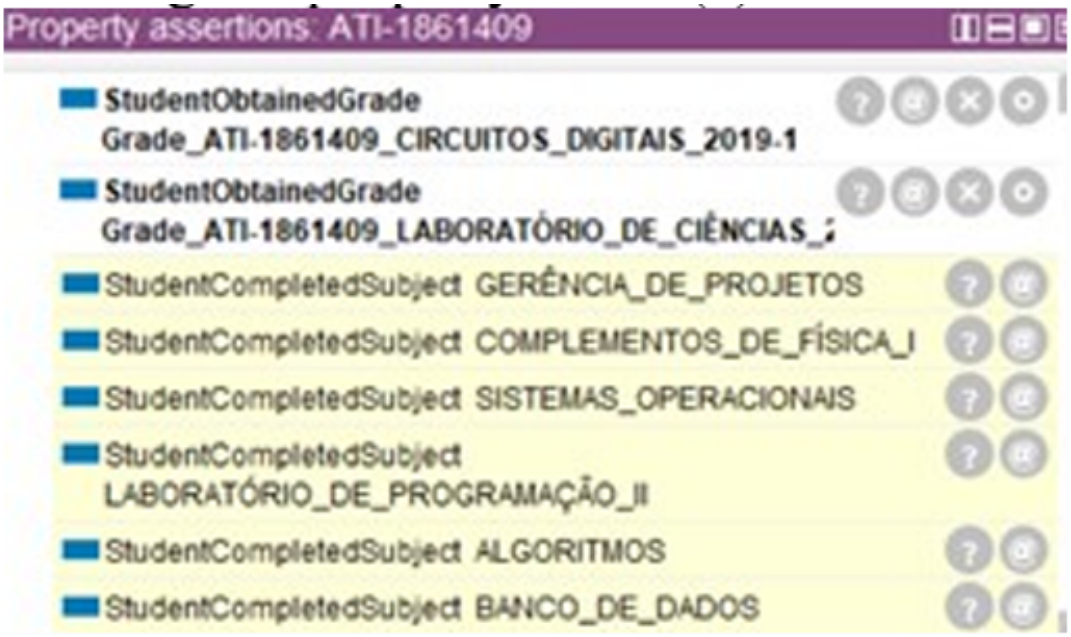

Figure 3.3: Example of properties inferred from the rules defined in the ontology.

modeling inconsistencies.

After this structural verification, SPARQL queries on the instantiated data were planned. These queries were used to verify whether information such as academic performance, student situation, admission method, and assistance was correctly represented and related. They also served to test the operation of the inferences foreseen in the model.

Finally, the practical usefulness of the ontology was evaluated by exploring the integrated database to identify patterns related to dropout. This stage involved cross-tabulating variables of interest, such as academic performance and student assistance, to verify whether the constructed structure enabled interpretable analyses of the phenomenon.

The detailed results of this stage are presented in Chapter 4 and were published in (AZY et al., 2024).

3.5.5 Synthesis of the First Cycle

The initial cycle achieved its goal by consolidating institutional data into a unified structure that could handle consistent queries and support inference. This stage demonstrated that it was possible to integrate data from different sources into a single base, meeting the requirements defined for the initial phase of the research.

At the same time, this cycle revealed an important limitation: the analyses produced were retrospective. Although they allowed for identifying patterns and relationships between variables, they did not yet produce individual risk estimates or explanations associated with these estimates, which, in a context of cause analysis for dropout, is of great importance.

This limitation motivated the expansion of the architecture in the second cycle, with the incorporation of machine learning and explainability, preserving the previously built foundation and broadening its analytical capacity.

3.6 Second DSR Cycle: Integration of ML and XAI

The second cycle aimed to expand EducAAr to meet the requirements FR04, FR05, and FR06, related to predictive analysis, explainability, and visualization. It also addressed non-functional requirements associated with transparency, privacy, reliability, extensibility, and computational feasibility.

In this stage, the architecture began to integrate machine learning models and explainability mechanisms. The ontology built in the previous cycle served as the foundation for data organization but was expanded to accommodate new variables and relationships necessary for predictive modeling.

The definition of this second cycle was also guided by the findings of the literature review, Chapter 2. In Axis 1, identified in Chapter 2, the studies pointed to the recurrence of tree ensembles in dropout prediction (PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023). In Axis 2, it became clear that explainability plays an important role in the interpretation of predictions, especially in institutional contexts (BARANYI; MOLONTAY, 2020; NAGY; MOLONTAY, 2024; ZANELLATI; GORI; FURLANELLO, 2024). Based on this, an explicit stage for explaining the results was included in the architecture.

3.6.1 Overview of the EducAAr Architecture in the Second Cycle

Figure 3.4 shows the version of the architecture resulting from the second cycle. In this configuration, the ontology continues to serve as the foundation for organizing information, while new layers act on these data to produce predictive analyses, explanations, and analysis-support resources.

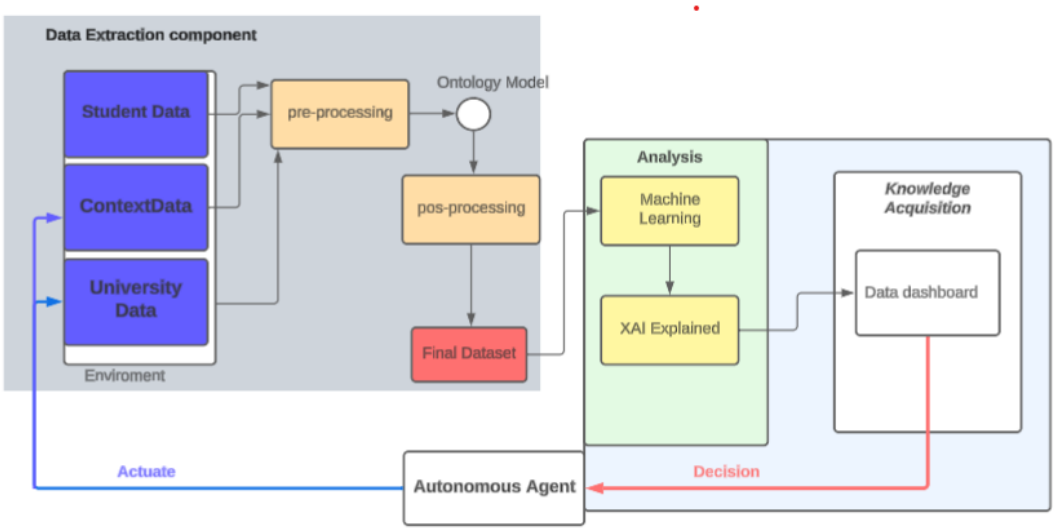

Figure 3.4: EducAAr architecture in the second cycle: integration between organized database, predictive analysis, explainability, and analysis support.

The architecture was organized into four main components:

1. data extraction and integration component;
2. analysis layer;
3. knowledge acquisition and analysis support;
4. institutional analysis support agent.

These components structure the processing workflow that goes from data acquisition to the presentation of results for institutional use.

3.6.2 Ontology Expansion

The initial modeling of the ontology was aimed at organizing data and enabling retrospective analyses. With the incorporation of the predictive stage, it became necessary to refine the representation of some academic variables.

One of the changes occurred in the way admission was represented. In the first cycle, the main distinction was between affirmative-action and non-affirmative-action students. In the second, this information was reorganized into two dimensions: selection method (e.g., Vestibular, PISM, SISU, or others) and type of quota associated with the position. This change enabled more precise representation of access modalities over time.

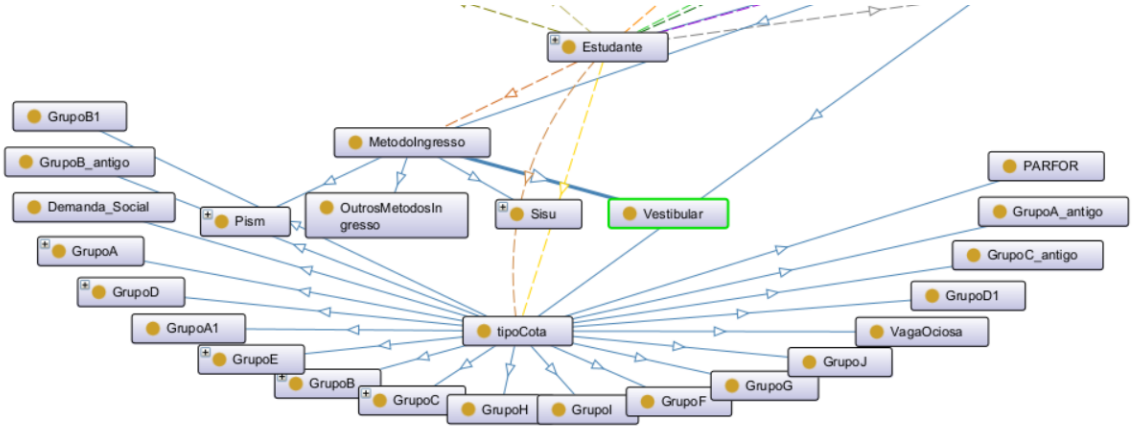

Figure 3.5: Expanded representation of the admission method.

The representation of student assistance and institutional support was also broadened in the second cycle. In the first cycle, although the original institutional data already contained more detailed records about scholarships, projects, and institutional programs, this information was used in an aggregated way, indicating only whether the student had received some type of benefit. In the second cycle, this level of detail was explicitly represented in the ontology. The original records, which included information such as the project position, project, scholarship type, modality, remuneration status, and start and end dates of the student's participation, were associated with instances of *VagaProjeto*. Each *VagaProjeto* instance represents the student's link to a specific scholarship, project, or institutional program during a given period. This instance was connected to a *Projeto* instance, which in turn was linked to a *Bolsa* instance. Thus, the ontology did not cre-

ate this information artificially, but organized and represented a level of detail that was already available in the institutional data. This reformulation made it possible to identify not only whether the student received support, but also when the support occurred, for how long it remained active, and under which modality, allowing these records to be crossed with the academic periods used in the predictive stage.

Figure 3.6 presents an example, in Portuguese, considering that the dataset is in Portuguese. Each *VagaProjeto* was linked to an instance of *Projeto* (project), and each project to an instance of *Bolsa* (scholarship). With this, it became possible to record not only whether the student received a benefit, but also when, for how long, and in what modality.

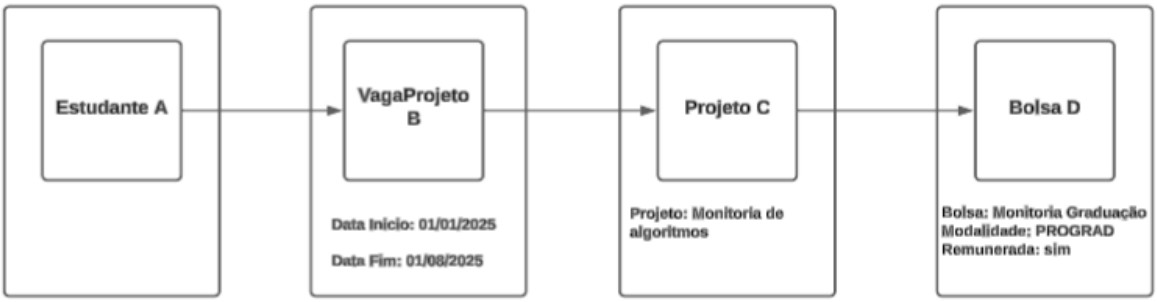

Figure 3.6: Structure of the relationship between VagaProjeto, Projeto, and Bolsa (in portuguese).

This reformulation enabled linking participation in institutional programs to the academic periods completed, which became important for the predictive stage.

Adjustments were also made in the representation of the academic history. The *Turma* (class) class came to record, in addition to year and semester, the start and end dates. For each student and class taken, an instance of *DesempenhoTurma* (class performance) was created, responsible for storing the final situation, the obtained performance, and the period taken.

With these changes, the ontology came to represent the academic trajectory in more detail, creating the necessary conditions for the predictive modeling stage.

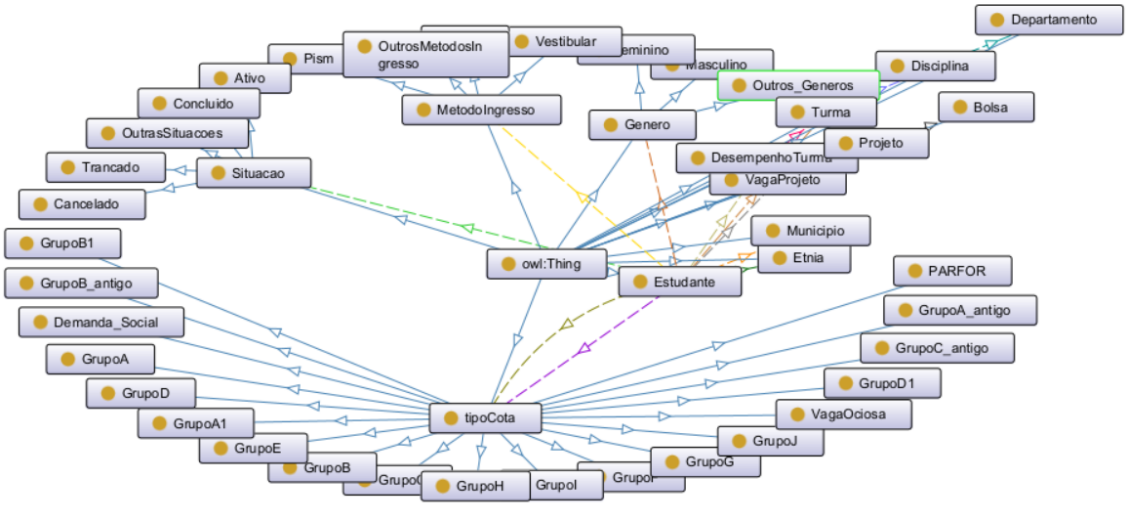

Figure 3.7: Expanded ontology in the second DSR cycle.

3.6.3 Data Extraction and Integration Component

The extraction and integration component is responsible for obtaining, processing, and organizing the data used in the second cycle. This stage includes extraction, cleaning, handling of missing values, and anonymization without violating privacy.

The expanded ontology continued to serve as a foundation for organization throughout this process, integrating information from different institutional systems and producing an integrated dataset for the following stages.

Loading of the Dataset

As in the first cycle, this stage starts with a previously anonymized institutional dataset. The data were organized into three main files: registration and admission information, academic history by class, and links to scholarships and projects.

Each student record gave rise to an instance of the `Estudante` (Student) class, associated with information on admission, ethnicity, gender, municipality, and academic situation. The history was converted into instances of `Turma`, `Disciplina` (Course Subject), and `DesempenhoTurma`. The scholarship records were represented by instances of `VagaProjeto`, `Projeto`, and `Bolsa`.

At the end of the process, the consolidated ontology was exported in OWL/RDF format.

Vectorization and Data Preparation

The use of the ontology before generating the tabular format was important because it not only allowed gathering the data into a common structure but also derived relationships that were not explicitly represented in the original files. Instead of treating each file in isolation, the ontology enabled the student to connect their academic trajectory to the associated institutional experiences.

As an example, by means of *property chains*, it was possible to infer the relationship `StudentReceivedScholarship` from the sequence `StudentOccupiesProjectPosition`, `ProjectPositionRefersToProject`, and `ProjectBelongsToScholarship`. In practice, this allowed directly identifying whether the student had any link to a scholarship, without the need to manually traverse all the connections between position, project, and scholarship at each stage of the analysis.

In this way, the ontology served as a preparatory stage for integrating and deriving relationships before the data were transformed into tabular attributes.

After the ontology was instantiated, the information was converted to the tabular format used for training the classification models. Each student was represented by a record containing fixed attributes and variables calculated at each period.

Among the fixed attributes are admission information, sociodemographic profile, and quota categories. To maintain comparability across records, the quota groups were reorganized into binary categories, including racial quota, income quota, public-school quota, quota for persons with disabilities, and open competition.

The academic variables were organized by academic term, considering approvals, failures by grade, failures due to insufficient attendance, course withdrawals, and other recorded situations. Participation in scholarships and the receipt of student assistance were also represented by a period, given the temporal intersection between the validity of the link and the reference period.

Datasets corresponding to the windows $p2$, $p3$, and $p4$ were generated, representing slices built up to the second, third, and fourth periods of the program, respectively. These datasets were organized incrementally based on the amount of academic history available in each window. The definition of these windows aligns with the literature on

early dropout prediction, which highlights the program's initial periods as especially relevant for identifying early signs associated with dropout risk (PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025).

The first semester, by itself, may concentrate initial trajectory adjustments, such as program reoptions and unstable enrollment patterns. For this reason, the period $p2$, which represents the second period, was adopted as the initial milestone of the analysis, since it already allows observing a complete academic cycle and generating more consistent performance variables.

The use of windows p2, p3 and p4 allowed for the evaluation of the models at different points along the academic trajectory. Students without complete records for the corresponding window were excluded from the dataset for that stage, ensuring consistency in the variables used.

Variables Used in Predictive Modeling

The variables used in modeling were organized into five main groups: admission and access information, quota information, sociodemographic profile, academic performance by period, and institutional support by period.

The variable *status* was defined as the binary target variable, where *status*=1 indicates a dropout student and *status*=0 indicates a graduate student. Active students were not included in supervised training because they do not yet have a final outcome, which would compromise the adjustment and calibration of the probabilities. After training, the model can be applied to active students to estimate prospective risk, without these cases being part of the training.

For temporal variables, the notation x was adopted to represent the reference academic period. Thus, the variable *periodo_x_disciplinas_ri*, for example, represents the number of course subjects in which the student failed due to insufficient attendance in period x. Similarly, variables related to scholarships and student assistance indicate whether the corresponding institutional support was active in the reference period.

