



UNIVERSIDADE FEDERAL DO MARANHÃO

Programa de Pós-Graduação em Ciência da Computação

Ronald César Carneiro da Silva

*Modelagem Preditiva Explicável para Apoio à Prevenção da
Evasão em Cursos de Graduação*

São Luís
2026

Ronald César Carneiro da Silva

Modelagem Preditiva Explicável para Apoio à Prevenção da Evasão em Cursos de Graduação

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Programa de Pós-Graduação em Ciência da Computação

Universidade Federal do Maranhão

Orientador: Prof. Dr. Francisco Glaubos Nunes Clímaco

Coorientador: Profa. Dra. Alana de Araújo Oliveira Meireles Teixeira

São Luís - MA

2026

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).
Diretoria Integrada de Bibliotecas/UFMA

César Carneiro da Silva, Ronald.

Modelagem Preditiva Explicável para Apoio à Prevenção da Evasão em Cursos de Graduação / Ronald César Carneiro da Silva. - 2026.

77 f.

Coorientador(a) 1: Alana de Araújo Oliveira Meireles.

Orientador(a): Francisco Glaubos Nunes Clímaco.

Dissertação (Mestrado) - Programa de Pós-graduação em Ciência da Computação/ccet, Universidade Federal do Maranhão, Universidade Federal do Maranhão, 2026.

1. Inteligência Artificial. 2. Evasão Acadêmica. 3. Modelagem Preditiva. I. Oliveira Meireles, Alana de Araújo. II. Glaubos Nunes Clímaco, Francisco. III. Título.

Ronald César Carneiro da Silva

Modelagem Preditiva Explicável para Apoio à Prevenção da Evasão em Cursos de Graduação

Dissertação apresentada como requisito parcial para obtenção do título de Mestre em Ciência da Computação, ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal do Maranhão.

Prof. Dr. Francisco Glaubos Nunes Clímaco
Orientador
Universidade Federal do Maranhão

Profa. Dra. Alana de Araújo Oliveira Meireles Teixeira
Coorientador
Universidade Federal do Maranhão

Prof. Dr. Tiago Bonini Borchardt
Examinador Interno
Universidade Federal do Maranhão

Prof. Dr. Sergio Teixeira de Carvalho
Examinador Externo
Universidade Federal de Goiás

São Luís - MA
2026

Aos meus imprescindíveis ... Dedicatória

Agradecimentos

Primeiramente, agradeço a Deus por me conceder força, disposição e resiliência nos momentos de incerteza ao longo desta jornada.

À minha família, pelo apoio constante, em especial à minha esposa e à minha filha, que estiveram ao meu lado, me encorajando e dando sentido a cada etapa deste desafio.

Ao meu orientador, Prof. Dr. Francisco Glaubos, pela orientação, pela paciência e pela compreensão diante das minhas dificuldades, sempre contribuindo de forma decisiva para a realização desta pesquisa.

À minha coorientadora, Profa. Dra. Alana Oliveira, pela sensibilidade, pelo olhar técnico e pelo equilíbrio nas discussões, contribuindo de maneira fundamental para o refinamento das ideias e o direcionamento desta pesquisa.

A todos os professores do Programa de Pós-Graduação em Ciência da Computação, que fizeram parte da minha formação acadêmica e contribuíram para o meu desenvolvimento ao longo do mestrado.

Por fim, à Universidade Federal do Maranhão que me deu oportunidade para aprimorar o meu desenvolvimento pessoal e profissional.

"Educação não transforma o mundo. Educação muda as pessoas. Pessoas transformam o mundo."

(Paulo Freire)

Resumo

A evasão acadêmica ainda é um desafio recorrente nas instituições de ensino superior, gerando impactos pessoais para os estudantes e consequências sociais e institucionais relevantes. Nos últimos anos, modelos de aprendizado de máquina têm sido utilizados para estimar o risco de desligamento a partir de dados acadêmicos, demográficos e socioeconômicos dos discentes. Entretanto, apesar do bom desempenho preditivo, muitos desses modelos operam com baixa transparência, o que dificulta interpretar seus resultados. Nesse cenário, abordagens de Inteligência Artificial Explicável (XAI) surgem como alternativa para tornar as previsões mais compreensíveis e úteis na prática.

Esta dissertação investiga como técnicas de Inteligência Artificial Explicável podem ser aplicadas a modelos preditivos para estimar o risco de desligamento acadêmico em cursos de graduação, com o objetivo de identificar os fatores que mais influenciam as decisões dos modelos e analisar o potencial das explicações geradas como suporte ao planejamento de estratégias institucionais de prevenção à evasão.