Table 3.2: Variables used in predictive modeling

Group	**Variable**	**Description**	**Type**
Dependent	*status*	Binary target variable. Value 1 indicates dropout and value 0 indicates completion. Active students were not used in supervised training because they did not yet have a final academic outcome.	Binary
Admission	*nota_ingresso*	Admission grade associated with the student's entry into the program, used when available for the considered admission modalities.	Numeric
	tipo_ingresso_pism	Indicates whether the student entered through PISM, the serial admission process used at UFJF.	Binary
	tipo_ingresso_sisu	Indicates whether the student entered through SISU.	Binary
	tipo_ingresso_vestibular	Indicates whether the student entered through the traditional entrance exam.	Binary
	tipo_ingresso_outros	Groups other admission modalities represented in the institutional records.	Binary
Quotas	*cota_racial*	Indicates whether the student was associated with a racial quota category.	Binary
	cota_ampla_concorrencia	Indicates whether the student entered through open competition, without quota reservation.	Binary

Group	Variable	Description	Type
	cota_escola_publica	Indicates whether the student was associated with a public-school quota category.	Binary
	cota_renda	Indicates whether the student was associated with an income-based quota category.	Binary
	cota_pcd	Indicates whether the student was associated with a quota category for persons with disabilities.	Binary
Profile	*genero_masculino*	Indicates whether the student record was associated with the male gender category.	Binary
	genero_feminino	Indicates whether the student record was associated with the female gender category.	Binary
	genero_outros	Groups other or unavailable gender records, according to the categories present in the institutional data.	Binary
	etnia_branca	Indicates whether the student record was associated with the white ethnicity category.	Binary
	etnia_parda	Indicates whether the student record was associated with the brown ethnicity category.	Binary
	etnia_preta	Indicates whether the student record was associated with the black ethnicity category.	Binary

Group	Variable	Description	Type
	etnia_outra	Groups other, undeclared, or less frequent ethnicity records according to the institutional data.	Binary
Academic performance	*periodo_x_ disciplinas_ aprovadas*	Number of course subjects approved by the student in academic period x.	Integer
	periodo_x_ disciplinas_ reprovadas	Number of course subjects failed by grade in academic period x.	Integer
	periodo_x_ disciplinas_ ri	Number of course subjects in which the student failed due to insufficient attendance in academic period x.	Integer
	periodo_x_ disciplinas_ trancadas	Number of course subjects withdrawn by the student in academic period x.	Integer
	periodo_x_ disciplinas_ outros_status	Number of course subjects with other academic situations recorded in academic period x.	Integer
Institutional support	*periodo_x_ bolsa_ remunerada*	Indicates whether the student had a paid scholarship or paid institutional project link active in academic period x.	Binary

Group	Variable	Description	Type
	periodo_x_ bolsa_n_ remunerada	Indicates whether the student had an unpaid scholarship or unpaid institutional project link active in academic period x.	Binary
	periodo_x_ ae	Indicates whether the student received student assistance in academic period x.	Binary

The variables were derived from the integrated institutional records represented in the ontology and later transformed into tabular attributes for modeling. Binary variables indicate the presence or absence of a given characteristic in the student record, while integer variables count academic events in each reference period. Variables related to gender, ethnicity, quotas, scholarships, and student assistance were maintained because they may help reveal contextual patterns associated with dropout risk. However, as discussed in the threats to validity, these variables must not be interpreted as causal explanations or as criteria for individual judgment.

3.6.4 Analysis Layer

The analysis layer is responsible for the construction, evaluation, and interpretation of the predictive models. It is in this layer that the requirements FR04 and FR05 are mainly addressed.

The problem was treated as a binary classification task, in which students are classified as dropout or graduates. This definition was adopted because the analysis aims to distinguish, directly, the two academic outcomes most relevant to the assessment of dropout risk: the definitive interruption of the academic trajectory and the completion of the program. By structuring the problem this way, it becomes possible to estimate the probability that a student belongs to the at-risk group, thereby favoring the early identification of cases that require institutional follow-up.

Based on recent literature (PALACIOS et al., 2021; VAARMA; LI, 2024; BETTAHI; BELOUADHA; HARROUD, 2025; SANTOS; PONTI; RODRIGUES, 2023), tree-based models were adopted: XGBoost, LightGBM, and CatBoost (CHEN; GUESTRIN, 2016; KE et al., 2017; PROKHORENKOVA et al., 2018).

Predictive Modeling

Analyses are always performed with multiple repetitions and stratified partitions, reducing dependence on a single data split. In each repetition, the dataset is split into training (70%) and test (30%) sets, preserving the dropout-to-graduate ratio. In general, 30 repetitions are performed, each one with a different seed[2].

Within the training set, a new stratified split is made between the internal training and validation sets. This separation is adopted to avoid information leakage: hyperparameters must be tuned only on the internal training set, while validation is used for calibration and the selection of the decision threshold. In boosting models, the validation set may also serve for *early stopping*.

Hyperparameter tuning is done by random search with cross-validation on the internal training set. As selection criterion, *Average Precision* (BERGSTRA; BENGIO, 2012; SAITO; REHMSMEIER, 2015) is used, since it is more appropriate for scenarios with class imbalance and a focus on the correct identification of dropout.

Since the dropout class is the minority class, a weighting scheme must be applied during training, with weights proportional to the ratio of negatives to positives in the internal training set. This strategy aims to increase the penalization of errors in the minority class without resorting to resampling.

After training, the model's probabilities must be calibrated using isotonic regression on the validation set (ZADROZNY; ELKAN, 2002). The objective is to make these probabilities more reliable for later use. Also on the validation set, the decision *threshold* is defined, that is, the minimum probability above which a student is classified as a dropout.

[2]Seed is the value used to define the randomness of each partitioning of the dataset. Thus, by employing different seeds in the 30 repetitions, the evaluation considers variations in the composition of the training and test sets, reducing the influence of a single specific split.

The final decision must be produced by a weighted ensemble. The choice of the ensemble is based on the idea that different models can capture complementary patterns in a multifactorial problem such as dropout. Thus, each model receives a weight proportional to its performance in identifying the positive class on the validation set, measured by $F1_{pos}$. The student's final probability is calculated by the weighted average of the calibrated probabilities, and the ensemble's *threshold* is tuned separately.

The overall predictive modeling flow adopted in this work is summarized in Figure 3.8. The figure shows the sequence from the initial stratified partitioning of the data to hyperparameter tuning, calibration, threshold definition and final ensemble prediction.

Evaluation Strategy and Metrics for Predictive Modeling

Model evaluation is performed on the test set, which must be kept separate from all stages of tuning, calibration, and threshold definition. The reported metrics are calculated from the calibrated probabilities and thresholds defined exclusively on the validation set.

Different metrics must be used, since the dropout problem introduces class imbalance (SOKOLOVA; LAPALME, 2009). Accuracy is used as a general view of performance, but should not be taken in isolation as the main criterion, since it can excessively favor the majority class.

For a more adequate analysis of the class of interest, the F1-score must be used, including both its weighted version and the F1 of the positive class. $F1_{pos}$ is adopted as the main indicator to evaluate the model's ability to correctly identify dropout students.

The ROC curve and the area under the curve (ROC-AUC) must also be considered, which measure the capacity for separation between classes (FAWCETT, 2006). In addition, the Precision–Recall curve and its *Average Precision* are used, especially in imbalanced contexts (SAITO; REHMSMEIER, 2015).

Since the architecture uses probabilities as dropout risk estimates, the calibrated probabilistic quality must also be evaluated. For this purpose, the *Brier Score* is used (BRIER, 1950), which measures the mean squared error between the predicted probabilities and the observed outcomes. The lower this value, the better the alignment between the estimated risk and the observed outcome.

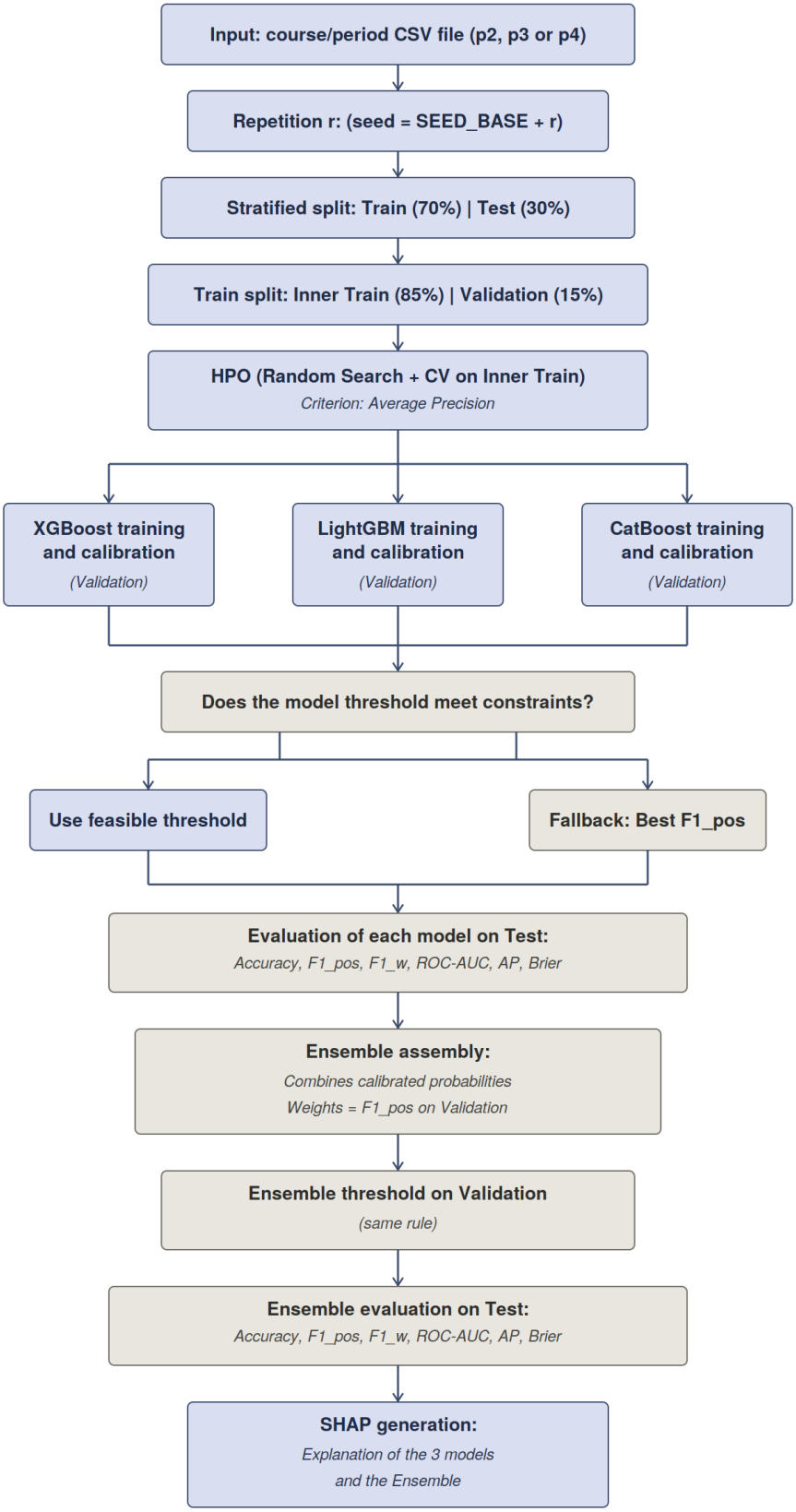

Figure 3.8: Predictive modeling flow.

Model Explainability – SHAP

In addition to risk prediction, EducAAr incorporates an explainability stage. The objective of this stage is to explain why the model assigns a given level of risk to each student, making the analysis more understandable and auditable.

As the main technique, SHAP (*SHapley Additive exPlanations*) was adopted (LUNDBERG; LEE, 2017). This choice is due to the fact that the method allows interpreting predictions at two levels. At the local level, it shows the contribution of each variable to the prediction of a specific case. At the global level, it allows aggregating these contributions and observing which factors weigh more in the model's behavior throughout the analyzed set.

In the adopted interpretation, positive SHAP values shift the prediction toward dropout, while negative values indicate association with retention.

In the case of the ensemble, the explanations are constructed in a way compatible with the very logic of model combination. First, SHAP values are computed for each base model using *TreeSHAP*, with a reference set sampled from the training data. Then, these values are aggregated using a weighted average, with the same weights as those assigned to the models in the ensemble. Thus, the final explanation reflects the joint contribution of the variables within the same logic employed in prediction.

Probabilistic calibration is also used to improve the reliability of the probabilities, but the SHAP explanations are derived directly from the base models, since calibration does not alter the internal contribution structure of the variables.

Results and Integration with the Simulation Panel

After modeling and explanation generation, the results are organized for use in the EducAAr panel. This stage is important because it transforms the experimental results into reusable artifacts within the architecture.

Among the items used by the panel are: the trained and calibrated ensemble, the consolidated metrics, the records by repetition, evaluation figures, SHAP values, the dataset corresponding to the explanations, and global rankings of the most important variables.

This organization allows the panel to present three types of reading: a global view of the factors most associated with dropout and retention; an individual analysis of instances; and simulations, in which hypothetical changes in the variables result in a new risk estimate accompanied by an explanation. The next chapter presents screenshots of the panel.

3.6.5 Knowledge Acquisition and Analysis Support

The results produced in the analysis layer are made available in an interactive panel, designed as the visualization and interpretation layer of the EducAAr architecture. This stage seeks to make the outputs of the architecture more accessible by gathering, in a single environment, information on predictive performance, global explainability, local explanations, and scenario simulation.

In this sense, the panel is not only an auxiliary interface, but a component of the proposed solution. It materializes the results produced by the ontology, machine learning models, probabilistic calibration, and SHAP explanations, organizing them into a format that can be consulted and interpreted more directly.

The panel organizes navigation by program and by the analyzed period. For each slice, consolidated indicators of ensemble performance are presented, such as $F1_{pos}$, AUC-ROC, *Average Precision*, and calibrated *Brier Score*, as well as the confidence interval for $F1_{pos}$, summary tables, confusion matrices, and performance curves. Information is also displayed on the best-observed run and the series of metrics throughout the repetitions, which allows evaluating not only the level of performance achieved but also the stability of the results.

On the explanatory axis, the panel presents a global analysis using SHAP values, including a ranking of the most influential variables and a *summary plot* that allows identifying, in each period, which attributes most contributed to shifting the prediction toward dropout or retention. With this, the interpretation is not limited to the final predicted value, but also incorporates evidence on the model's most relevant factors.

In addition to the global analysis, the system incorporates a scenario simulation module. In this environment, the user can input academic and contextual attributes of

a student profile – such as admission method, gender, ethnicity, quotas, performance by period, and links to scholarship or student assistance – and obtain the estimated dropout probability, the predicted class, and the corresponding local explanation. This explanation is presented through a graph and a summary of the factors that most increased or reduced the estimated risk, allowing a more concrete examination of how different combinations of attributes affect the ensemble's output.

The panel also has a module for explanation assisted by a LLM (Large Language Model). In this case, the language model is not used to perform prediction, access the institutional database, or produce new evidence. Its role is restricted to generating a textual synthesis from structured outputs previously produced by the architecture.

The prompt is automatically assembled from aggregated and controlled information, such as the program, the analyzed period, the model used, the most relevant variables according to SHAP, the direction of the signals associated with dropout and retention, and cautionary instructions against causal interpretations. Therefore, the LLM operates over a summarized representation of the explanatory results, rather than over raw student data.

The ontology contributes to this process by organizing the educational concepts and relationships that are later transformed into tabular variables and explanatory summaries. In this way, the prompt is restricted by the vocabulary and structure defined in the architecture, reducing the risk of producing interpretations disconnected from the modeled educational domain. Thus, the LLM-assisted explanation does not replace the specialist's analysis, but works as a textual layer over the technical results generated by EducAAr.

The objective of this stage is not to replace institutional interpretation or validate institutional decision-making, but to provide organized technical elements that support the analysis of dropout risk. Figure 3.9 represents the global panel for exploratory analysis, presenting the main explainability results used to support the interpretation of the model.

Figure 3.9: Explanatory panel for dropout risk analysis support (in portuguese).

3.6.6 Institutional Analysis Support Agent

The institutional analysis support agent is the component that connects the results produced by the architecture to possible institutional readings of dropout risk. While the analysis layer produces metrics, probabilities, and explanations, this component organizes these results so that programs, periods, or student profiles that demand greater attention can be identified.

Through the panel, academic teams can consult the results generated by the architecture and analyze patterns related to academic performance, admission, scholarships, student assistance, quotas, gender, and ethnicity. However, the architecture does not prescribe actions automatically and was not evaluated, in this dissertation, as an intervention tool with managers or academic teams.

Its role is to provide organized evidence that may support subsequent institutional analysis. Any pedagogical, administrative, or assistance-related action must depend on human interpretation, institutional context, and complementary information not captured by the model.

Therefore, the agent should be understood as a bridge between technical results

and institutional interpretation, not as an autonomous decision-making component.

3.6.7 Validation of the Second Cycle

The validation of the second cycle was structured to verify whether the architecture met the requirements defined for this stage, especially FR04 and FR05.

To evaluate FR04, controlled experiments were designed with multiple repetitions and data stratification. Predictive performance was measured using a complementary set of metrics, including accuracy, F1-Score, ROC-AUC, *Average Precision*, and Brier Score (SOKOLOVA; LAPALME, 2009; FAWCETT, 2006; SAITO; REHMSMEIER, 2015; BRIER, 1950). Beyond the mean values, the stability of the results and of the *threshold* across repetitions was observed.

For FR05, validation focused on the system's capacity to produce usable explanations. For this, global explanations were extracted from the variable rankings generated by SHAP (LUNDBERG; LEE, 2017), and these results were compared with patterns observed in the literature and with interpretations consistent with academic reality.

The technical integration between the produced results and the visualization panel was also considered, ensuring that metrics, probabilities, and explanations were displayed in a structured and coherent manner.

The results of this evaluation are presented in Chapter 4.

3.6.8 Synthesis of the Second Cycle

In the second cycle, EducAAr was improved to incorporate predictive analysis and explainability. The integration of classification models, probabilistic calibration, and SHAP enabled transforming academic data into risk estimates accompanied by interpretable explanations.

With this, the architecture came to gather, within a single structure, data organization, prediction, explanation, and presentation of results. This expansion responds to the requirements defined for this stage of the research.

3.7 Chapter Conclusion

This chapter presented the methodology adopted for developing the EducAAr architecture, structured into two cycles of *Design Science Research*. Based on the gaps identified in the literature review, the architecture requirements were defined and the procedures of construction, expansion, and validation of the artifact were described.