A pesquisa foi conduzida a partir de dados de cursos de graduação presenciais da Universidade Federal do Maranhão - UFMA e adotou uma abordagem analítica estruturada que combina análise exploratória, modelagem preditiva segmentada por curso e técnicas de explicabilidade global e local. Foram utilizados diferentes algoritmos de aprendizado de máquina para estimar o risco de desligamento e métodos da XAI como *SHAP*, *Permutation Importance*, *LIME* e *Anchors* foram aplicados para interpretar as previsões e analisar a contribuição das variáveis nas decisões dos modelos.

Os resultados indicam que os modelos preditivos apresentam bom desempenho na identificação de estudantes em risco, embora com variações entre cursos. As análises explicativas mostraram convergência entre técnicas ao apontar três dimensões acadêmicas como centrais na estrutura das previsões: a eficiência na conversão da carga horária matriculada em aprovações (IECH), o ritmo de progressão ao longo do curso (IEPL) e o desempenho global representado pela média de conclusão (MC). No contexto analisado, as variáveis demográficas apresentaram influência menos significativa nas previsões.

Além disso os resultados sugerem que integrar modelagem preditiva e técnicas de explicabilidade torna o risco de desligamento mais interpretável, ao evidenciar padrões de desempenho acadêmico associados ao evento. As explicações ajudam a compreender quais fatores sustentam as estimativas de risco em diferentes cursos. Com isso, a abordagem pode apoiar a gestão acadêmica na identificação de perfis que demandam maior atenção e no planejamento de ações institucionais voltadas à permanência estudantil.

Palavras-chave: Inteligência Artificial; Evasão Acadêmica; Modelagem Preditiva; XAI; SHAP.

Abstract

Academic dropout remains a recurring challenge in higher education institutions, generating personal impacts for students as well as relevant social and institutional consequences. In recent years, machine learning models have been used to estimate the risk of academic dropout based on students' academic, demographic, and socioeconomic data. However, despite their good predictive performance, many of these models operate with limited transparency, making it difficult to interpret their results. In this context, Explainable Artificial Intelligence (XAI) approaches emerge as an alternative to make predictions more understandable and practically useful.

This dissertation investigates how Explainable Artificial Intelligence techniques can be applied to predictive models to estimate the risk of academic dropout in undergraduate programs, aiming to identify the factors that most influence model decisions and to analyze the potential of the generated explanations as support for planning institutional strategies to prevent dropout.

The research was conducted using data from on-campus undergraduate programs at the Federal University of Maranhão (UFMA) and adopted a structured analytical approach that combines exploratory analysis, course-segmented predictive modeling, and global and local explainability techniques. Different machine learning algorithms were used to estimate dropout risk, and XAI methods such as SHAP, Permutation Importance, LIME, and Anchors were applied to interpret predictions and analyze the contribution of variables to model decisions.

The results indicate that predictive models show good performance in identifying students at risk, although with variations across programs. The explanatory analyses revealed convergence among techniques in identifying three academic dimensions as central to the structure of predictions: efficiency in converting enrolled workload into course approvals (IECH), the pace of progression throughout the program (IEPL), and overall academic performance represented by the completion GPA (MC). In the analyzed context, demographic variables showed less significant influence on predictions.

Furthermore, the results suggest that integrating predictive modeling with explainability techniques makes dropout risk more interpretable by highlighting academic performance patterns associated with the event. The explanations help clarify which factors sustain risk estimates across different programs. Consequently, this approach can support academic management in identifying student profiles that require greater attention and in planning institutional actions aimed at promoting student retention.

Keywords: Artificial Intelligence; Academic Dropout; Predictive Modeling; Explainable AI; SHAP.

Lista de ilustrações

Figura 1 – Fluxograma da metodologia proposta.	36
Figura 2 – Situação final dos estudantes por curso (2011–2025)	44
Figura 3 – Distribuição do tempo de permanência (em períodos) dos alunos desligados por curso. As linhas pontilhadas indicam os quartis (Q1 e Q3) e a mediana (Q2), utilizados para definição das faixas de desligamento precoce, típico e tardio.	47
Figura 4 – Matrizes de correlação de Pearson entre os indicadores acadêmicos (MC, IECH, IEPL e IEA) no subgrupo de estudantes desligados, por curso.	50
Figura 5 – Distribuição de AUC-PR por Modelo e Curso (via Bootstrap)	53
Figura 6 – Distribuição das probabilidades previstas de desligamento por curso.	55
Figura 7 – Importância global das variáveis estimada por meio da média dos valores absolutos de SHAP (mean SHAP value) para cada curso	58
Figura 8 – Importância das variáveis estimada por Permutation Importance, mensurada pela queda média na AUC-PR após permutação aleatória das variáveis no conjunto de teste	59
Figura 9 – Explicações locais (<i>SHAP, LIME e Anchors</i>) para o caso <i>Verdadeiro Positivo</i> na faixa de Alto Risco (CCO).	62
Figura 10 – Explicações locais (<i>SHAP, LIME e Anchors</i>) para o caso <i>Falso Positivo</i> na faixa de Alto Risco (CCO).	63
Figura 11 – Regiões empíricas de risco – Faixa Alto Risco (CCO)	64
Figura 12 – Explicações locais (<i>SHAP, LIME e Anchors</i>) para o caso <i>Verdadeiro Negativo</i> na faixa de Alto Risco (CCO).	66
Figura 13 – Explicações locais (<i>SHAP, LIME e Anchors</i>) para o caso <i>Falso Negativo</i> na faixa de Alto Risco (CCO).	67
Figura 14 – Regiões empíricas de risco – Faixa Baixo Risco (CCO)	68

Lista de tabelas

Tabela 1 – Métodos de explicabilidade aplicados neste estudo.	29
Tabela 2 – Síntese comparativa de estudos empíricos sobre predição de evasão acadêmica com métodos de XAI.	34
Tabela 3 – Variáveis utilizadas no estudo	38
Tabela 4 – Distribuição das variáveis demográficas por curso e situação final	45
Tabela 5 – Distribuição da variável faixa etária por curso e situação final	46
Tabela 6 – Distribuição da variável região de origem por curso e situação final . . .	47
Tabela 7 – Distribuição do tempo de permanência dos alunos desligados por curso	48
Tabela 8 – Estatísticas descritivas dos indicadores de desempenho por curso e situação acadêmica	49
Tabela 9 – Desempenho dos modelos preditivos por curso (AUC-PR com IC 95%).	52
Tabela 10 – Matriz de classificação dos modelos selecionados por curso (classe positiva = Desligado, limiar = 0,5).	54
Tabela 11 – Perfis de risco estimados pelo modelo por curso com intervalos 10–90 dos indicadores e métricas agregadas de desempenho.	56
Tabela 12 – Correlação de Spearman entre os rankings globais dos indicadores por curso.	59
Tabela 13 – Atributos do caso <i>Verdadeiro Positivo</i> na faixa de Alto Risco (CCO). . . .	61
Tabela 14 – Atributos do caso <i>Falso Positivo</i> na faixa de Alto Risco (CCO).	62
Tabela 15 – Atributos do caso <i>Verdadeiro Negativo</i> na faixa de Baixo Risco (CCO). . . .	65
Tabela 16 – Atributos do caso <i>Falso Negativo</i> na faixa de Baixo Risco (CCO). . . .	66

Lista de abreviaturas e siglas

AUC-PR	<i>Área sob a Curva Precisão-Revocação</i>
BICT	<i>Bacharelado Interdisciplinar em Ciência e Tecnologia</i>
CCO	<i>Ciência da Computação</i>
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i>
EDA	<i>Análise Exploratória de Dados</i>
ENF	<i>Enfermagem</i>
IA	<i>Inteligência Artificial</i>
IEA	<i>Índice de Eficiência Acadêmica</i>
IECH	<i>Índice de Eficiência em Carga Horária</i>
IEPL	<i>Índice de Eficiência em Períodos Letivos</i>
LIME	<i>Local Interpretable Model-Agnostic Explanations</i>
MC	<i>Média de Conclusão</i>
ML	<i>Machine Learning</i>
MLP	<i>Multilayer Perceptron</i>
OECD	<i>Organisation for Economic Co-operation and Development</i>
ODO	<i>Odontologia</i>
PI	<i>Permutation Importance</i>
SHAP	<i>SHapley Additive exPlanations</i>
SIGAA	<i>Sistema Integrado de Gestão de Atividades Acadêmicas</i>
SMOTE	<i>Synthetic Minority Over-sampling Technique</i>
XAI	<i>Explainable Artificial Intelligence</i>