The evaluation of the architecture involved both structural and computational procedures. The structural evaluation concerned the ontological modeling, consistency verification, and querying of the integrated data. The computational evaluation involved the preparation of datasets, construction of predictive models, analysis of performance metrics, probabilistic calibration, and interpretation of the explanations produced by SHAP.

The first cycle focused on the organization and integration of data through the ontology. The second cycle expanded on this foundation by incorporating machine learning models, explainability techniques, and visualization resources for analysis support.

In the next chapter, the results of the architecture evaluation are presented, based on the application of these components to real higher education data.

4 Evaluation of the EducAAr Architecture

As described in Chapter 3, the development of EducAAr was conducted based on Design Science Research (HEVNER, 2007), in two main cycles. This chapter presents the evaluation carried out in each of them.

In what follows, the two cycles are presented separately, with the description of the objective, the adopted solution, and the obtained results.

4.1 Evaluation of the First DSR Cycle – Modeling and Integration

The first cycle of EducAAr aimed to organize educational data that were originally dispersed across different institutional databases. Information on academic history, admission methods, sociodemographic characteristics, and student assistance was scattered across non-integrated, heterogeneous databases. This lack of integration and heterogeneity hindered broader analyses of students' trajectories and the patterns associated with dropout.

An ontology was then developed as the canonical model of the architecture, gathering the concepts, attributes, and relationships of the educational domain into a single structure. Based on this canonical model, it was possible to perform queries, infer new relationships among the data, and conduct exploratory analyses of these integrated data.

The results of this cycle were published in (AZY et al., 2024), in a study conducted in the context of the Information Systems program at UFJF. The evaluation used anonymized data in compliance with the LGPD (BRASIL, 2018), with a temporal range from 2013 to 2023. The analyzed set comprised 439 students and 13.518 instances integrated into the ontology presented in Section 3.5, including information on grades, student support, and admissions by quota.

4.1.1 Results

The evaluation of the first cycle was conducted along two fronts. The first consisted of verifying the structural consistency of the ontology. The second focused on extracting knowledge from the instantiated data.

In the structural validation, the HermiT *reasoner* was used to identify possible conflicts between classes, cardinality violations, or modeling inconsistencies. The execution did not identify structural errors, indicating that the ontology was logically coherent and capable of supporting queries and inferences, as shown in Figure 3.3.

After this stage, SPARQL queries were performed on the integrated data. The objective was to verify whether the constructed structure enabled the retrieval of relevant information about students' academic trajectories and the identification of patterns associated with dropout and completion outcomes. We present below some examples of relevant information that could be extracted from the integrated data, and which were not possible before the canonical ontological model – the discovery of new relationships and the integration of heterogeneous data.

SPARQL1: Academic Performance of dropout Students

```
SELECT ?subclasse (COUNT(?student)
AS ?count) WHERE { ?student
rdf:type ont:Student. ?student
ont:StudentHasStatus ?status.
?status rdf:type ont:Withdrawn.
?student
ont:StudentHasAcademicPerformance
?performance.

?performance rdf:type ?subclasse.
?subclasse rdfs:subClassOf
ont:AcademicPerformance.}GROUP BY
?subclasse
```

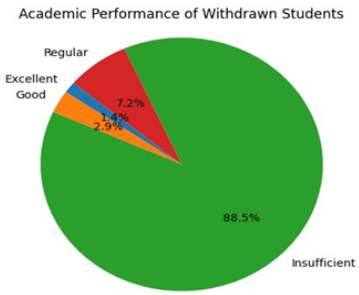

Figure 4 - Academic performance of dropout students

Figure 4.1: Distribution of academic performance among dropout students (AZY et al., 2024).

Figure 4.1 presents the distribution of academic performance among students classified as dropouts. About 88.5% of these students were categorized as *Insufficient* performance, while 7.2% were classified as *Regular*, 2.9% as *Good*, and 1.4% as *Excellent*. This result shows a strong association between low academic performance and dropout, reinforcing the relevance of this type of variable for analysis of the student trajectory.

SPARQL2: Students who did not receive student assistance

```
SELECT ?subclasse (COUNT(?student) AS
?count)WHERE { ?student rdf:type
ont:Student. ?student
ont:StudentReceivedStudentAssistance
?assistance. ?assistance rdf:type
ont:NotReceived.

  ?student ont:StudentHasStatus ?status.
?status rdf:type ?subclasse. ?subclasse
rdfs:subClassOf ont:Status.

FILTER (?subclasse = ont:Completed ||
?subclasse = ont:Withdrawn)}GROUP BY
?subclasse
```

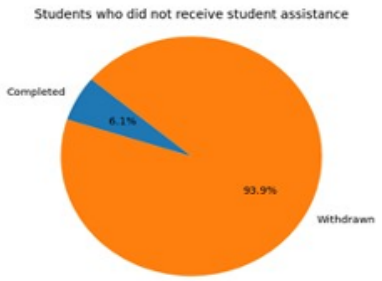

Figure 5 - Without financial support

Figure 4.2: Distribution of academic outcomes among students without student assistance.

Figure 4.2 presents the results related to student assistance. In this analysis, assistance was defined as the receipt of any institutional benefit of a financial nature throughout the program, including aid, food vouchers, student passes, or combinations of these benefits.

The results showed an important difference between the groups. Among students who received assistance, a higher proportion completed than in the group without assistance. Among those without assistance, dropout reached 93.9%, with only 6.1% of completion. This result suggests that student assistance is associated with retention and is a relevant variable for institutional analyses, especially when considered alongside academic indicators.

In general terms, the SPARQL queries showed that the ontology was able to integrate different types of data and to retrieve useful information for the analysis of dropout. The main contribution of this stage was not only to show previously known associations, but also to organize them into a structured base capable of being queried, validated, and reused in later stages of the architecture.

4.1.2 Lessons learned and limitations

The results of the first cycle show that the ontological modeling was effective in organizing academic data from different sources into a single structure. The use of the *reasoner* ensured logical consistency and enabled the inference of new properties and relationships, as illustrated in Figure 3.3, while the SPARQL queries allowed identification of relevant

patterns associated with dropout, with emphasis on academic performance and student assistance. With this, the first cycle partially answers the research question by demonstrating that it is possible to integrate and structure heterogeneous educational data into a canonical model capable of supporting analyses of student trajectories and factors related to dropout.

Despite these results, the stage still presents clear limitations. The analyses performed are retrospective, that is, focused on students who have already dropped out or graduated. This means that the first cycle did not yet produce individual risk estimates or provide predictive mechanisms.

Another limitation is in the way some variables are represented. Student assistance, for example, still appeared in an aggregated form, without detailing the duration, continuity, or the moment of granting. Similarly, the initial representation of quotas remained simplified, given the diversity of modalities in the institutional context.

It was also observed that some variables presented low contribution in the exploratory analyses, such as the student's nativity, indicating the need to reassess their relevance or their form of use. In addition, the analyses of this stage tended to obscure particularities between programs, reinforcing the importance of more specific cuts in later stages of the research.

Thus, the first cycle fulfilled its role by demonstrating that it was possible to gather, organize, and query educational data in an integrated manner. At the same time, its limitations made evident the need to expand the architecture to incorporate predictive mechanisms and interpretable explanations, which directly motivated the second cycle of EducAAr.

4.2 Evaluation of the Second DSR Cycle – ML and XAI

The second cycle evaluated the capacity of the EducAAr architecture to produce dropout risk estimates and interpretable explanations for these estimates, as described in Chapter 3. The evaluation was conducted across 16 undergraduate courses at the Federal

University of Juiz de Fora, totaling 7.731 students in the considered sample. For the construction of the datasets used in this cycle, institutional data on students, academic history, and scholarships were integrated.

Admissions originating from ABI (Basic Admission Area) and BI (Interdisciplinary Bachelor's) courses were not considered. Although these paths are part of the university trajectory, their curricular organization differs from that of the specific courses analyzed, hindering a direct comparison among students

In this chapter, the results are presented with a cut based on the Ciência da Computação - Noturno program, chosen because it aligns with the research scope, provides an adequate sample size, and shows greater consistency in the records available in the adopted cut. The results of the other programs are available in a public GitHub repository GitHub.

4.2.1 Dataset and execution conditions

The experiments were conducted with institutional data from the Federal University of Juiz de Fora. To obtain access to the data required for this research, the study was submitted to the UFJF Research Ethics Committee, which is responsible for evaluating and authorizing research involving institutional data related to individuals. Although the final datasets used in this dissertation did not contain directly identifiable student information, the access request followed the formal institutional approval process required for this type of research.

The approval process lasted approximately eight months until the final authorization was granted by the Ethics Committee (CAAE **86818225.9.0000.5147**), in compliance with the LGPD (BRASIL, 2018). During this process, successive interactions with the committee were necessary to clarify the research protocol, the data access procedures, the anonymization strategy, and the conditions under which the institutional data would be handled. These interactions were important to define the limits of data use and to ensure that the analysis would be conducted only within the authorized scope.

After approval, access to the data was provided in a restricted research environment. The extraction, organization, and preparation of the data, from the instantiation

of the ontology to the generation of the tabular datasets, were carried out within this controlled environment. Only the anonymized tabular files used for model training and explanation generation were authorized to leave this environment.

Table 4.1 presents the distribution of students by program in the second cycle, based on the $p2$ window datasets[3].

Table 4.1: Number of students per program considered in the second cycle, based on the $p2$ window datasets (in portuguese).

Program	**Number of students**
Ciências Biológicas - Bacharelado	181
Ciências Biológicas - Licenciatura	168
Ciências Econômicas - Integral	353
Ciências Econômicas - Noturno	319
Ciência da Computação - Integral	76
Ciência da Computação - Noturno	203
Direito - Integral	771
Direito - Noturno	730
Enfermagem	504
Engenharia Civil	850
Engenharia de Produção	420
Farmácia	706
Medicina	1302
Pedagogia	549
Psicologia	413
Sistemas de Informação	186
Total	**7731**

The evaluation focused on windows $p2$, $p3$, and $p4$, since they correspond to early stages of the academic trajectory in which it is already possible to observe signs related to dropout risk. As the main evidence of this section, the Ciência da Computação - Noturno program was considered, with 203 students in the temporal cut from 2010 to 2023.

Regarding admission, only SISU[4] and PISM[5] were considered. This choice was made because these are the predominant admission methods for students during the analyzed period and exhibit greater regularity in institutional records. Furthermore, the grades from these two systems could be used on the same scale, without the need for

[3]The $p2$ window corresponds to the second period of the program; $p3$ and $p4$ correspond, respectively, to the third and fourth periods.

[4]Sistema de Seleção Unificada (Unified Selection System).

[5]Serial assessment process used at UFJF.

additional normalization. Modalities such as transfers, graduate admissions, or vacant slots due to program changes were not included, as these cases do not capture the entire academic trajectory in the records used and occur infrequently.

4.2.2 Experimental protocol applied to the Ciência da Computação - Noturno program

The experimental protocol was executed separately on the windows $p2$, $p3$, and $p4$, with 30 repetitions per window and a random seed varied at each execution. In each repetition, the data were partitioned into training and test in a stratified manner. From the training set, an additional split was performed for hyperparameter tuning, calibration, and threshold definition, as described in Chapter 3.

4.2.3 ML Results (CC Noturno, $p2$)

Table 4.2 presents the aggregated results (*mean* and *standard deviation*) of the 30 repetitions in period $p2$ for the three base models and for the ensemble.

Table 4.2: Aggregated results in the Ciência da Computação - Noturno program ($p2$, 30 repetitions).

Model	$\mathbf{F1_{pos}}$	ROC-AUC	Avg. Precision	Brier (cal.)	Threshold
XGBoost	0.824 ± 0.046	0.840 ± 0.046	0.881 ± 0.040	0.163 ± 0.036	0.680 ± 0.186
LightGBM	**0.825 ± 0.039**	0.832 ± 0.045	0.872 ± 0.037	0.167 ± 0.034	0.737 ± 0.186
CatBoost	0.817 ± 0.047	0.844 ± 0.053	0.880 ± 0.047	0.161 ± 0.038	0.705 ± 0.201
Ensemble	0.814 ± 0.053	**0.863 ± 0.041**	**0.911 ± 0.031**	**0.151 ± 0.030**	0.598 ± 0.173

In period $p2$, LightGBM presented the highest mean value of $F1_{pos}$. This metric is important because it summarizes the balance between precision and recall for the positive class, indicating how well the model correctly identifies dropout students. The ensemble, in turn, achieved the highest ROC-AUC, the highest *Average Precision*, and the lowest calibrated Brier score. This is relevant because ROC-AUC indicates better separation between dropouts and graduates; *Average Precision* is especially useful in imbalanced datasets with a focus on the positive class; and lower Brier scores indicate more reliable probabilities after calibration.

In the best execution observed for the ensemble, corresponding to repetition 2, the

following were obtained: $F1_{pos}$ = 0.892, ROC-AUC= 0.903, *Average Precision*= 0.959, and calibrated Brier= 0.104, with *threshold*= 0.521 and weights of 0.321 (XGBoost), 0.333 (LightGBM), and 0.346 (CatBoost).

In general terms, the results of $p2$ show that, already in the first periods, the ensemble demonstrates good capacity for separating the classes and good probabilistic quality. This is important for institutional analysis, since the architecture not only classifies students but also provides more consistent risk probabilities.

4.2.4 ML Results (CC Noturno, $p3$)

Table 4.3 presents the aggregated results (*mean* and *standard deviation*) of the 30 repetitions in period $p3$ for the three base models and for the ensemble.

Table 4.3: Aggregated results in the Ciência da Computação - Noturno program ($p3$, 30 repetitions).

Model	**$F1_{pos}$**	**ROC-AUC**	**Avg. Precision**	**Brier (cal.)**	**Threshold**
XGBoost	0.809 ± 0.052	0.833 ± 0.057	0.857 ± 0.053	0.176 ± 0.045	0.670 ± 0.189
LightGBM	**0.819 ± 0.062**	0.848 ± 0.050	0.869 ± 0.043	0.164 ± 0.038	0.639 ± 0.178
CatBoost	0.791 ± 0.070	0.839 ± 0.044	0.862 ± 0.045	0.173 ± 0.040	0.694 ± 0.209
Ensemble	0.809 ± 0.066	**0.861 ± 0.042**	**0.897 ± 0.036**	**0.160 ± 0.037**	0.597 ± 0.163

In period $p3$, LightGBM again presented the highest mean value of $F1_{pos}$, indicating a slight advantage in the direct identification of dropout students. Even so, the ensemble maintained the lead in ROC-AUC, *Average Precision*, and calibrated Brier. The pattern observed in $p2$, therefore, was maintained: the base models remain competitive at identifying the positive class, but the ensemble remains stronger in metrics that assess separation between classes and the reliability of probabilities.

In the best execution observed for the ensemble, corresponding to repetition 11, the following were obtained: $F1_{pos}$ = 0.906, ROC-AUC= 0.945, *Average Precision*= 0.966, and calibrated Brier= 0.087, with *threshold*= 0.534 and weights of 0.331 (XGBoost), 0.334 (LightGBM), and 0.334 (CatBoost).

With the incorporation of more academic history, that is, as the analysis advances to subsequent periods and incorporates more variables, the general behavior of the system does not change. The ensemble continues to show greater stability in the metrics most

useful for dropout risk analysis, especially because it combines strong discrimination with calibrated probabilities.

4.2.5 ML Results (CC Noturno, $p4$)

Table 4.4 presents the aggregated results (*mean* and *standard deviation*) of the 30 repetitions in period $p4$ for the three base models and for the ensemble.

Table 4.4: Aggregated results in the Ciência da Computação - Noturno program ($p4$, 30 repetitions).

Model	$\mathbf{F1_{pos}}$	ROC-AUC	Avg. Precision	Brier (cal.)	Threshold
XGBoost	0.824 ± 0.048	0.881 ± 0.046	0.883 ± 0.049	0.143 ± 0.035	0.608 ± 0.182
LightGBM	0.804 ± 0.074	0.872 ± 0.039	0.869 ± 0.042	0.153 ± 0.033	0.740 ± 0.204
CatBoost	**0.832 ± 0.050**	0.875 ± 0.042	0.872 ± 0.047	0.145 ± 0.036	0.738 ± 0.215
Ensemble	0.816 ± 0.053	**0.899 ± 0.032**	**0.910 ± 0.033**	**0.135 ± 0.029**	0.655 ± 0.160

In period $p4$, CatBoost presented the highest mean value of $F1_{pos}$, indicating a better balance between precision and recall of the dropout class in this window. Even so, the ensemble again achieved the highest ROC-AUC, the highest *Average Precision*, and the lowest calibrated Brier score. This shows that the pattern observed in the previous windows is also maintained in $p4$: the base models alternate in the lead in $F1_{pos}$, while the ensemble stands out more consistently in the metrics that indicate the capacity to separate the groups and produce more adjusted probabilities.

In the best execution observed for the ensemble, corresponding to repetition 20, the following were obtained: $F1_{pos} = 0.929$, ROC-AUC= 0.963, and calibrated Brier= 0.059, with *threshold*= 0.727 and weights of 0.328 (XGBoost), 0.328 (LightGBM), and 0.345 (CatBoost). In this repetition, the *Average Precision* was 0.944, while the highest value observed for this metric throughout the repetitions was 0.953.

With this, the last window reinforces the reading already observed in $p2$ and $p3$: the combination of the models produces more reliable probabilities and more stable behavior for institutional use.

4.2.6 Synthesis of predictive results in the windows $p2$, $p3$, and $p4$

The results from windows $p2$, $p3$, and $p4$ are consistent with the validation strategy defined in Chapter 3. In general terms, the base models alternate in the lead for $F1_{pos}$, while the ensemble outperforms them more frequently in ROC-AUC, *Average Precision*, and Brier Score.

This result is relevant because $F1_{pos}$ shows the capacity to correctly identify the class of greatest interest, which in this case is that of dropout students. ROC-AUC, in turn, shows how well the model separates dropouts from graduates. *Average Precision* reinforces this reading in an imbalanced scenario, and the Brier Score indicates whether the predicted probabilities are well calibrated to the observed outcomes. Since the institutional objective is not only to label students, but also to work with more reliable risk estimates, the ensemble's better performance on these metrics reinforces its practical usefulness for analysis.