Sumário

1	INTRODUÇÃO	15
1.1	Objetivos	16
1.1.1	Objetivos Específicos	16
1.1.2	Contribuições	17
1.2	Organização do Trabalho	17
2	FUNDAMENTAÇÃO TEÓRICA	19
2.1	Evasão no ensino superior	19
2.1.1	Conceitos e definições operacionais	19
2.1.2	Fatores associados à evasão	20
2.2	Aprendizado de máquina aplicado à predição de evasão	21
2.2.1	Aprendizado supervisionado em contextos educacionais	21
2.2.2	Risco acadêmico, desbalanceamento e limites preditivos	22
2.3	Inteligência Artificial Explicável em modelos preditivos educacionais	23
2.3.1	XAI em contextos educacionais: requisitos, desafios e limites	24
2.3.2	Explicabilidade e interpretabilidade: distinções conceituais	24
2.3.3	Métodos de explicabilidade pós-hoc: categorias e adequação	25
2.3.3.1	LIME (Local Interpretable Model-Agnostic Explanations)	26
2.3.3.2	SHAP (SHapley Additive exPlanations)	26
2.3.3.3	Anchors (Explicações Baseadas em Regras)	27
2.3.3.4	Permutation Importance (PI)	28
3	TRABALHOS RELACIONADOS	30
3.1	Predição de evasão no ensino superior com aprendizado de máquina	30
3.1.1	Estudos clássicos e linhas de base interpretáveis	30
3.1.2	Modelos mais complexos e limitações de transparência	31
3.2	XAI aplicada à evasão acadêmica	31
3.3	Síntese comparativa e posicionamento do estudo	33
4	METODOLOGIA	35
4.1	Visão Geral da Metodologia e Caracterização do Estudo	35
4.2	Compreensão dos Dados	36
4.3	Preparação dos Dados para Modelagem	39
4.4	Modelagem Preditiva	40
4.5	Explicabilidade dos Modelos	41

5	RESULTADOS E DISCUSSÃO	43
5.1	Caracterização exploratória dos dados e padrões iniciais	43
5.1.1	Distribuição dos estudantes por curso e situação acadêmica	43
5.1.2	Distribuição das variáveis demográficas	44
5.1.3	Variáveis derivadas e caracterização da trajetória acadêmica	45
5.1.4	Associações entre indicadores de desempenho e implicações para a modelagem	48
5.1.5	Síntese integradora e critérios de seleção de variáveis	50
5.2	Desempenho preditivo e estratificação por risco	51
5.2.1	Comparação de modelos por curso	51
5.2.2	Análise complementar de erros e limiar fixo (0,5)	53
5.2.3	Faixas de risco e leitura aplicada das probabilidades	55
5.3	Explicabilidade dos Modelos	57
5.3.1	Explicabilidade Global do Risco por Curso	57
5.3.2	Explicabilidade Local Estratificada por Faixa de Risco	60
5.3.2.1	Explicabilidade Local - Faixa de Alto Risco	61
5.3.2.1.1	Caso 1: acerto na predição(<i>Verdadeiro Positivo</i>)	61
5.3.2.1.2	Caso 2: erro na predição(<i>Falso Positivo</i>)	61
5.3.2.1.3	Síntese da faixa de alto risco	63
5.3.2.2	Explicabilidade Local – Faixa de Baixo Risco	64
5.3.2.2.1	Caso 1: acerto na predição(<i>Verdadeiro Negativo</i>)	65
5.3.2.2.2	Caso 2: erro na predição(<i>Falso Negativo</i>)	65
5.3.2.2.3	Síntese da faixa de baixo risco	67
5.4	Síntese dos resultados	68
6	CONCLUSÃO	70
	REFERÊNCIAS	72

1 Introdução

A evasão acadêmica continua sendo uma preocupação no ensino superior brasileiro, com impactos direto sobre a gestão universitária, o desempenho dos estudantes e a própria sociedade. Relatórios recentes da Organisation for Economic Co-operation and Development (OECD) indicam que as taxas de conclusão da graduação seguem desiguais entre países e grupos sociais, refletindo barreiras estruturais associadas à permanência estudantil (OECD, 2023). No contexto brasileiro, dados do Instituto Semesp apontam que, em 2023, a evasão em cursos presenciais de universidades públicas atingiu 23,1%, resultando em perdas financeiras expressivas e comprometendo políticas voltadas à democratização do acesso e da permanência no ensino superior (SEMESP, 2025).

No âmbito institucional, o cenário é igualmente preocupante. Dados do Painel Estatístico do Censo da Educação Superior, disponibilizados pelo Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (Inep, 2025), indicam que, em 2024, a taxa de desistência acumulada entre estudantes da Universidade Federal do Maranhão atingiu aproximadamente 46%. Esse percentual, de natureza agregada e descritiva, sinaliza o tamanho do problema no contexto da instituição e reforça a relevância de estudos que investiguem, de forma estruturada, os fatores associados à evasão acadêmica e suas possíveis estratégias de mitigação.

Além dos impactos institucionais, a não conclusão do ensino superior está associada a consequências individuais de longo prazo, como menor empregabilidade, redução de renda e prejuízos ao bem-estar (NEUGEBAUER et al., 2025). Nesse cenário, torna-se relevante dispor de instrumentos que viabilizem a identificação precoce de estudantes em risco, de modo a subsidiar intervenções pedagógicas e ações de apoio mais eficazes e direcionadas.