This behavior indicates that the ensemble offers greater robustness for institutional analysis, mainly because the objective is not only to classify correctly, but also to produce more reliable probabilities to support dropout risk interpretation. In addition, the weighted combination of the models reduces dependence on a single algorithm, thereby reducing oscillations associated with data partitioning.

Together, these results show that the predictive module of EducAAr meets requirement FR04, as the architecture analyzed structured data and produced predictions with consistent performance across the evaluated windows. At the same time, these results provide a basis for the next stage, dedicated to the explanatory analysis of the factors most associated with dropout risk and retention.

4.3 Ensemble Explainability (XAI)

Dropout prediction, in isolation, is not enough to support institutional analysis. Knowing that a student has a high probability of dropping out does not, by itself, answer the main point of interest of the analysis: understanding why this risk was assigned. For the

model's results to support institutional analysis, it is necessary to identify which factors contributed to this classification. It is precisely in this gap that explainability fits, by allowing the analysis of which characteristics most influenced the model's decision.

In this stage, the analysis turns to how the ensemble supports its dropout estimates. For this, SHAP values were used (LUNDBERG; LEE, 2017), calculated separately for each base model and aggregated according to the combination rule described in Chapter 3. In this interpretation, positive SHAP values shift the prediction toward dropout, while negative values indicate an association with retention.

The analysis was organized into two parts. First, the global view is presented, focusing on the most influential variables in each window and how this pattern changes across $p2$, $p3$, and $p4$. Then, the local view is presented, showing how these variables appear in cases with higher and lower dropout propensities.

4.3.1 Global view related to the evolution of factors between $p2$, $p3$, and $p4$

$p2$: greater weight of more immediate performance

In $p2$, the ensemble's explanation focuses mainly on the academic performance of the first two periods. The most important variables were *periodo_2_disciplinas_aprovadas*, *periodo_1_disciplinas_aprovadas*, *nota_ingresso*, *periodo_2_bolsa_remunerada*, and *periodo_2_disciplinas_reprovadas*. Among them, *periodo_2_disciplinas_aprovadas* appears in first place, indicating the current period's weight in differentiating between risk and retention.

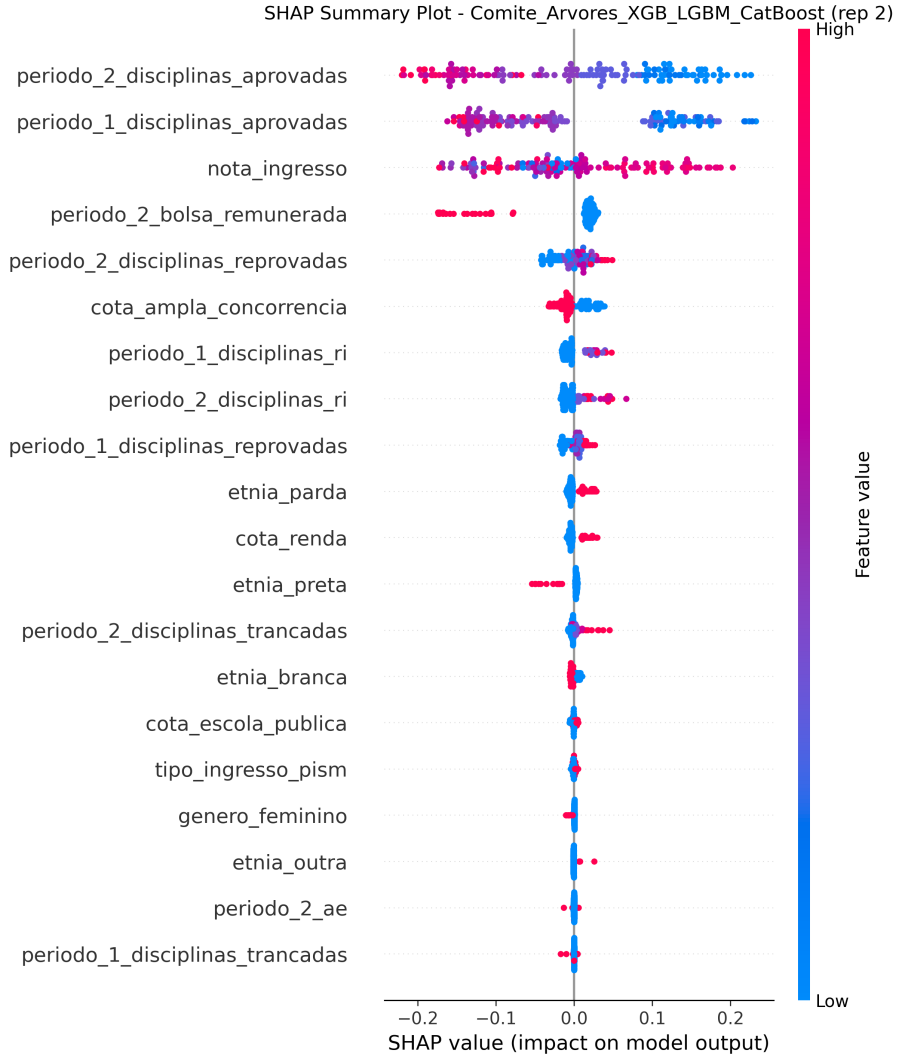

Figure 4.3: Global SHAP summary of the ensemble in period $p2$ (in portuguese).

The reading of $p2$ is relatively direct. More approvals in *periodo_2_disciplinas_aprovadas* and *periodo_1_disciplinas_aprovadas* shift the estimate toward retention, while failures in *periodo_2_disciplinas_reprovadas* push the prediction toward dropout. The presence of *periodo_2_bolsa_remunerada* also appears to be associated with lower risk, though with less impact than the academic variables. The variable *nota_ingresso* appears among the most relevant, but without the same stability as that observed for approvals and failures.

This first cut already shows an important pattern: even at the beginning of the trajectory, the ensemble relies mainly on concrete signals of academic performance.

*p*3: stronger entry of failures due to insufficient attendance

In *p*3, the explanation no longer focuses solely on immediate approvals and failures, and begins to incorporate failures due to insufficient attendance more strongly. The variables of greatest importance were *periodo_3_disciplinas_ri*, *periodo_1_disciplinas_aprovadas*, *periodo_3_disciplinas_aprovadas*, *periodo_2_disciplinas_aprovadas*, and *periodo_3_bolsa_remunerada*. The main change in relation to *p*2 is in *periodo_3_disciplinas_ri*, which moves to occupy the first position in the global ranking.

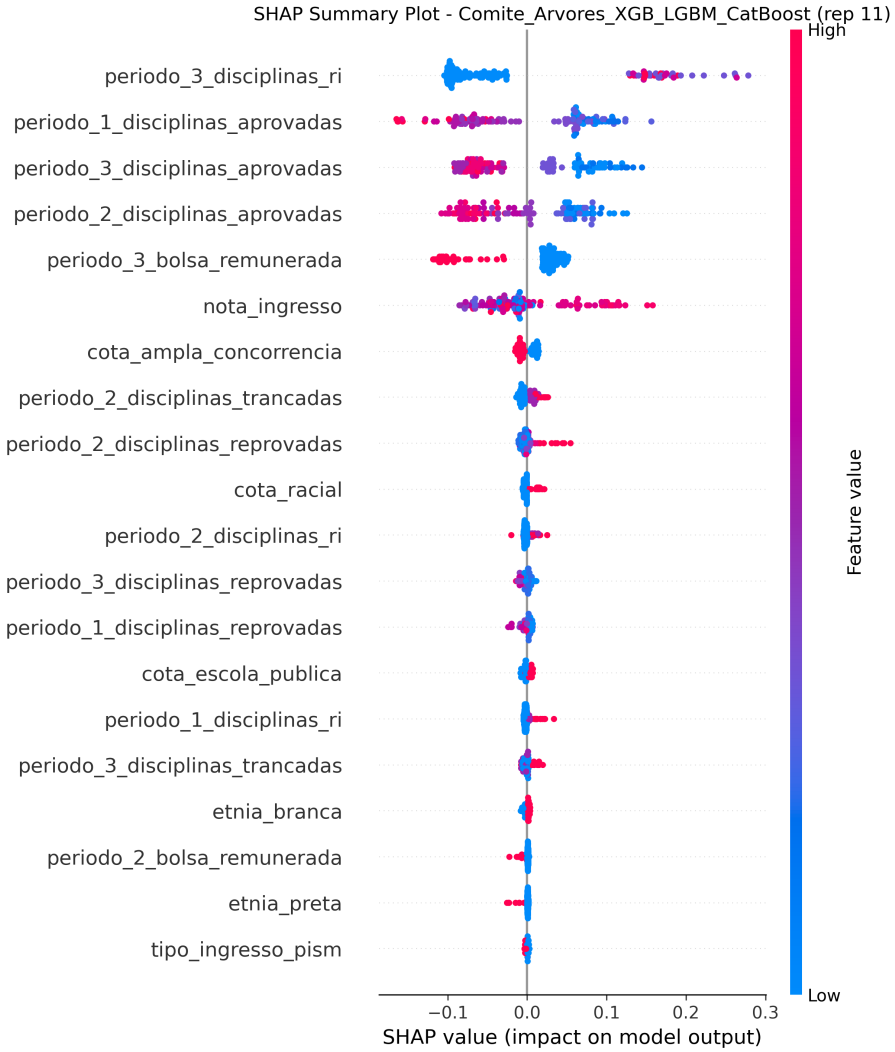

Figure 4.4: Global SHAP summary of the ensemble in period *p*3 (in portuguese).

At this point, the model begins to more clearly separate students who maintain attendance and approval from those who show signs of disengagement from academic activities. The presence of *periodo_3_disciplinas_ri* increases the estimated risk, whereas its absence reduces it. At the same time, approvals in *periodo_3_disciplinas_aprovadas*,

periodo_2_disciplinas_aprovadas, and *periodo_1_disciplinas_aprovadas* continue to play a relevant role, showing that accumulated performance still structures the explanation.

The presence of *periodo_3_bolsa_remunerada* also appears associated with retention. Even so, the most striking point in $p3$ is the entry of failure due to insufficient attendance as the variable of greatest weight.

$p4$: more evident accumulated trajectory

In $p4$, the ensemble's explanation comes to depend even more on the accumulated academic trajectory. The most important variables were *periodo_4_disciplinas_aprovadas*, *periodo_3_disciplinas_ri*, *periodo_1_disciplinas_aprovadas*, *nota_ingresso*, and *periodo_3_disciplinas_aprovadas*. The first position is now occupied by *periodo_4_disciplinas_aprovadas*, indicating that the performance of the current period is gaining even more centrality.

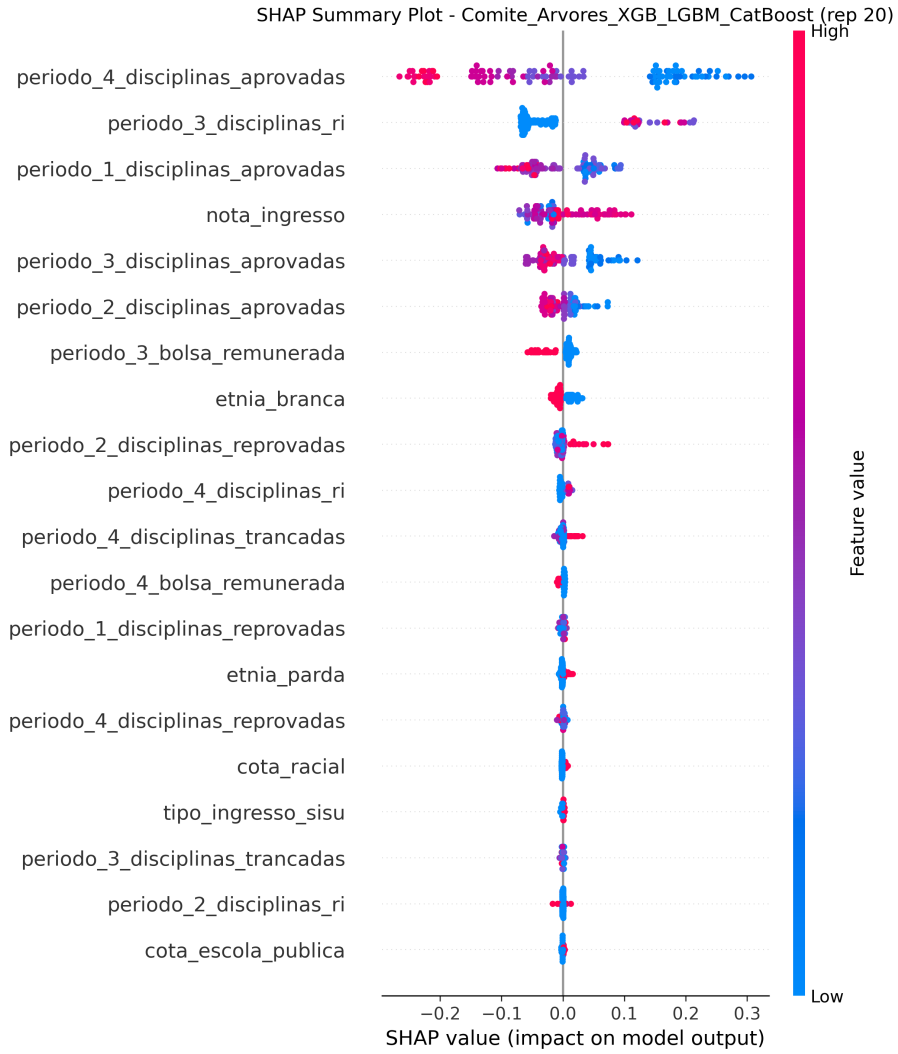

Figure 4.5: Global SHAP summary of the ensemble in period $p4$ (in portuguese).

Here, what weighs most is the combination between what happens in $p4$ itself and the accumulated signals from previous periods. Few approvals in *periodo_4_disciplinas_aprovadas* raise the risk, while more approvals reduce it. At the same time, *periodo_3_disciplinas_ri* remains among the most influential variables, showing that the failures due to insufficient attendance recorded in the previous period continue to reverberate in the model's explanation. The variables *periodo_1_disciplinas_aprovadas* and *periodo_3_disciplinas_aprovadas* indicate that the approval history continues to accompany the student's trajectory.

In this window, the presence of *nota_ingresso* remains relevant, without shifting the main focus of the explanation, which remains on the academic variables.

Global comparison among the periods

Analyzed together, the three periods show a clear pattern. The ensemble primarily supports its estimates using variables linked to academic performance. In $p2$, the greatest weight falls on more immediate approvals and failures. In $p3$, failure due to insufficient attendance gains prominence and begins to distinguish more clearly between students with lower risk and those showing stronger signs of dropout risk. In $p4$, performance in the current period appears at the top, but is already connected with accumulated signals from previous periods.

In other words, what changes between the windows is not the axis of the explanation, but the type of signal that gains more strength at each moment of the trajectory.

4.3.2 Local view: typical patterns in individual cases

After the global reading, it is important to observe how these variables appear in concrete cases. For this, examples of higher and lower propensity to drop out were selected in each period. The objective here is not to exhaustively describe each chart, but to show how the ensemble organizes, in individual situations, the main signals of risk and retention.

Profiles of higher and lower propensity to dropout

Window $p2$**.** Figures 4.6 and 4.7 show two contrasting cases in $p2$. In the case of higher propensity to dropout, the prediction starts from $E[f(X)] = 0.643$ and reaches $f(x) = 0.916$. What weighs the most is the low performance in the first two periods, especially in *periodo_2_disciplinas_aprovadas*= 0 and *periodo_1_disciplinas_aprovadas*= 2. The presence of *periodo_1_disciplinas_reprovadas*= 5 reinforces this picture.

In the case of lower propensity, the prediction drops to $f(x) = 0.113$. In this profile, the highlights are *periodo_2_disciplinas_aprovadas*= 7, *periodo_1_disciplinas_aprovadas*= 7, and *periodo_2_bolsa_remunerada*= 1, in addition to the absence of failures. The contrast between the two cases is clear: in $p2$, students with higher risk already present very weak performance from the very beginning.

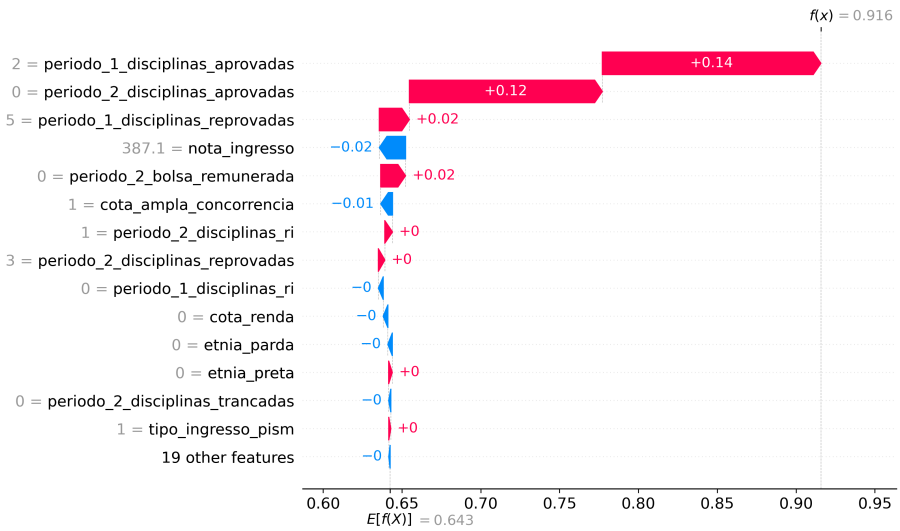

Figure 4.6: Local SHAP explanation for a student with higher propensity to dropout in period $p2$ (in portuguese).