Nos últimos anos, técnicas de Inteligência Artificial (IA), especialmente métodos de *Machine Learning* (ML), têm sido investigadas como ferramentas de apoio à gestão acadêmica. Estudos anteriores utilizaram algoritmos de árvores de decisão, redes neurais e métodos de *ensemble* para estimar o risco de desligamento a partir de informações acadêmicas e demográficas (PÉREZ; CASTELLANOS; CORREAL, 2018; ULLOA-CÁZAREZ; GARCÍA-DÍAZ; SORIANO-EQUIGUA, 2021; NIYOGISUBIZO et al., 2022; KOCSIS; MOLNÁR, 2025). Embora tais abordagens apresentem bom desempenho preditivo, revisões da literatura destacam que a limitada transparência desses modelos pode comprometer sua adoção prática em ambientes educacionais, nos quais interpretabilidade, confiabilidade e responsabilidade ética são requisitos centrais para o uso de sistemas baseados em IA (TETE et al., 2022; RODRIGUES et al., 2024; NAGY; MOLONTAY, 2024).

Nesse contexto, métodos de Inteligência Artificial Explicável (XAI) surgem como alternativa para aproximar modelos preditivos do cotidiano institucional. Técnicas como *SHapley Additive exPlanations* (SHAP) e *Local Interpretable Model-Agnostic Explanations* (LIME) possibilitam compreender como variáveis distintas contribuem para as estimativas do modelo, favorecendo análises mais transparentes (NAGY; MOLONTAY, 2024). Apesar desse potencial, ainda são escassos os estudos que investigam, de forma sistemática, como explicações podem ser integradas à análise da evasão acadêmica e utilizada para apoiar o planejamento de ações institucionais voltadas à permanência estudantil.

Diante dessas lacunas, esta dissertação investiga a aplicação de técnicas de Inteligência Artificial Explicável (XAI) em modelos preditivos no contexto da Universidade Federal do Maranhão, com base em dados de cursos presenciais. Neste trabalho, a evasão é caracterizada operacionalmente pelo desligamento do curso, conforme registro institucional. O estudo adota uma abordagem analítica estruturada e reproduzível, combinando modelagem preditiva segmentada por curso e a aplicação integrada de métodos globais e locais de explicabilidade, buscando interpretações contextualizadas dos resultados.

1.1 Objetivos

O objetivo geral desta dissertação é investigar como técnicas de Inteligência Artificial Explicável podem ser aplicadas a modelos preditivos de evasão acadêmica em cursos de graduação, com o propósito de tornar as decisões desses modelos mais interpretáveis e identificar os principais fatores associados ao risco de desligamento acadêmico no contexto da Universidade Federal do Maranhão (UFMA).

1.1.1 Objetivos Específicos

Especificamente, este trabalho busca os seguintes objetivos aplicados ao problema da evasão acadêmica em cursos de graduação:

- analisar padrões de desempenho acadêmico associados ao desligamento de estudantes em cursos de graduação da UFMA;
- desenvolver modelos preditivos de evasão acadêmica utilizando dados institucionais, por meio da construção de modelos independentes para cada curso de graduação, considerando as diferenças nos padrões de desempenho acadêmico e nas trajetórias estudantis entre os cursos analisados.
- aplicar técnicas de Inteligência Artificial Explicável, incluindo métodos globais e locais, para interpretar as previsões dos modelos e analisar a contribuição das variáveis nas decisões algorítmicas;

- analisar como os padrões explicativos identificados podem contribuir para a compreensão dos perfis de risco de evasão e para o apoio à gestão acadêmica no acompanhamento do desempenho estudantil.

1.1.2 Contribuições

Destacam-se como principais contribuições desta dissertação:

- a aplicação integrada de técnicas de aprendizado de máquina e Inteligência Artificial Explicável para análise do risco de evasão acadêmica utilizando dados institucionais reais de cursos de graduação;
- a proposição de um processo analítico estruturado que combina análise exploratória, modelagem preditiva segmentada por curso e técnicas de explicabilidade global e local;
- a utilização e comparação de diferentes métodos de explicabilidade, incluindo *SHAP*, *Permutation Importance*, *LIME* e *anchors*, para interpretar o comportamento dos modelos preditivos;
- a identificação de padrões de desempenho acadêmico associados ao risco de evasão, evidenciando o papel de indicadores acadêmicos institucionais na estrutura das previsões dos modelos;
- a produção de evidências interpretativas que podem apoiar gestores acadêmicos na compreensão do risco de evasão e no acompanhamento de trajetórias acadêmicas dos estudantes.