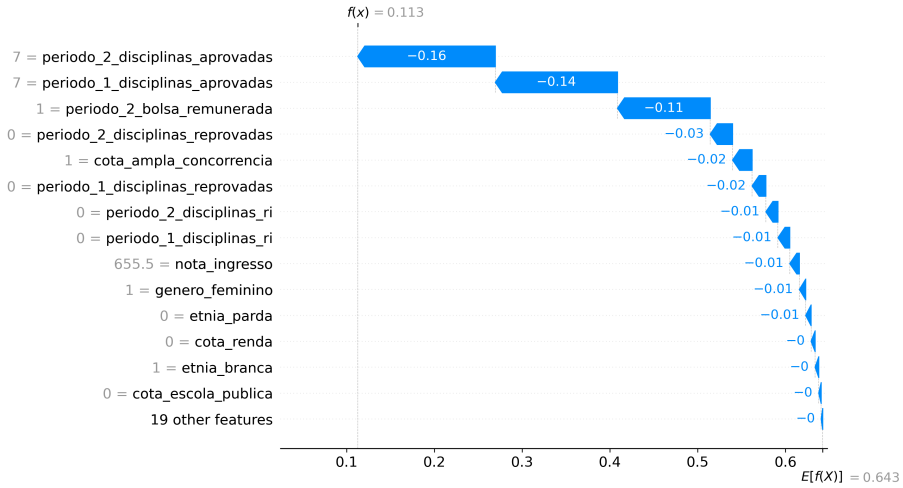

Figure 4.7: Local SHAP explanation for a student with a lower propensity to drop out in period $p2$ (in portuguese).

Window $p3$**.** In $p3$, the difference between the cases appears more strongly in failure due to insufficient attendance. In the profile of higher propensity to dropout (Figure 4.8), the prediction rises from $E[f(X)] = 0.602$ to $f(x) = 0.909$. The main factor is *periodo_3_disciplinas_ri*$= 1$, which shifts the estimate significantly toward risk. In addition, signs of only moderate performance appear, such as *periodo_3_disciplinas_aprovadas*$= 2$ and the absence of a paid scholarship in $p3$.

In the case of lower propensity (Figure 4.9), the prediction drops to $f(x) = 0.250$. The absence of *periodo_3_disciplinas_ri*, the good performance in *periodo_2_disciplinas_aprovadas*$= 5$ and *periodo_3_disciplinas_aprovadas*$= 3$, in addition to a *nota_ingresso*$=$ 575.7, help sustain the movement toward retention. Here, the local reading follows the pattern observed in the global analysis: what most separates the profiles is the presence or absence of failure due to insufficient attendance in the current period.

Window $p4$**.** In $p4$, the contrast between the cases becomes even stronger. In the profile of higher propensity to dropout (Figure 4.10), the prediction rises from $E[f(X)] = 0.607$ to $f(x) = 0.973$. The strongest factor is *periodo_4_disciplinas_aprovadas*$= 0$, which indicates the absence of approvals in the current period. Next, *periodo_3_disciplinas_ri*$= 2$ also weighs heavily, showing that the accumulated failures due to insufficient attendance continue to reverberate in the estimate. The rest of the history maintains the same direction: few approvals and fragile performance throughout the trajectory.

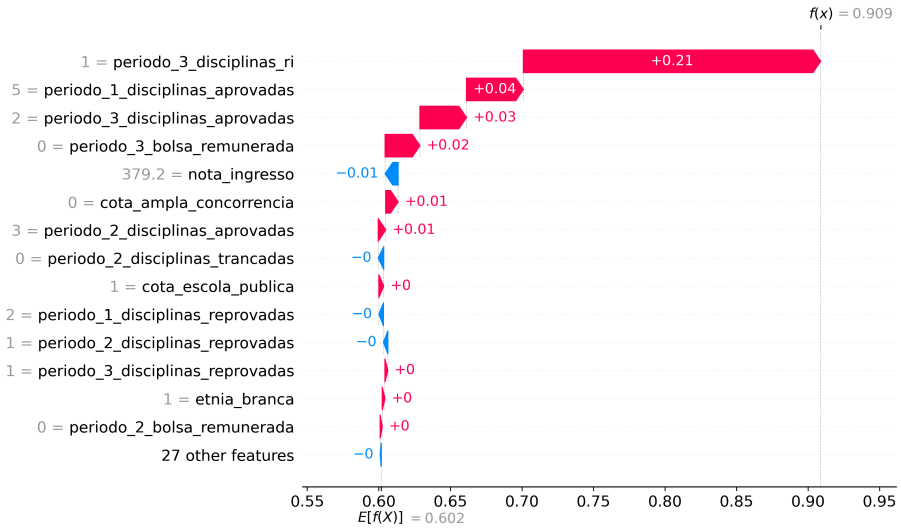

Figure 4.8: Local SHAP explanation for a student with higher propensity to dropout in period $p3$ (in portuguese).

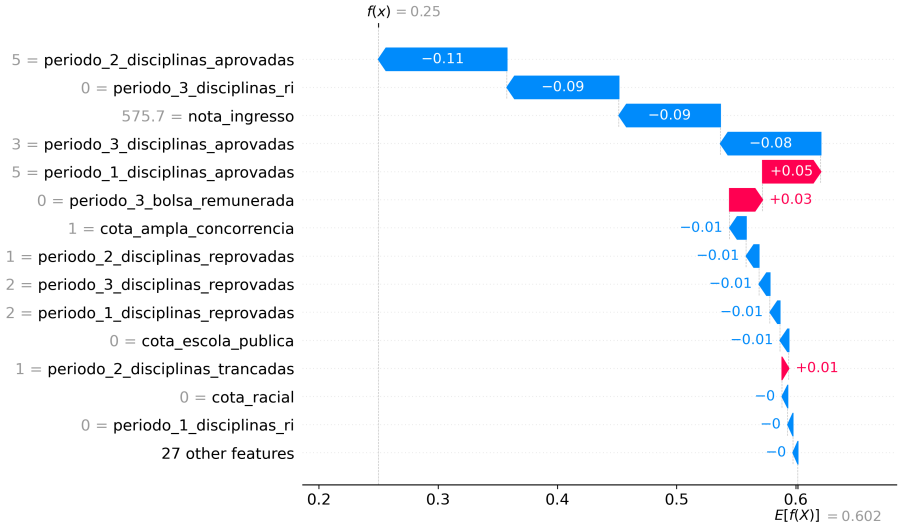

Figure 4.9: Local SHAP explanation for a student with a lower propensity to drop out in period $p3$ (in portuguese).

In the case of lower propensity (Figure 4.11), the prediction drops to $f(x) = 0.110$. Here, what sustains retention is a positive and continuous history, with emphasis on *periodo_4_disciplinas_aprovadas*$= 5$, absence of *periodo_3_disciplinas_ri*, high approvals in previous periods, and presence of paid scholarship in $p3$.

On $p4$, the current period remains decisive, but its reading already appears clearly articulated with the accumulated history of previous periods.

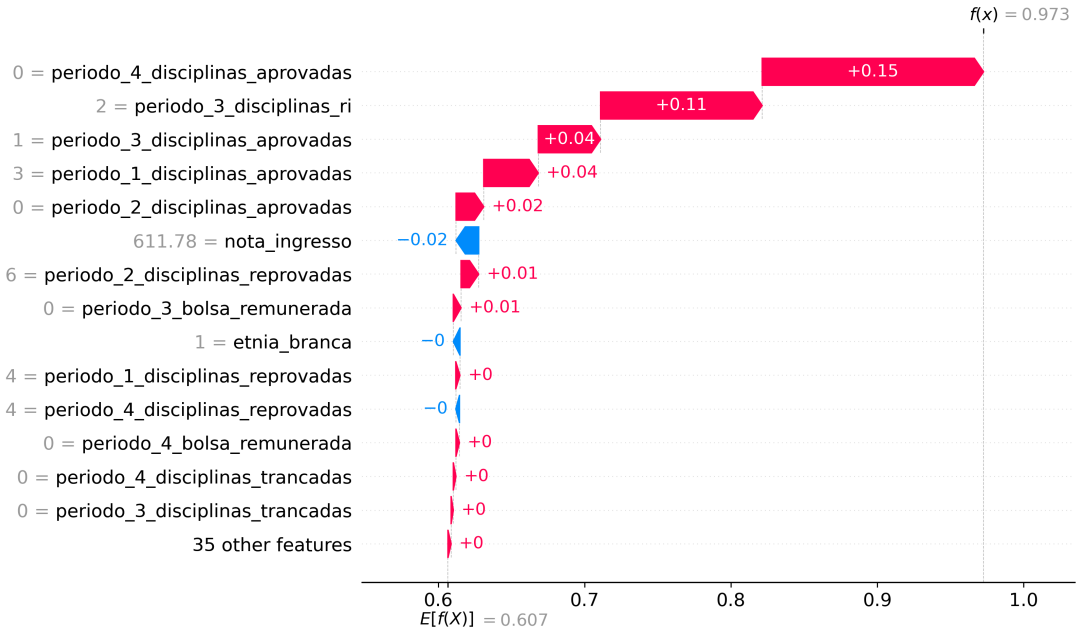

Figure 4.10: Local SHAP explanation for a student with higher propensity to dropout in period $p4$ (in portuguese).

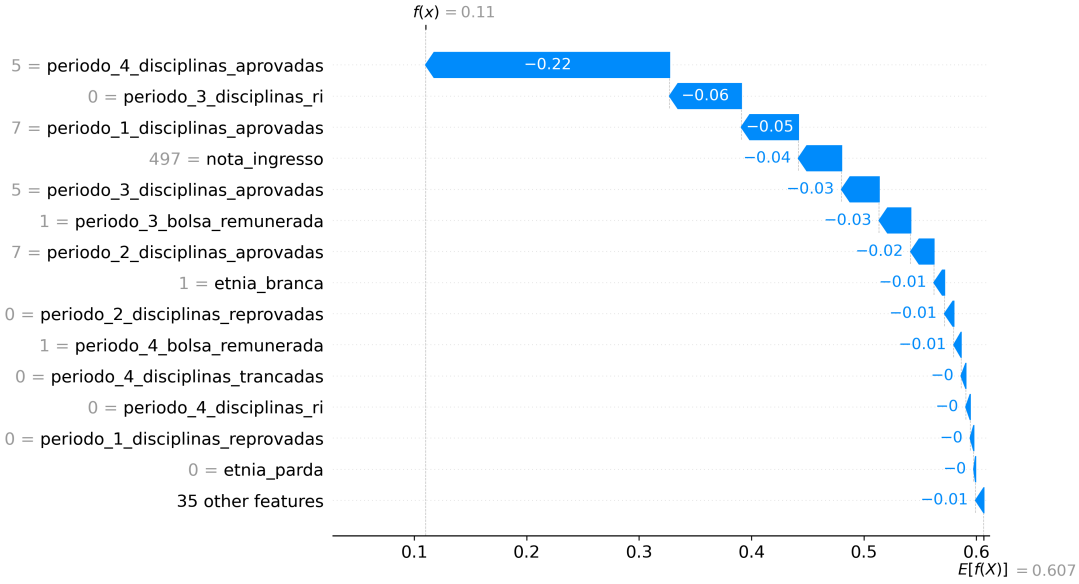

Figure 4.11: Local SHAP explanation for a student with lower propensity to dropout in period $p4$ (in portuguese).

4.3.3 Discussion

Both the global analysis and the local reading point in the same direction. The ensemble explains its estimates primarily based on the student's academic performance, with emphasis on approvals, failures, and failures due to insufficient attendance. Variables of admission and profile continue to appear in the explanations, but with a complementary role.

As the window incorporates more history, the explanation does not change its axis; it gains depth. Risk and retention continue to be differentiated mainly by the academic trajectory built in the first periods of the program.

With this, the second cycle also meets the requirement FR05 in the program analyzed in greater detail, since the architecture provides interpretable explanations for the predictions. Subsequently, this reading is broadened through the triangulation of the explainability results across the 16 programs considered in the second cycle.

4.4 Triangulation of explainability results in the 16 programs

After a detailed analysis of the ensemble's explainability in the Ciência da Computação - Noturno program, a triangulation of the results observed across the 16 course offerings considered in the second cycle was conducted. The objective was to verify to what extent the pattern identified in the case study was also maintained in the analyzed set, and to identify relevant differences between programs, areas, and observation windows.

This triangulation was conducted based on two complementary elements: the global rankings of variable importance and the direction of effects observed in the SHAP *summary plots* in the windows $p2$, $p3$, and $p4$. While the previous analysis allowed for a deeper examination of the ensemble's explanatory logic in a specific program, this section broadens that reading and shows what is repeated, what changes, and what appears only in particular contexts.

In general terms, the triangulation confirmed the pattern already observed in Ciência da Computação - Noturno: the ensemble's explanation remained concentrated, above all, on academic performance variables. At the same time, it showed that the explanatory logic is not restricted to that core. In different programs and windows, variables of admission, quotas, student assistance, scholarships, and student profile also emerged, some closer to retention and others more associated with dropout.

4.4.1 Recurring pattern across programs

Considering the 16 programs analyzed, a triangulation of results was performed to identify recurring or conflicting patterns across programs. The most stable pattern of the triangulation was in the approval variables. In $p2$, *periodo_2_disciplinas_aprovadas* ranked among the five most important variables in all 16 programs, while *periodo_1_disciplinas_aprovadas* appeared in 14 and *nota_ingresso* in 15. In $p3$, *periodo_3_disciplinas_aprovadas* was present in the 16 programs, *periodo_2_disciplinas_aprovadas* and *nota_ingresso* in 14, and *periodo_1_disciplinas_aprovadas* in 12. In $p4$, *periodo_4_disciplinas_aprovadas* appeared in 15 programs, *nota_ingresso* in 15, *periodo_3_disciplinas_aprovadas* in 12, *periodo_1_disciplinas_aprovadas* in 10, and *periodo_2_disciplinas_aprovadas* in 9.

The direction of these variables was the most uniform of the set. In the summary plots, higher approval values tended to shift the explanation toward retention, while lower approval values approached dropout. This behavior was repeated across the three windows, confirming that the main explanatory factor in the ensemble lies in the student's approval history.

The variables of academic instability appeared with less frequency, but with equally consistent direction. In $p2$, *periodo_2_disciplinas_reprovadas* and *periodo_2_disciplinas_ri* were among the five most important variables in 6 programs each. In $p3$, *periodo_3_disciplinas_reprovadas*, *periodo_2_disciplinas_ri*, and *periodo_3_disciplinas_ri* continued to appear in different programs. In $p4$, course withdrawals, failures, and failure due to insufficient attendance (RI) records remained present in part of the set. In the charts, higher values of these variables tended to shift the explanation toward dropout.

Thus, the triangulation confirms the same interpretative axis already observed in the case study: academic performance remains the main organizing principle of the explanation. What changes between the windows is the depth of this reading. In p2, the weight falls more on immediate signals. In $p3$ and $p4$, the explanation more clearly incorporates the accumulation of the trajectory.

4.4.2 Complementary factors

Although the explanatory core of the ensemble remains academic, some variables recur outside this pattern and help qualify the reading of the results.

The *nota_ingresso* was the main complementary factor. It appeared among the five most important variables in 15 programs in $p2$, 14 in $p3$, and 15 in $p4$. Its frequency was therefore high in all windows. The direction, however, was not as uniform as that of the approvals. In programs such as Farmácia and Direito - Integral, higher values of *nota_ingresso* tended to be associated with higher retention. In Psicologia, Pedagogia, and the two Ciências Biológicas offerings, especially in $p4$, higher values were more strongly associated with dropout. This indicates that the *nota_ingresso* is a transversal variable in the ensemble's explanation, but with a behavior that is more dependent on the context of each program.

Student assistance also appeared as complementary evidence in some programs. In $p2$, *periodo_2_ae* entered among the five most important variables in Farmácia and Pedagogia, while Psicologia highlighted *periodo_1_ae*. In Engenharia Civil, *periodo_2_ae* also appeared just below the main core. In $p4$, Pedagogia again presented *periodo_4_ae* as one of the most relevant factors. In the charts, these variables tended, in these programs, to shift the explanation toward retention. Thus, student assistance was not configured as a general pattern of the set, but emerged as a protective factor in specific contexts.

The variables of paid scholarship also deserve emphasis. In Ciência da Computação - Noturno, *periodo_3_bolsa_remunerada* appeared among the five most important variables in $p3$ and, in the chart, its presence approached retention. The same occurred in Ciências Biológicas - Licenciatura, with *periodo_3_bolsa_remunerada*, and in Ciências Biológicas - Bacharelado, with *periodo_4_bolsa_remunerada*. In Psicologia, however, *periodo_3_bolsa_remunerada* appeared on $p4$ closer to dropout. This shows that scholarships did not produce a single direction in the set, although in some of the programs they appeared to be associated with retention.

The variables of admission and quotas also emerged in a localized but relevant manner. The *tipo_ingresso_pism* appeared among the five most important variables in 5 programs in $p2$, 4 in $p3$, and 3 in $p4$. In Engenharia Civil, it remained relevant in the

three windows. In Pedagogia, Direito - Integral, and Ciências Econômicas - Noturno, when it appeared, it also tended more toward the retention side. The quotas, in turn, showed distinct behaviors. In Direito - Integral, *cota_renda* appeared in $p2$ closer to retention, while *cota_racial* appeared closer to dropout in $p2$ and again emerged in $p4$ with the same direction. In Ciências Econômicas - Noturno, *cota_renda* also appeared in $p2$ associated with retention. In Pedagogia, *cota_racial* emerged in $p4$ closer to dropout. These occurrences show that quotas entered the ensemble's explanation, but without producing a single pattern across programs.

The same occurred with gender and ethnicity. In Ciências Econômicas - Noturno, *genero_masculino* appeared among the central variables in $p2$, $p3$, and $p4$, always with direction closer to dropout. In the same program, *genero_feminino* appeared in $p2$ and $p3$ more associated with retention. In Direito - Integral, *etnia_branca* appeared in $p4$ with direction closer to retention, while *etnia_parda* appeared closer to dropout. In Pedagogia, *etnia_preta* emerged in $p3$ with displacement toward retention, while *etnia_parda* appeared in $p4$ closer to dropout. In Psicologia, *etnia_parda* also appeared on $p4$ with a direction toward dropout. In Ciências Biológicas - Licenciatura, *etnia_parda* again emerged in $p4$ with the same direction. This evidence shows that the profile variables did not dominate the triangulation, but also cannot be treated as marginal.

4.4.3 Synthesis by area and particularities

In the Exact Sciences and Computing programs – Ciência da Computação - Integral, Ciência da Computação - Noturno, Sistemas de Informação, Engenharia Civil, and Engenharia de Produção – the dominant pattern was the centrality of approvals, accompanied by signs of academic instability. In this area, RI, failures and course withdrawals occurred more frequently than in the others. In Ciência da Computação - Noturno, *periodo_3_disciplinas_ri* occupied the first position in $p3$ and remained relevant in $p4$. In Engenharia de Produção, *periodo_4_disciplinas_ri* and *periodo_2_disciplinas_trancadas* remained among the most important variables in $p4$. In Engenharia Civil, in addition to *periodo_2_disciplinas_ri*, *tipo_ingresso_pism* and *periodo_2_ae* appeared, all with direction closer to retention. In this area, therefore, the ensemble's explanation continued to be

strongly academic, but with greater sensitivity to signs of trajectory rupture.

In the Health and Biological Sciences programs – Ciências Biológicas - Bacharelado, Ciências Biológicas - Licenciatura, Enfermagem, Farmácia, and Medicina – the *nota_ingresso* appeared with high frequency and remained relevant up to more advanced windows. Even so, its direction was not unique. In Farmácia, it appeared closer to retention in $p2$, while in Ciências Biológicas - Bacharelado and Ciências Biológicas - Licenciatura, especially in $p4$, it shifted more toward dropout. In this same area, student assistance and paid scholarship also gained ground. In Farmácia, *periodo_2_ae* appeared in $p2$ with direction closer to retention. In Enfermagem, *periodo_2_ae* emerged in $p3$ with the same direction. In Ciências Biológicas - Licenciatura, *periodo_3_bolsa_remunerada* appeared associated with retention. In Ciências Biológicas - Bacharelado, the same occurred with *periodo_4_bolsa_remunerada*. Medicina was the main particularity of the area, because *periodo_1_disciplinas_aprovadas* remained the most important variable in $p2$, $p3$, and $p4$, indicating that, in this program, the student's initial history continued to weigh more than in the rest of the set.

In the Humanities and Applied Social Sciences programs – Direito - Integral, Direito - Noturno, Pedagogia, Psicologia, Ciências Econômicas - Integral, and Ciências Econômicas - Noturno – academic performance also remained central, but it was in this area that variables of profile, assistance, quotas, and admission appeared more clearly. Ciências Econômicas - Noturno concentrated the clearest gender pattern: *genero_masculino* appeared in the three windows closer to dropout, while *genero_feminino* appeared in $p2$ and $p3$ closer to retention. In Direito - Integral, *cota_renda* appeared in $p2$ associated with retention, while *cota_racial* and *etnia_parda* appeared closer to dropout, and *etnia_branca* approached retention in $p4$. In Pedagogia, *periodo_2_ae* and *periodo_4_ae* pointed to retention, *etnia_preta* appeared in $p3$ in the same direction, while *etnia_parda* and *cota_racial* emerged in $p4$, closer to dropout. In Psicologia, *periodo_1_ae* appeared in $p2$ associated with retention, but in $p4$, *periodo_3_bolsa_remunerada*, *etnia_parda*, *cota_renda*, and *periodo_4_disciplinas_trancadas* shifted toward dropout. This was, therefore, the most heterogeneous area of the triangulation.

To make the similarities and differences between the groups of programs more

visible, Table 4.5 synthesizes the main patterns observed in the triangulation by area, highlighting which factors appear most recurrently in the explanation of dropout and retention in each set of programs.

Table 4.5: Patterns of dropout explainability by area

Area	**Common pattern**	**Closer to dropout**	**Closer to retention / highlights**
Exact Sciences, Engineering, and Computing	Predominance of academic performance, especially accumulated approvals, with explanation centered on the academic trajectory.	RI, failures, and course withdrawals appeared more frequently, indicating greater sensitivity to signs of trajectory rupture.	Approvals remained as the main retention factor. In some programs, *tipo_ingresso_pism* and *periodo_2_ae* also appeared.
Health and Biological Sciences	Academic performance remained central, but the *nota_ingresso* appeared with high frequency and remained relevant up to more advanced windows.	In part of the programs, especially in the Biological Sciences, the *nota_ingresso* appeared closer to dropout, with less stable direction than that of approvals.	Approvals, student assistance, and paid scholarship emerged as factors closer to retention. In Medicina, the initial history maintained greater weight.
Humanities and Applied Social Sciences	Academic performance also remained relevant, but the area was more heterogeneous, with clearer entry of contextual factors.	Quotas, gender, ethnicity, scholarship, and course withdrawals appeared with greater prominence, depending on the program. In Ciências Econômicas - Noturno, the gender pattern was the most evident.	More varied protective signals also emerged, such as approvals, student assistance, and *cota_renda*, making this the area in which contextual factors most complemented the academic reading.

4.4.4 Most evident exceptions

Among the analyzed programs, four cases stood out more clearly.

In Ciências Econômicas - Noturno, the recurring presence of *genero_masculino* and *genero_feminino* among the most important variables was not episodic. It appeared as early as $p2$, remained in $p3$, and, in the case of *genero_masculino*, continued in $p4$. Furthermore, the charts showed distinct directions between these two variables: *genero_masculino* closer to dropout and *genero_feminino* closer to retention. This is the clearest gender exception in the analyzed set.

In Direito - Integral, profile and quota variables occupied a relevant space in the

explanation. In $p2$, *cota_renda* and *cota_racial* were already among the most important factors, with opposite directions. In $p4$, *etnia_branca* entered a prominent position, with direction toward retention, while *etnia_parda* remained closer to dropout. In this program, the ensemble preserved social and profile signals more visibly than in most of the set.

In Psicologia, the explanation also differed from the rest. In $p2$, *periodo_1_ae* appeared among the most important variables with direction toward retention. In $p4$, however, *periodo_3_bolsa_remunerada*, *etnia_parda*, *cota_renda*, and *periodo_4_disciplinas_trancadas* came to compose the explanation of dropout. This result shows that, in this program, the final explanation was less focused on performance and more clearly incorporated contextual factors.

In Medicina, the main particularity was the persistence of *periodo_1_disciplinas_aprovadas* as a central variable up to $p4$. While in most programs the explanatory weight was being shifted toward more recent periods, in Medicina the initial history remained at the top of the explanation. This indicates that, in this program, the performance at the beginning of the trajectory was more relevant to the ensemble's explanatory logic.

4.4.5 Synthesis of the triangulation

The results of the triangulation show four main points. The first is that the most stable explanatory axis of the ensemble, across the 16 programs analyzed, was academic performance variables, especially accumulated approvals. The second is that this axis becomes more accumulated from $p2$ to $p4$, more strongly incorporating the history of previous periods. The third is that the *nota_ingresso* was the main transversal variable outside the academic core, although with less stable direction across programs. The fourth is that student assistance, paid scholarship, admission type, quotas, gender, and ethnicity appeared as complementary components of the explanation, in some cases associated with retention and, in others, closer to dropout.

This triangulation broadens the findings from the case study of Ciência da Computação - Noturno. The central pattern identified in that program was maintained in the analyzed set, but the comparison between areas and offerings showed that the ensemble's explainability also preserves contextual signals of admission, institutional support, and

student profile. With this, the evaluation of the second cycle is not limited to demonstrating that the ensemble produces interpretable explanations in a single case. It also shows that these explanations maintain coherence across a broader set of programs, strengthening the fulfillment of requirement FR05 and providing a more comparable basis for institutional reading within the scope of FR06.

4.5 Evaluation of the EducAAr Analysis Support Panel

In addition to prediction and explainability, the second cycle of EducAAr also includes the presentation of results in an analysis-support panel. The panel evaluation seeks to verify whether the artifacts produced by the architecture were organized in an accessible manner for consultation, interpretation, and simulation.

This evaluation does not correspond to a user study with managers or academic teams. Its focus is technical and architectural: to verify whether the results generated by the predictive and explanatory layers were integrated into a coherent interface, capable of presenting metrics, explanations, simulations, prompts, and technical details in a unified environment.

The panel was built from the results exported at the end of predictive modeling, including aggregated metrics, records by repetition, evaluation figures, SHAP values, global rankings of importance, and the trained and calibrated ensemble. With this, the interface came to gather, in a single environment, the main results of the second cycle.

Although the panel operates on the results of the 16 analyzed course offerings, the detailed presentation of this stage uses the Ciência da Computação – Noturno program as a reference, maintaining the same line adopted throughout this chapter. This choice allows evaluating the panel's operation without repeating, for all programs, the analyses already presented.

4.5.1 Panel organization

The panel was organized into four main tabs: Summary, Explainability, Simulation, and Technical details. This division was adopted to separate the more synthetic reading of

the results, the global explanatory analysis, the exploration of individual scenarios, and the complementary technical elements of the evaluation.

In the Summary tab, the panel presents a synthesis of the selected period. It displays central metrics of the ensemble, such as $F1_{pos}$, ROC-AUC, calibrated Brier Score, and number of analyzed students. The tab also includes a chart of the most important variables in the period, lists of the main signals associated with dropout risk and retention, and a summarized visualization of the global importance ranking.

Figure 4.12: Summary tab of the EducAAr panel for the Ciências da Computação – Noturno program in period $p3$ (in portuguese).

It also offers an optional LLM-assisted natural-language explanation. This expla-

nation is generated only when the user activates the corresponding button in the interface and is intended to complement the panel reading by reorganizing the main patterns identified in the analyzed period into text.

In the submission to the language model, the system does not send the complete dataset or individual student data. The prompt is assembled from a global summary of the period. Initially, the panel calculates the importance of all variables based on the mean absolute SHAP value. Then, it sums these importance values to obtain a global total and selects the 10 most relevant variables for the period. For each variable, the global contribution percentage is calculated by dividing its importance by the sum of all variables' importances and multiplying the result by 100.

These 10 variables compose the "Main variables by weight" block of the prompt. From this same selected set, the system also generates the "Signals associated with risk" and "Signals associated with retention" blocks, based on the mean direction of SHAP contributions. When the admission grade appears among the highlighted variables, the prompt also includes a specific note of caution regarding its interpretation.

Thus, the textual explanation is derived from a summarized set of global factors for the period: program, analyzed period, model used, interpretation instruction, 10 most important variables, risk signals, retention signals, and observations on the admission grade. The interface also allows the prompt to be visualized, making this resource more transparent to the user.

Figure 4.12 presents the Summary tab of the panel for the Ciências da Computação – Noturno program in period $p3$. The Box below illustrates the content sent to the language model in this functionality.

Box – Example of content sent to the language model for Ciências da Computação – Noturno in period $p3$

You are helping to interpret an academic dropout prediction panel.

Program: Ciências da Computação - Noturno

Analyzed period: p3

Model: Calibrated tree ensemble (XGBoost + LightGBM + CatBoost)

Explain in objective language, aimed at supporting academic analysis: 1. what are the main signals of dropout risk; 2. which factors are associated with retention; 3. how to interpret these findings with caution, without treating them as causality; 4. what type of institutional analysis could be prioritized based on these signals.

Main variables by weight: - failures due to insufficient attendance in the 3rd period: 22.9% of the global weight - approved course subjects in the 1st period: 16.4% of the global weight - approved course subjects in the 3rd period: 14.3% of the global weight - approved course subjects in the 2nd period: 12.8% of the global weight - paid scholarship in the 3rd period: 10.3% of the global weight - admission grade: 9.8% of the global weight - open competition: 2.1% of the global weight - withdrawn course subjects in the 2nd period: 1.8% of the global weight - failed course subjects in the 2nd period: 1.3% of the global weight - racial quota: 0.9% of the global weight

Signals associated with risk: - High values of failures due to insufficient attendance in the 3rd period - Low values of approved course subjects in the 1st period - Low values of approved course subjects in the 3rd period - Low values of approved course subjects in the 2nd period - Absence of paid scholarship in the 3rd period - High values of admission grade - Absence of open competition - High values of withdrawn course subjects in the 2nd period - Presence of racial quota

Signals associated with retention: - Low values of failures due to insufficient attendance in the 3rd period - High values of approved course subjects in the 1st period - High values of approved course subjects in the 3rd period - High values of approved course subjects in the 2nd period - Presence of paid scholarship in the 3rd period - Low values of admission grade

Important note: - The admission grade may present a non-homogeneous distribution in the summary plot. In such cases, its reading should be done with caution and together with the other variables.

Use a clear academic tone, without exaggeration.

In the Explainability tab, the panel presents the global SHAP chart of the selected period. This visualization allows observing which variables most contributed to shifting the ensemble's prediction toward dropout or retention, offering a more detailed view of the ensemble's global behavior.

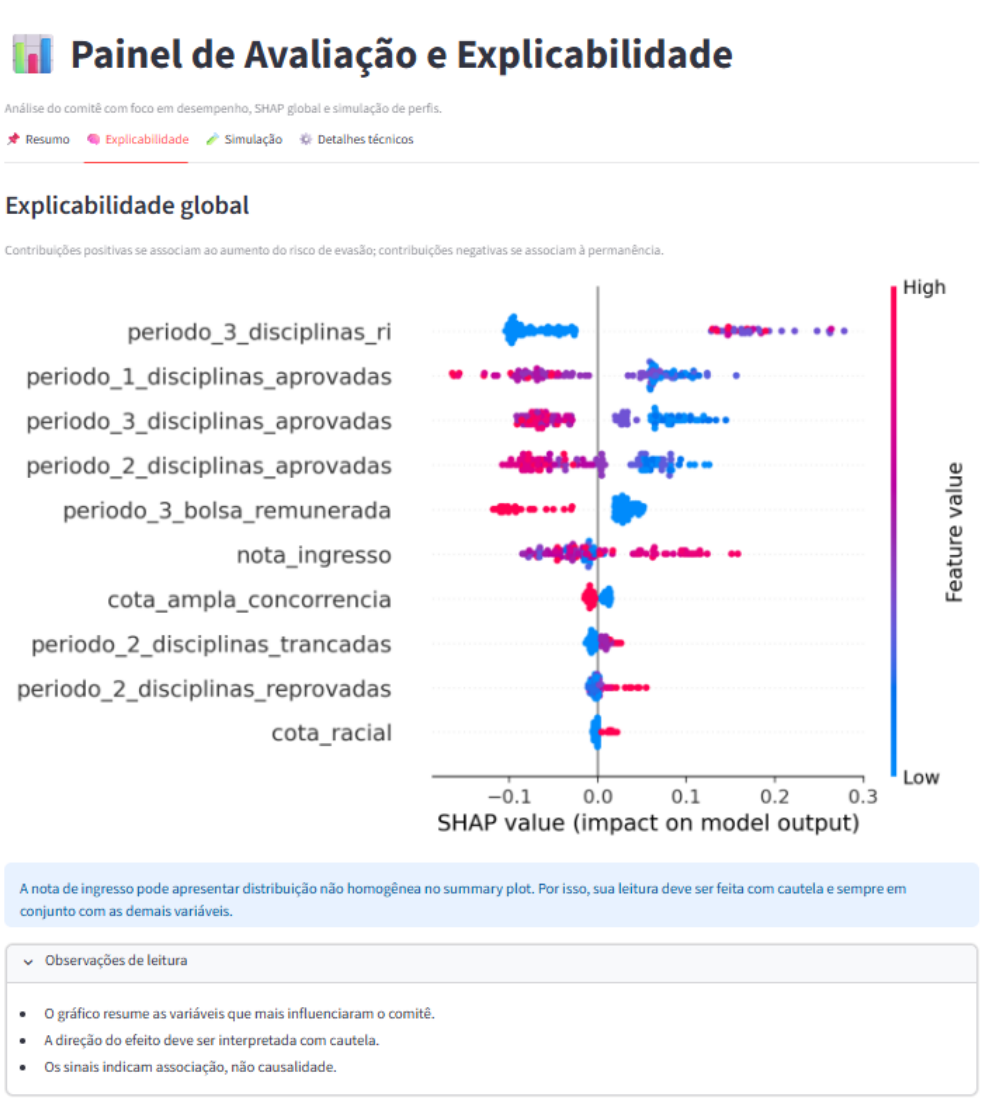

Figure 4.13: Explainability tab of the EducAAr panel, with the global SHAP chart for Ciências da Computação – Noturno in period $p3$ (in portuguese).

In the Simulation tab, the panel allows entering values for variables of admission, gender, quotas, ethnicity, and academic performance by period. Based on these inputs, the system calculates the estimated dropout probability, the predicted class, and the local explanation of the informed profile.

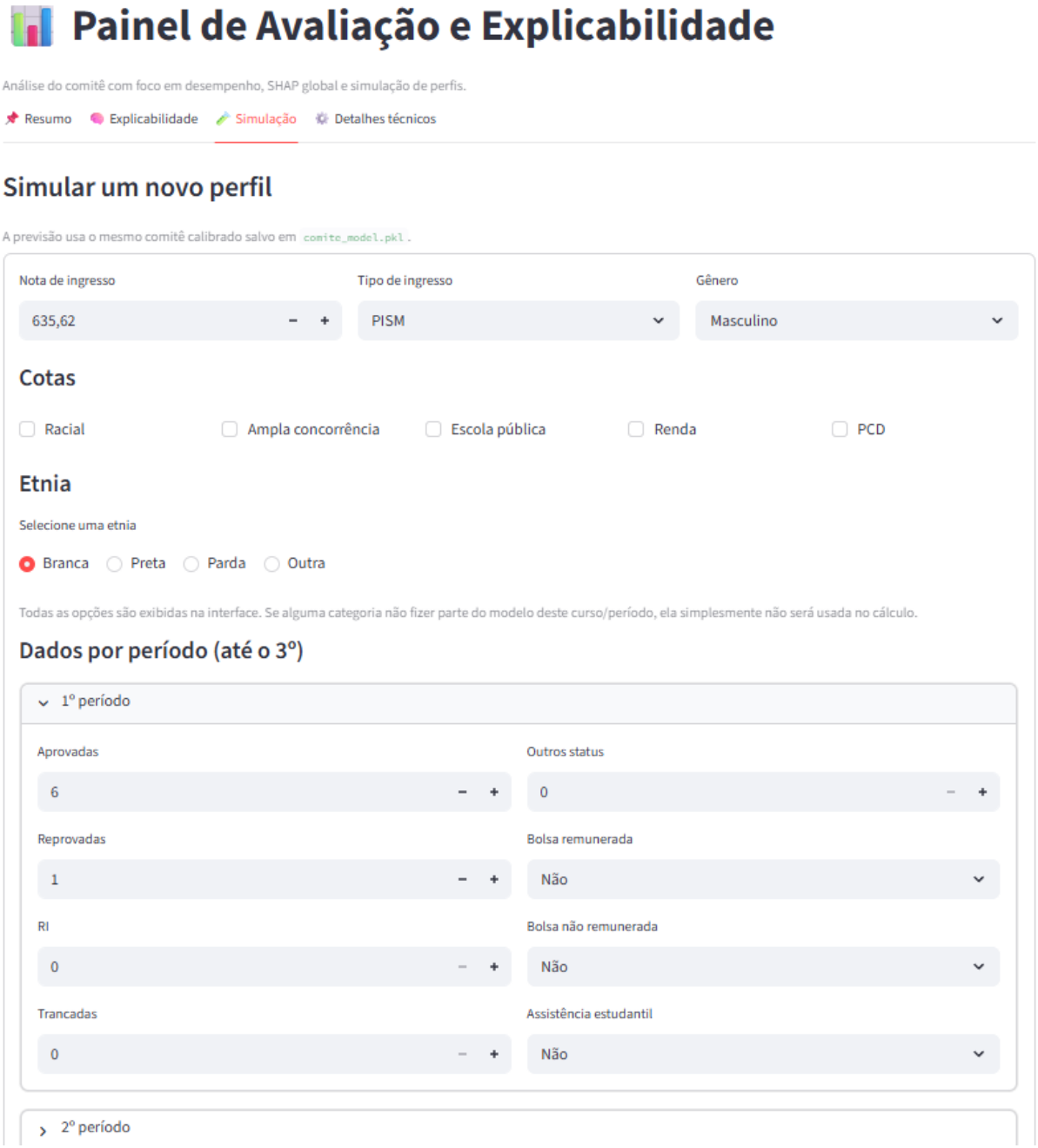

Figure 4.14: Form of the Simulation tab of the EducAAr panel for Ciências da Computação – Noturno in period $p3$ (in portuguese).