1.2 Organização do Trabalho

Este trabalho está estruturado da seguinte forma:

- O Capítulo 2 aborda a fundamentação teórica necessária para o desenvolvimento do estudo, incluindo conceitos relacionados a evasão, aprendizado de máquina e técnicas de explicabilidade.
- O Capítulo 3 apresenta os trabalhos relacionados, situando esta pesquisa no contexto da literatura sobre evasão acadêmica, modelagem preditiva e Inteligência Artificial Explicável aplicada à educação.
- O Capítulo 4 descreve a metodologia adotada, detalhando as etapas do processo analítico utilizado para preparação dos dados, desenvolvimento dos modelos preditivos e aplicação das técnicas de explicabilidade.

- O Capítulo 5 apresenta os resultados obtidos nos experimentos, integrando a avaliação do desempenho dos modelos com as análises de explicabilidade global e local.
- O Capítulo 6 apresenta as considerações finais da pesquisa, destacando os principais resultados do estudo, suas limitações e as perspectivas para trabalhos futuros.

2 Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica que sustenta as escolhas conceituais e analíticas adotadas nesta dissertação. Inicialmente, são discutidos os conceitos de evasão no ensino superior e os principais fatores associados ao abandono acadêmico. Em seguida, são apresentados os fundamentos do uso de técnicas de aprendizado de máquina na predição de evasão, destacando seus paradigmas, potencialidades e limitações. Por fim, discute-se o papel da Inteligência Artificial Explicável (XAI) com ênfase em seus princípios conceituais e nas distinções teóricas que orientam sua utilização em modelos preditivos aplicados à educação.

2.1 Evasão no ensino superior

2.1.1 Conceitos e definições operacionais

A evasão no ensino superior refere-se, de modo geral, à saída antecipada do estudante de seu curso ou instituição antes da conclusão, caracterizando o rompimento do vínculo acadêmico sem a obtenção do diploma (KOCISIS; MOLNÁR, 2025). Embora essa definição seja amplamente aceita, as definições operacionais de evasão podem variar conforme o órgão responsável pela produção de dados e os objetivos analíticos de cada estudo (Universidade Pública do Brasil no Sul, 2011).

No âmbito das estatísticas oficiais brasileiras, o Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira (INEP) adota uma definição abrangente, segundo a qual é considerado evadido o estudante que não permanece matriculado no curso no ano subsequente, sem ter concluído a graduação, independentemente do motivo do desligamento (SILVA; MARIANO, 2021). Essa definição permite o monitoramento agregado do fenômeno em escala nacional. No entanto, tende a agrupar situações distintas sob uma mesma categoria analítica, o que pode limitar análises mais contextualizadas.

Buscando maior refinamento conceitual, a Associação Nacional dos Dirigentes das Instituições Federais de Ensino Superior (ANDIFES) propõe uma tipologia que distingue três níveis de evasão: evasão de curso, evasão institucional e evasão do sistema. A evasão de curso corresponde à saída definitiva do estudante de um curso específico sem concluí-lo; a evasão institucional refere-se ao desligamento do estudante da instituição de ensino superior, ainda que venha a ingressar em outra; e a evasão do sistema caracteriza o afastamento do estudante do sistema de ensino superior como um todo, de forma temporária ou definitiva.

Essa distinção é relevante, pois permite diferenciar formas de evasão com implicações analíticas e institucionais distintas. Considerando os objetivos desta dissertação e a natureza

dos dados analisados, adota-se neste estudo a definição de evasão de curso como a saída definitiva do estudante de seu curso de origem sem a respectiva conclusão, conforme a tipologia proposta pela ANDIFES (Brasil. Ministério da Educação, 1997).

2.1.2 Fatores associados à evasão

A evasão no ensino superior é frequentemente descrita na literatura como um fenômeno multifatorial e complexo. O abandono da graduação raramente ocorre em função de um único motivo isolado, estando geralmente associado à combinação de diferentes fatores que se acumulam ao longo da trajetória acadêmica do estudante (TINTO, 1975; FILHO et al., 2007; SANTOS; PILATTI; BONDARIK, 2022; BARROSO et al., 2022; KOCSIS; MOLNÁR, 2025). Nesse sentido, compreender a evasão exige considerar a interação entre dimensões acadêmicas, institucionais, socioeconômicas e individuais. Conforme argumenta Filho et al. (2007), a diversidade das razões que levam ao abandono torna o problema difícil de ser enfrentado por meio de soluções simples ou isoladas.