The interface also presents a SHAP waterfall chart and a list with the factors that most increased or reduced the estimated risk. With this, the panel is not limited to a global reading of the results but allows exploration of hypothetical profiles individually.

Figure 4.15: Result of the simulation of a student profile, with estimated risk, predicted class, and local SHAP explanation (in portuguese).

Finally, the Technical details tab gathers the complementary elements of the evaluation. It presents the ensemble's aggregated performance, the best-execution data, figures associated with this execution, and the variation in the metrics across repetitions. This organization keeps this information accessible in the panel, but without overloading the central reading and simulation tabs.

Figure 4.16: Technical details tab of the EducAAr panel for Ciências da Computação – Noturno in period $p3$ (in portuguese).

4.5.2 Evaluation of use in the Ciência da Computação – Noturno program

In the Ciência da Computação – Noturno program, the panel allowed the gathering of the main results already discussed in the previous sections in a single interface. By selecting windows $p2$, $p3$, and $p4$, it was possible to directly access aggregated metrics, the global SHAP chart, simulation features, and complementary technical information,

without needing to consult separate files.

This point is relevant because, outside the panel, the results of the second cycle remain distributed in tables, images, and artifacts exported in different formats. The interface reduces this dispersion and organizes the reading into a single flow, facilitating the transition between predictive performance, global explainability, and individual analysis.

In practice, this allowed observing, in the same environment, that the ensemble maintained good performance on separation and calibration metrics while, at the same time, supporting its estimates mainly on variables linked to academic performance. It was also possible to simulate student profiles and verify how changes in approvals, failures, failure due to insufficient attendance, and scholarship participation altered the estimated risk and its explanation.

Thus, the evaluation in the Ciência da Computação – Noturno program shows that the panel manages to transform the analytical results of the second cycle into a more direct visualization, articulating global reading, assisted synthesis, and exploration of individual cases.

4.5.3 Panel scope

Although the detailed presentation of this evaluation was made based on Ciência da Computação – Noturno, the panel was structured to reuse the same organizational standard in the other analyzed programs. Switching program and period in the interface triggers the loading of the corresponding artifacts, preserving the same visualization logic.

This means that the panel was not built for a single case study, but as a reusable component of the architecture. Its function is to serve as an access layer to the results of the second cycle, regardless of the selected program, provided that the necessary artifacts have been previously generated.

4.5.4 Synthesis of the panel evaluation

The panel evaluation shows that the second cycle of EducAAr did not limit itself to generating predictions and explanations but also included a structured presentation of these results. The panel gathered, in the same environment, a synthetic reading of the

period, global explainability, profile simulation, and technical consultation, reducing the dispersion of the artifacts produced by the architecture.

With this, the panel meets the FR06 requirement by offering an interface that organizes results in a more accessible way for dropout risk analysis. Its role is not to replace the specialist's interpretation, nor was it validated as an institutional decision-making tool. Rather, it makes the reading of the results produced by EducAAr more direct by integrating predictive metrics, explainability, simulation, prompt transparency, and technical details in the same environment.

4.6 Threats to Validity

The results of this dissertation should be interpreted considering some threats to validity and ethical aspects related to the use of institutional educational data. In this section, these issues are discussed across internal validity, external validity, construct validity, ethical considerations, and reliability. This organization makes explicit the main limits related to the research design, the data sample, the representation of the investigated phenomenon, the use of sensitive variables, and the possibility of reproducing the adopted procedures.

4.6.1 Internal Validity

Internal validity concerns the factors that may influence the results obtained during the development and evaluation of the architecture. In this dissertation, the first threat is inherent to the predictive analysis's design. Windows $p2$, $p3$, and $p4$ were adopted to focus the observation on the initial periods of the program, with the objective of bringing the analysis closer to an early dropout prediction scenario. However, this choice results in the results representing a specific cut of the academic trajectory rather than the full complexity of the student journey.

In addition, students without complete records for the corresponding window were excluded from the dataset for that stage. This decision was necessary to maintain consistency in the variables used in each cut, but it may affect the composition of the

analyzed datasets. Thus, the results obtained in each window should be understood as associated with the set of students who had sufficient information at that moment of the trajectory.

Another internal threat is the splitting of the data into training and test sets across repetitions. Although a protocol with multiple repetitions and stratified partitions was adopted to reduce dependence on a single data split, this procedure does not completely eliminate possible variations arising from sample composition. This point is especially relevant in programs with fewer students, where small changes in data splitting can lead to greater oscillation in the metrics.

The risk of model overfitting must also be considered. Although appropriate models for tabular data and repeated evaluation procedures were used, the classifiers may still capture patterns specific to the dataset used, which do not necessarily repeat in other datasets. For this reason, the results should be interpreted together with the evaluation metrics, the explanatory analyses, and the limitations of the adopted cut.

4.6.2 External Validity

External validity concerns the extent to which results can be generalized to other contexts. The evaluation of the EducAAr architecture was conducted using anonymized institutional data from the Federal University of Juiz de Fora, encompassing 16 undergraduate course offerings. Although this cut allowed evaluating the architecture in a concrete institutional context, the results should not be automatically generalized to other institutions, programs, or historical periods.

Another threat lies in the cut adopted for admission data. In the predictive stage, only SISU and PISM were considered, as they are the predominant methods during the analyzed period, exhibit greater regularity in institutional records, and allow the use of grades on the same scale. This choice was adequate to preserve comparability among the analyzed cases, but it restricts the scope of the study, as other types of admission, such as transfers, re-admissions, vacant slots, and admissions for graduates, were excluded from the modeling.

The differences between the analyzed programs must also be considered. The 16

offerings have distinct sizes, profiles, and academic dynamics. Thus, the patterns observed in certain programs may not repeat with the same intensity in others. The architecture was proposed to be reusable and adaptable, but its application in new contexts requires new data instantiations, reevaluation of the models, and analysis of the generated explanations.

Finally, this dissertation did not longitudinally follow the application of institutional interventions based on the generated predictions. Thus, the work shows that EducAAr can support dropout risk analysis, but does not, at this stage, measure the practical effect of this support on concrete student retention actions.

4.6.3 Construct Validity

Construct validity concerns how the investigated phenomenon was represented in the study. In this dissertation, an important threat concerns the quality of the database used. Although procedures for organization, integration, and treatment of information were adopted through the ontology and the generation of the tabular datasets, the results still depend on the quality of the available institutional records. Inconsistencies, absences, filling errors, or differences in records across systems may reverberate throughout the construction of the variables and the estimates produced by the models.

Another issue concerns the representation of dropout as a binary target variable. The variable *status* was defined based on dropout and graduate students, while active students were not included in the supervised training because they do not yet have a final outcome. This decision avoids training the models with cases without known completion, but it also simplifies the academic trajectory by reducing it to two main outcomes. Therefore, intermediate situations or more complex trajectories may not be completely represented in the modeling.

There are also threats associated with the transformation of the data integrated by the ontology into tabular attributes. The ontology allowed for organizing relationships among students, admissions, academic history, scholarships, and other institutional links. However, the vectorization stage required transforming these relationships into variables used by the machine learning models. This transformation may reduce some of

the semantic richness of the ontological model, since more complex relationships must be represented using numerical or binary attributes.

In addition, some variables used in the analysis, such as scholarship participation, student assistance, failures, course withdrawals, and performance by period, depend directly on the way this information was recorded in the institutional systems. Thus, the absence of a given record does not necessarily mean the absence of the phenomenon in the student's trajectory, but may reflect limitations of the available database.

4.6.4 Ethical Considerations in the Use of Sensitive Variables

Another important aspect concerns the use of variables related to student profile and institutional support, such as gender, ethnicity, quotas, and student assistance. These variables were included because they are part of the institutional records and because they can help reveal contextual patterns associated with dropout risk. However, their use requires caution, since these attributes may be related to social inequalities, institutional conditions, and historical differences in access and permanence.

In this dissertation, these variables are not interpreted as causal explanations of dropout, nor as attributes that should justify individual judgments about students. When gender, ethnicity, quotas, or student assistance appear in the explanations, the result indicates only that the model used these variables as part of its predictive structure in a specific dataset and context. Therefore, such findings must be interpreted together with academic, institutional, and social information, avoiding any stigmatizing or deterministic reading.

The inclusion of these variables also has an analytical role. By making their influence visible through explainability techniques, the architecture allows these effects to be inspected rather than hidden inside the model. This visibility is important because it helps identify whether the model is relying on sensitive or contextual variables and supports a more critical interpretation of the results.

For this reason, the use of EducAAr should always preserve human interpretation, institutional responsibility, and ethical caution. The architecture should not be used to classify students in a punitive way or to restrict opportunities. Its purpose is to organize

evidence for dropout risk analysis, while preserving the need for contextual reading and careful institutional judgment.

4.6.5 Reliability

Reliability refers to the ability to reproduce the adopted procedures and obtain consistent results from the same methodological design. In this dissertation, this threat was sought to be reduced by explicitly defining the stages of extraction, integration, ontology instantiation, generation of tabular datasets, model training, and explanation production.

In the second cycle, the use of multiple repetitions with different seeds aimed to reduce dependence on a single data split. This procedure helps assess the stability of the models and metrics across different partitions. Even so, in programs with smaller datasets, the number of available students may make the splits between training and test less stable, hindering the calibration of probabilities and the definition of more robust thresholds.

Another threat to reliability concerns the complete replication of the process in other institutional contexts. Although the architecture has been built to integrate educational data through a canonical model, its application at another institution would require new mappings of the source databases, assessment of the adequacy of the available variables, and validation of the consistency of the loaded data. Therefore, reproducing the approach depends not only on the code and models, but also on the availability and quality of institutional data.

Finally, the explanations produced by explainability techniques should also be interpreted with caution. The explanatory values indicate the contribution of variables to the model's estimates, but should not be treated as causal proof of dropout. Thus, EducAAr should be understood as an architecture for supporting institutional analysis, capable of indicating patterns and factors associated with dropout risk, but not as an automatic decision-making mechanism.

5 Conclusion

This dissertation began with the problem of student dropout in higher education and the difficulty of analyzing this phenomenon using data that are, in practice, distributed across different systems, files, and formats. In addition, even when it is possible to predict dropout risk, prediction in isolation does not answer an important question for institutional use: why does a given student, program, or profile appear associated with higher risk?

In this context, the EducAAr (*Educational Analysis Architecture*) was proposed to integrate educational data, apply machine learning models, generate explanations for the resulting estimates, and organize these results in a visualization panel. The research was conducted in two cycles of Design Science Research. The first DSR cycle focused on constructing an ontology as a canonical model for data integration. The second DSR cycle expanded on this foundation with predictive models, explainability, and a panel to support institutional analysis.

The research question that guided the work was:

> *How can an architecture based on a canonical model for educational data integration combine machine learning, explainability, and visualization mechanisms to produce interpretable analyses of student dropout risk in higher education?*

The approach developed throughout the dissertation shows that this integration can be achieved through a layered architecture, in which the ontology organizes the data and their relationships, machine learning models produce risk estimates, explainability techniques help interpret these estimates, and the panel presents the results in a more accessible manner for analysis.

In relation to the specific objectives defined in the Introduction, they are considered to have been fulfilled in the following way:

- Build an ontology as a canonical model to integrate educational data from different

institutional sources. This objective was met in the first DSR cycle. The ontology enabled the representation of students, academic history, admission methods, performance, student assistance, scholarships, and academic situation within a single structure. The validation with the HermiT *reasoner* showed that the modeling was logically consistent, and the SPARQL queries enabled retrieving information previously dispersed across the source databases. Thus, the ontology fulfilled the role of canonical model for organizing and integrating the educational data used in the architecture.

- Transform the integrated educational data into structured datasets suitable for predictive modeling. This objective was fulfilled in the second DSR cycle through the vectorization process. The information represented in the ontology was transformed into tabular attributes, preserving fixed variables related to admission, student profile, quotas, scholarships, and student assistance, as well as temporal variables related to academic performance in the analyzed periods. This process enabled the construction of the datasets used in the windows $p2$, $p3$, and $p4$.

- Incorporate and technically evaluate machine learning models to estimate student dropout risk. This objective was fulfilled in the second cycle. The XGBoost, LightGBM, and CatBoost models were used, combined in a calibrated ensemble. The models were applied to initial windows of the academic trajectory, represented by $p2$, $p3$, and $p4$. The results showed that the architecture produced dropout risk estimates from the integrated data, with consistent performance across the evaluated metrics, especially ROC-AUC, *Average Precision*, and calibrated Brier Score.

- Apply explainability techniques to identify the factors most associated with the risk estimates produced by the models. This objective was met with the use of SHAP. The explanations made it possible to observe which variables shifted the estimates toward dropout or retention. In the Ciência da Computação – Noturno program, analyzed in greater detail, the main explanatory factors were linked to academic performance, especially approvals, failures, and failures due to insufficient attendance. The triangulation across the 16 programs confirmed this pattern, but

also showed that variables such as admission grade, student assistance, scholarships, quotas, gender, and ethnicity appear in certain contexts.

- Organize predictive metrics, explanatory results, and simulated student profiles in a visualization panel. This objective was fulfilled with the development of the EducAAr panel. The panel gathered metrics, explainability charts, profile simulations, prompt transparency, and technical details of the executions. With this, the results no longer remained scattered across files and were organized in a single interface, allowing a more direct reading of model performance, explanatory factors, and simulated cases.

- Analyze how the integration between ontology, predictive modeling, explainability, and visualization contributes to a structured architectural workflow for dropout risk analysis. This objective was fulfilled through the evaluation of the two DSR cycles. The first cycle demonstrated the role of the ontology in organizing and integrating institutional data, while the second cycle showed how this integrated basis could be connected to prediction, explainability, and visualization. Together, the cycles showed that EducAAr operates as an integrated workflow rather than as an isolated predictive model.

The evaluation of the architecture showed that EducAAr is not limited to a predictive model. The first cycle demonstrated that the ontology could integrate educational data and support queries over information previously dispersed across different sources. The second cycle showed that this integrated foundation could be used to generate risk estimates, explain the factors associated with these estimates, and organize the results in a panel for consultation and interpretation.

The explanatory analysis showed that academic performance was the main interpretation axis. In $p2$, the greatest weight was on initial approvals and failures. In $p3$, failure due to insufficient attendance gained prominence. In $p4$, the performance of the current period appeared combined with accumulated signals from previous periods. This reading also appeared in the triangulation across the 16 programs, indicating that the academic history of the first periods plays a central role in the explanation of dropout

risk.

At the same time, the triangulation showed that dropout should not be interpreted based on a single variable or type of data. In some programs, factors linked to admission grade, student assistance, scholarships, quotas, gender, and ethnicity emerged. These factors did not appear with the same strength in all programs, but they help to show that the institutional analysis also needs to consider the context of each area and of each offering.

These findings also reinforce the need for ethical caution in the interpretation of the results. Variables such as gender, ethnicity, quotas, and student assistance should not be read as causal explanations or as criteria for individual judgment, but as contextual signals that require institutional and social interpretation.

The main contributions of this dissertation can be synthesized as follows:

- proposal of the EducAAr architecture, bringing together data integration, machine learning, explainability, and visualization in the same structure;
- construction of an ontology as a canonical model to integrate educational data from different institutional sources;
- expansion of the architecture across two DSR cycles, starting from data organization to a structure with prediction, explanation, and result visualization;
- application of a tree-based ensemble of models to estimate dropout risk in early windows of the program;
- use of SHAP to interpret the factors associated with dropout and retention estimates;
- analysis of the explanatory patterns across 16 course offerings, allowing the observation of both recurring factors and particularities among areas;
- development of a panel to organize metrics, explanations, and simulations in an interface aimed at supporting the interpretation of dropout risk analyses.

Despite these contributions, EducAAr should not be understood as an automatic solution to the dropout problem. The architecture organizes data, estimates risks, and

provides explanations, but the interpretation of the results still depends on the institutional context and on analyses by managers, coordination teams, and academic staff. The model's explanations indicate associations relevant to prediction but should not be treated as causal evidence of dropout.

As future work, the following stand out:

- incorporate new sources of information, such as data on pedagogical follow-up, participation in academic activities, and records from virtual environments;
- evaluate the panel with managers, program coordinations, and teams responsible for student follow-up;
- study forms of continuous updating of the models over time;
- deepen the analysis of the contextual factors observed in the triangulation, especially student assistance, quotas, gender, and ethnicity;
- conduct longitudinal studies, after institutional validation of the panel, to analyze how the architecture may support retention strategies and how these strategies relate to dropout and retention outcomes.

Thus, this dissertation demonstrated that integrating ontology, machine learning, explainability, and visualization can support a more structured analysis of dropout in higher education. EducAAr contributes by organizing the data, producing risk estimates, and making the factors associated with these estimates more visible, offering a foundation for more consistent institutional analyses.

References

ABOUELNOUR, S. et al. Machine learning in higher education: Predicting and mitigating student dropout. In: . Piscataway, NJ, USA: IEEE, 2024.

AGUAYO-MAURI, S. et al. Human vs machine learning: Best approach to early detect university dropout rates. In: *Lecture Notes in Educational Technology*. Cham, Switzerland: Springer, 2025. Part F642, p. 1129–1138.

AJOODHA, R.; JADHAV, A.; DUKHAN, S. Forecasting learner attrition for student success at a south african university. In: *ACM International Conference Proceeding Series*. New York, NY, USA: Association for Computing Machinery, 2020. p. 19–28.

AKTER, T. et al. Dropout prediction of university students in bangladesh using machine learning. In: . Piscataway, NJ, USA: IEEE, 2024.

ALCAUTER, I.; MARTÍNEZ-VILLASEÑOR, L.; PONCE, H. E. Explaining factors of student attrition at higher education. *Computacion y Sistemas*, v. 27, n. 4, p. 929–940, 2023.

ALI, S. et al. The enlightening role of explainable artificial intelligence in medical & healthcare domains: A systematic literature review. *Computers in Biology and Medicine*, v. 166, p. 107555, 2023.

AQIB, A. I. et al. Explainable artificial intelligence (xai) in the veterinary and animal sciences field. In: . [s.n.], 2023. p. 33–56. Disponível em: ⟨https://www.scopus.com/inward/record.uri?eid=2-s2.0-85164037792&partnerID=40&md5=a00aa86c43588ebf1c406747471a525b⟩.

AZY, W. et al. 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)*. Porto Alegre, RS, Brasil: SBC, 2024. p. 3038–3048. Disponível em: ⟨https://sol.sbc.org.br/index.php/sbie/article/view/31466⟩.

BARANYI, M.; NAGY, M.; MOLONTAY, R. Interpretable deep learning for university dropout prediction. In: . New York, NY, USA: Association for Computing Machinery, 2020. p. 13–19.

BARANYI, R.; MOLONTAY, R. Interpretable deep learning for university dropout prediction. *Computers & Education: Artificial Intelligence*, v. 1, p. 100008, 2020. Disponível em: ⟨https://www.sciencedirect.com/science/article/pii/S2666920X20300083⟩.

BARRAMUÑO, M.; MEZA-NARVÁEZ, C.; GÁLVEZ-GARCÍA, G. Prediction of student attrition risk using machine learning. *Journal of Applied Research in Higher Education*, v. 14, n. 3, p. 974–986, 2022.

BELLO, F. A. et al. Using machine learning methods to identify significant variables for the prediction of first-year informatics engineering students dropout. In: *Proceedings–International Conference of the Chilean Computer Science Society, SCCC*. Piscataway, NJ, USA: IEEE, 2020. v. 2020-November.

BERENS, J. et al. Crossing individual university boundaries: a comprehensive approach to predicting dropouts in the higher education system. *Higher Education*, 2025.

BERGSTRA, J.; BENGIO, Y. Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, v. 13, n. 10, p. 281–305, 2012. Disponível em: ⟨https://jmlr.org/papers/v13/bergstra12a.html⟩.

BETTAHI, A.; BELOUADHA, F. Z.; HARROUD, H. A modular and explainable machine learning pipeline for student dropout prediction in higher education. *Algorithms*, v. 18, n. 10, p. 662, 2025.

BRANDENBURG, J. M. et al. Can surgeons trust ai? perspectives on machine learning in surgery and the importance of explainable artificial intelligence (xai). *Langenbeck's Archives of Surgery*, v. 410, n. 1, 2025.

BRASIL. *Lei nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD)*. 2018. Disponível em: ⟨https://www.planalto.gov.br/ccivil_03/_ato2015-2018/2018/lei/l13709.htm⟩.

BRIER, G. W. Verification of forecasts expressed in terms of probability. *Monthly Weather Review*, v. 78, n. 1, p. 1–3, 1950. Disponível em: ⟨https://journals.ametsoc.org/view/journals/mwre/78/1/1520-0493_1950_078_0001_vofeit_2_0_co_2.xml⟩.

CARDONA, T. A. et al. Data mining and machine learning retention models in higher education. *Journal of College Student Retention: Research, Theory and Practice*, v. 25, n. 1, p. 51–75, 2023.

CASTRO, R. Q.; GARCIA, K. C.; PELAEZ, E. Predictive modeling and explainability for academic dropout risk detection using machine learning. In: *International Conference on eDemocracy and eGovernment, ICEDEG*. Piscataway, NJ, USA: IEEE, 2025. p. 114–122.

CHEN, T.; GUESTRIN, C. Xgboost: A scalable tree boosting system. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2016. p. 785–794. Disponível em: ⟨https://dl.acm.org/doi/10.1145/2939672.2939785⟩.

Connected Papers. *Connected Papers: Find and explore academic papers*. 2026. Disponível em: ⟨https://www.connectedpapers.com/⟩.

CROOK, B.; SCHLUTER, M.; SPEITH, T. Revisiting the performance-explainability trade-off in explainable artificial intelligence (xai). In: . Piscataway, NJ, USA: IEEE, 2023. p. 316–324.

DUMKA, A. et al. Methods, techniques, and application of explainable artificial intelligence. In: . Hershey, PA, USA: IGI Global, 2024. p. 337–354.

FAWCETT, T. An introduction to roc analysis. *Pattern Recognition Letters*, v. 27, n. 8, p. 861–874, 2006.

FERNÁNDEZ-LÓPEZ, M.; GÓMEZ-PÉREZ, A.; JURISTO, N. Methontology: From ontological art towards ontological engineering. In: *Proceedings of the AAAI Spring Symposium Series on Ontological Engineering*. Stanford, USA: AAAI Press, 1997.

GLANDORF, D. et al. Temporal and between-group variability in college dropout prediction. In: . New York, NY, USA: Association for Computing Machinery, 2024. p. 486–497.

GLIMM, B. et al. Hermit: An owl 2 reasoner. *Journal of Automated Reasoning*, v. 53, n. 3, p. 245–269, 2014.

GUNASEKARA, S.; SAARELA, M. Explainability in educational data mining and learning analytics: An umbrella review. In: PAASSEN, B.; EPP, C. D. (Ed.). *Proceedings of the 17th International Conference on Educational Data Mining*. Atlanta, Georgia, USA: International Educational Data Mining Society, 2024. p. 887–892. ISBN 978-1-7336736-5-5. Disponível em: ⟨https://educationaldatamining.org/edm2024/proceedings/2024.EDM-posters.104/index.html⟩. Acesso em: 2 jun. 2026.

GUTIERREZ-PACHAS, D. A. et al. Supporting decision-making process on higher education dropout by analyzing academic, socioeconomic, and equity factors through machine learning and survival analysis methods in the latin american context. *Education Sciences*, v. 13, n. 2, p. 154, 2023.

HAMIDA, S. U. et al. Exploring the landscape of explainable artificial intelligence (xai): A systematic review of techniques and applications. *Big Data and Cognitive Computing*, v. 8, n. 11, p. 149, 2024.

HEVNER, A. A three cycle view of design science research. *Scandinavian Journal of Information Systems*, v. 19, n. 2, p. 87–92, 01 2007. Disponível em: ⟨https://aisel.aisnet.org/sjis/vol19/iss2/4/⟩.

Instituto Semesp. *Mapa do Ensino Superior no Brasil 2025*. 15. ed. São Paulo, SP, Brasil: Instituto Semesp, 2025. Disponível em: ⟨https://www.semesp.org.br/mapa/edicao-15/⟩.

JAYARAMAN, J. D.; GERBER, S.; GARCIA, J. Supporting minority student success by using machine learning to identify at-risk students. In: . [s.n.], 2019. p. 584–587. Disponível em: ⟨https://www.scopus.com/inward/record.uri?eid=2-s2.0-85085979172&partnerID=40&md5=cb1f065847f3fbd611b1a94e0f18850b⟩.

JIMÉNEZ, O.; JESÚS, A.; WONG, L. R. Model for the prediction of dropout in higher education in peru applying machine learning algorithms: Random forest, decision tree, neural network and support vector machine. In: *Conference of Open Innovation Association, FRUCT*. Helsinki, Finland: FRUCT Oy, 2023. v. 2023-May, p. 116–124.

JINAD, R.; ISLAM, A. B.; SHASHIDHAR, N. K. Interpretability and transparency of machine learning in file fragment analysis with explainable artificial intelligence. *Electronics (Switzerland)*, v. 13, n. 13, p. 2438, 2024.

JOSHI, P. P.; BAGADE, A. M. Explainable artificial intelligence techniques for image classification models in diverse domains. In: . Piscataway, NJ, USA: IEEE, 2023.

KE, G. et al. Lightgbm: A highly efficient gradient boosting decision tree. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Red Hook, NY, USA: Curran Associates, Inc., 2017. Disponível em: ⟨https://proceedings.neurips.cc/paper/2017/hash/6449f44a102fde848669bdd9eb6b76fa-Abstract.html⟩.

KITCHENHAM, B.; BRERETON, P. A systematic review of systematic review process research in software engineering. *Information and Software Technology*, v. 55, n. 12, p. 2049–2075, 2013. Disponível em: ⟨https://doi.org/10.1016/j.infsof.2013.07.010⟩.

KOSTOPOULOS, G. et al. Early dropout prediction in distance higher education using active learning. In: . Piscataway, NJ, USA: IEEE, 2017. v. 2018-January, p. 1–6.

LAMY, J.-B. *Owlready2 Documentation*. 2024. Disponível em: ⟨https://owlready2.readthedocs.io/en/v0.50/⟩.

LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems (NIPS)*, v. 30, p. 4768–4777, 2017. Disponível em: ⟨https://proceedings.neurips.cc/paper/7062-a-unified-approach-to-interpreting-model-predictions.pdf⟩.

MAKSIMOVA, N.; PENTEL, A.; DUNAJEVA, O. Predicting first-year computer science students drop-out with machine learning methods: A case study. In: *Advances in Intelligent Systems and Computing*. Cham, Switzerland: Springer, 2021. v. 1329, p. 719–726.

MANRIQUE, R. et al. An analysis of student representation, representative features and classification algorithms to predict degree dropout. In: *Proceedings of the 9th International Conference on Learning Analytics and Knowledge*. New York, NY, USA: ACM, 2019. (LAK19), p. 401–410.

MARTINS, L. C. B. et al. Early prediction of college attrition using data mining. In: . Piscataway, NJ, USA: IEEE, 2017. v. 2017-December, p. 1075–1078.

MARTINS, T. et al. Explainable artificial intelligence (xai): A systematic literature review on taxonomies and applications in finance. *IEEE Access*, v. 12, p. 618–629, 2024.

MOHALE, V. Z.; OBAGBUWA, I. C. A systematic review on the integration of explainable artificial intelligence in intrusion detection systems to enhancing transparency and interpretability in cybersecurity. *Frontiers in Artificial Intelligence*, v. 8, p. 1526221, 2025.

MOISEEV, I.; BALABAEVA, K. Y.; KOVALCHUK, S. V. Open and extensible benchmark for explainable artificial intelligence methods. *Algorithms*, v. 18, n. 2, p. 85, 2025.

MOURÃO, É.; OLIVEIRA, L.; FIGUEIREDO, F. Hybrid search strategies for systematic reviews in computing education: combining database searches and snowballing. In: *Anais do Workshop de Informática na Escola (WIE)*. Porto Alegre, RS, Brasil: Sociedade Brasileira de Computação, 2020. Disponível em: ⟨https://sol.sbc.org.br/index.php/wie/article/view/11385⟩.

MURESAN, A.; CARDEI, M.; CARDEI, I. Predicting student success with heterogeneous graph deep learning and machine learning models. In: MILLS, C. et al. (Ed.). *Proceedings of the 18th International Conference on Educational Data Mining*. Palermo, Italy: International Educational Data Mining Society, 2025. p. 265–275. ISBN 978-1-7336736-6-2. Disponível em: ⟨https://educationaldatamining.org/EDM2025/proceedings/2025.EDM.long-papers.38/index.html⟩. Acesso em: 2 jun. 2026.

NAGY, B.; MOLONTAY, R. Interpretable dropout prediction: Towards xai-based personalized intervention. *International Journal of Artificial Intelligence in Education*, v. 34, p. 274–300, 2024. Disponível em: ⟨https://jedm.educationaldatamining.org/index.php/JEDM/article/view/744⟩.

NIZAM, T.; ZAFAR, S. Explainable artificial intelligence (xai): Conception, visualization and assessment approaches towards amenable xai. In: *Studies in Computational Intelligence*. Cham, Switzerland: Springer, 2023. v. 1072, p. 35–51.

NUGRAHA, B.; JNANASHREE, A. V.; BAUSCHERT, T. An efficient explainable artificial intelligence (xai)-based framework for a robust and explainable ids. In: . Piscataway, NJ, USA: IEEE, 2024. p. 173–181.

OQAIDI, K.; AOUHASSI, S.; MANSOURI, K. Towards a students' dropout prediction model in higher education institutions using machine learning algorithms. *International Journal of Emerging Technologies in Learning*, v. 17, n. 18, p. 103–117, 2022.

ORTIGOSSA, E. S.; GONÇALVES, T.; NONATO, L. G. Explainable artificial intelligence (xai)–from theory to methods and applications. *IEEE Access*, v. 12, p. 80799–80846, 2024.

OSORIO, J. K. H.; SANTACOLOMA, G. D. Predictive model to identify college students with high dropout rates; modelo preditivo para identificar estudantes universitários com alto risco de evasão; modelo predictivo para identificar estudiantes universitarios con alto grado de deserción. *Revista Electronica de Investigacion Educativa*, v. 25, 2023.

PALACIOS, C. A. et al. Knowledge discovery for higher education student retention based on data mining: Machine learning algorithms and case study in chile. *Entropy*, v. 23, n. 4, p. 485, 2021.

PIMENTEL, M. et al. Design science research: princípios, práticas e aplicações. In: *Anais do XXXIX Congresso da Sociedade Brasileira de Computação (CSBC 2020)*. Porto Alegre: Sociedade Brasileira de Computação, 2020. Disponível em: ⟨https://sol.sbc.org.br/index.php/csbc/article/view/10591⟩.

PRAJWAL, P.; SAHANA, L. R.; KANCHANA, V. Forecasting student attrition using machine learning. In: . Piscataway, NJ, USA: IEEE, 2024.

PROKHORENKOVA, L. et al. Catboost: Unbiased boosting with categorical features. In: *Advances in Neural Information Processing Systems (NeurIPS)*. Red Hook, NY, USA: Curran Associates, Inc., 2018. p. 6639–6649. Disponível em: ⟨https://proceedings.neurips.cc/paper/2018/hash/14491b756b3a51daac41c24863285549-Abstract.html⟩.

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*. New York, NY, USA: Association for Computing Machinery, 2016. Disponível em: ⟨https://dl.acm.org/doi/10.1145/2939672.2939778⟩.

SAITO, T.; REHMSMEIER, M. The precision-recall plot is more informative than the roc plot when evaluating binary classifiers on imbalanced datasets. *PLOS ONE*, v. 10, n. 3, 2015.

SALLOUM, S. A. et al. Predicting student retention in higher education using machine learning. In: *Communications in Computer and Information Science*. Cham, Switzerland: Springer, 2024. v. 2162 CCIS, p. 197–206.

SANTOS, R. S. S. D.; PONTI, M. A.; RODRIGUES, K. R. D. H. Analyzing college student dropout risk prediction in real data using walk-forward validation. In: *Lecture Notes in Computer Science*. Cham, Switzerland: Springer, 2023. v. 14195 LNAI, p. 291–305.

SHAFIK, W. Explainable artificial intelligence (xai) for trustworthy ai in 6g networks. In: . Hershey, PA, USA: IGI Global, 2025. p. 189–215.

SHARMA, J. et al. Explainable artificial intelligence (xai) approaches in predictive maintenance: A review. *Recent Patents on Engineering*, v. 18, n. 5, p. 18–26, 2024.

SHARMA, P. Utilizing explainable artificial intelligence to address deep learning in biomedical domain. In: . Boca Raton, FL, USA: CRC Press, 2023. p. 19–38.

SINGHAL, A. et al. Explainable artificial intelligence (xai) model for cancer image classification. *CMES–Computer Modeling in Engineering and Sciences*, v. 141, n. 1, p. 401–441, 2024.

SOKOLOVA, M.; LAPALME, G. A systematic analysis of performance measures for classification tasks. *Information Processing & Management*, v. 45, n. 4, p. 427–437, 2009.

Stanford Center for Biomedical Informatics Research. *Protégé Ontology Editor and Framework*. 2025. Disponível em: ⟨https://protege.stanford.edu/⟩.

VAARMA, M.; LI, H. Predicting student dropouts with machine learning: An empirical study in finnish higher education. *Technology in Society*, v. 76, p. 102474, 2024.

VEMULAPALLI, S. et al. Predicting student performance and academic success in higher education using a hybrid xgboost-lstm model. In: . Piscataway, NJ, USA: IEEE, 2025. p. 1501–1506.

WIRATSIN, I. O.; RAGKHITWETSAGUL, C. Effectiveness of explainable artificial intelligence (xai) techniques for improving human trust in machine learning models: A systematic literature review. *IEEE Access*, v. 13, p. 121326–121350, 2025.

ZADROZNY, B.; ELKAN, C. Transforming classifier scores into accurate multiclass probability estimates. In: *Proceedings of the 8th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. New York, NY, USA: Association for Computing Machinery, 2002. p. 694–699.

ZANELLATI, A.; GORI, M.; FURLANELLO, C. Balancing accuracy and explainability in predictive models for higher education. *IEEE Transactions on Learning Technologies*, v. 17, n. 2, p. 2140–2153, 2024. Disponível em: ⟨https://ieeexplore.ieee.org/document/10337773⟩.

ZHANG, C. A.; CHO, S.; VASARHELYI, M. A. Explainable artificial intelligence (xai) in auditing. *International Journal of Accounting Information Systems*, v. 46, p. 100572, 2022.