Entre os fatores socioeconômicos mais frequentemente associados à evasão destacam-se dificuldades financeiras, a necessidade de conciliar trabalho e estudo e a ausência de apoio material. Essas condições tendem a afetar de forma mais intensa estudantes em situação de maior vulnerabilidade social (AMBIEL; CORTEZ; SALVADOR, 2021; GOMES; HIRATA, 2022). No campo acadêmico, pesquisas indicam que a insuficiente preparação prévia, as dificuldades de adaptação às exigências do curso, os elevados índices de reprovação e o baixo rendimento acadêmico estão fortemente relacionados ao risco de abandono, especialmente em cursos caracterizados por currículos mais exigentes, como aqueles das áreas de Engenharias e Ciências Exatas (GOMES; HIRATA, 2022; NIEROTKA; SALATA; MARTINS, 2023).

Além desses aspectos, fatores institucionais também podem influenciar a permanência dos estudantes. Elementos como a organização curricular, as regras de progressão acadêmica e as condições de apoio oferecidas pelas instituições podem facilitar ou dificultar a continuidade dos estudos. De maneira complementar, características demográficas como sexo, faixa etária, cor ou raça, local de residência e forma de ingresso podem atuar como marcadores de desigualdades sociais e institucionais, associando-se indiretamente às trajetórias de permanência ou abandono no ensino superior (SILVA et al., 2022; BARROSO et al., 2022).

A literatura também destaca a relevância de fatores individuais, como identificação com o curso escolhido, motivação acadêmica, dificuldades de adaptação ao ambiente universitário e questões relacionadas à saúde mental (FILHO et al., 2007; AMBIEL; CORTEZ; SALVADOR, 2021). Embora esses aspectos tenham influência significativa na trajetória discente, sua mensuração sistemática costuma ser limitada em bases de dados administrativas, o que representa uma restrição recorrente em estudos quantitativos sobre

evasão.

A identificação sistemática desses fatores ao longo da trajetória acadêmica é fundamental para o acompanhamento de estudantes em situação de maior vulnerabilidade. Informações relacionadas ao desempenho acadêmico, à progressão curricular e às condições institucionais podem fornecer subsídios importantes para o monitoramento do risco de evasão, permitindo que gestores e equipes pedagógicas atuem de forma mais preventiva e orientada (KOCISIS; MOLNÁR, 2025).

2.2 Aprendizado de máquina aplicado à predição de evasão

O aprendizado de máquina reúne abordagens computacionais voltadas à identificação de padrões em dados e à construção de modelos preditivos a partir de observações históricas. Essas abordagens têm sido amplamente utilizadas em situações nas quais as relações entre variáveis são complexas e difíceis de representar por regras explícitas (GÉRON, 2021). No contexto educacional, seu uso tem se ampliado como apoio ao monitoramento acadêmico, permitindo analisar grandes volumes de dados institucionais e estimar a probabilidade de eventos como evasão, retenção ou conclusão de cursos (NIYOGISUBIZO et al., 2022).

Ao mesmo tempo, a aplicação de aprendizado de máquina em educação impõe limites metodológicos que precisam ser explicitados. As inferências produzidas pelos modelos dependem diretamente da qualidade, da representatividade e da abrangência dos dados institucionais utilizados. Em geral, esses dados não contemplam dimensões subjetivas, como motivação, engajamento ou condições psicossociais dos estudantes. Por isso, os resultados devem ser interpretados como instrumentos de apoio à decisão e ao acompanhamento institucional, e não como explicações causais ou determinísticas sobre a evasão (BONIFRO et al., 2020).

Nesse sentido, o aprendizado de máquina aplicado à predição de evasão pode ser entendido como um recurso analítico voltado à identificação de padrões de risco ao longo da trajetória acadêmica. Seu papel não é substituir a análise pedagógica ou institucional, mas oferecer evidências adicionais que auxiliem gestores e equipes acadêmicas no acompanhamento de estudantes em situação de maior risco.

2.2.1 Aprendizado supervisionado em contextos educacionais

O aprendizado supervisionado é um dos paradigmas centrais do aprendizado de máquina e tem sido amplamente utilizado em aplicações educacionais voltadas à predição de desfechos acadêmicos, como evasão, retenção e desempenho discente. Nesse paradigma, os modelos são treinados a partir de dados históricos rotulados, nos quais as variáveis de entrada estão associadas a uma saída previamente conhecida (BISHOP, 2006; HASTIE; TIBSHIRANI; FRIEDMAN, 2009).

No contexto educacional, essas variáveis de entrada geralmente representam características acadêmicas, demográficas ou administrativas dos estudantes, enquanto a saída corresponde a um desfecho observável, como a conclusão ou o desligamento do curso. A partir desse processo, os modelos aprendem padrões presentes nas trajetórias acadêmicas e passam a estimar o comportamento de estudantes com perfis semelhantes.

Segundo [Quimiz-Moreira et al. \(2025\)](#), a predição de evasão com aprendizado supervisionado envolve um conjunto diverso de técnicas, que vai de modelos mais simples, como regressão logística e árvores de decisão, até abordagens mais complexas, como florestas aleatórias e redes neurais artificiais. Em aplicações educacionais, essas técnicas são utilizadas para estimar, com base em dados institucionais rotulados, a probabilidade de um estudante apresentar risco de evasão. Por essa razão, o aprendizado supervisionado constitui a base de muitos sistemas de monitoramento acadêmico orientados à identificação precoce de situações de vulnerabilidade.

2.2.2 Risco acadêmico, desbalanceamento e limites preditivos

Na predição de evasão acadêmica, o conceito de risco deve ser compreendido como uma estimativa probabilística construída a partir de variáveis observadas e de padrões identificados em dados históricos. Isso significa que o risco não expressa uma condição fixa do estudante, mas uma probabilidade associada ao conjunto de informações disponíveis no momento da modelagem. Em contextos educacionais, essa distinção é importante porque as trajetórias acadêmicas são influenciadas por fatores múltiplos e dinâmicos, muitos dos quais não estão registrados nas bases institucionais ([BONIFRO et al., 2020](#)). Por essa razão, as estimativas produzidas pelos modelos devem ser interpretadas como apoio analítico ao acompanhamento acadêmico, e não como diagnósticos conclusivos sobre o comportamento futuro dos estudantes.

Outro aspecto central nesse tipo de modelagem é o desbalanceamento de classes. Em bases educacionais, é comum que a evasão represente a classe minoritária em relação à conclusão do curso. Esse cenário afeta diretamente o treinamento e a avaliação dos modelos, pois métricas globais como a acurácia podem sugerir bom desempenho mesmo quando o modelo falha justamente na identificação dos casos de maior interesse [Teodoro e Kappel \(2020\)](#). Em problemas dessa natureza, torna-se necessário adotar métricas e estratégias de avaliação mais sensíveis à assimetria entre as classes, de modo a considerar adequadamente o impacto dos erros de classificação.

Além disso, é fundamental distinguir predição estatística de inferência causal. Modelos de aprendizado de máquina identificam associações entre variáveis observadas e desfechos históricos, mas não permitem afirmar, por si só, que determinada variável causa a evasão ([TEODORO; KAPPEL, 2020](#)). Essa distinção é particularmente importante em educação, onde interpretações causais indevidas podem levar à estigmatização de

estudantes ou à formulação de intervenções baseadas em pressupostos não sustentados pelos dados. Assim, o uso de modelos preditivos exige cautela interpretativa, eles podem produzir evidências úteis para o monitoramento institucional, mas seus resultados precisam ser lidos dentro dos limites estatísticos e informacionais da base analisada.

2.3 Inteligência Artificial Explicável em modelos preditivos educacionais

A Inteligência Artificial Explicável (do inglês *Explainable Artificial Intelligence* – XAI) refere-se ao conjunto de abordagens desenvolvidas para tornar os resultados de modelos de aprendizado de máquina mais compreensíveis para os humanos (KAADOUD; FAHED; LENCA, 2021). Esses métodos buscam reduzir a opacidade associada a modelos preditivos complexos, oferecendo mecanismos que permitem interpretar como determinadas previsões são produzidas (MELO et al., 2022). Conforme discutido por Fiok et al. (2022), o interesse crescente por XAI está relacionado à difusão de técnicas de aprendizado de máquina cujo funcionamento interno nem sempre pode ser interpretado diretamente pelos usuários.

Nesse cenário, a explicabilidade surge como um complemento ao desempenho estatístico dos modelos. Em vez de apenas gerar estimativas, sistemas baseados em XAI procuram indicar quais fatores contribuíram para uma determinada previsão. Essa capacidade amplia a transparência dos resultados e permite examinar criticamente o comportamento dos modelos. Em aplicações nas quais as previsões influenciam processos de tomada de decisão, a possibilidade de interpretar as estimativas torna-se particularmente relevante, pois favorece a auditabilidade e a compreensão dos critérios utilizados pelo modelo (ALTUKHI; PRADHAN, 2025).

No domínio educacional, essa discussão assume implicações específicas. Modelos preditivos utilizados para estimar risco acadêmico integram informações demográficas, acadêmicas e institucionais para produzir estimativas sobre a trajetória dos estudantes. Esses resultados podem apoiar análises institucionais sobre permanência e evasão, mas sua interpretação exige compreender quais padrões foram identificados nos dados.

Dessa forma, a explicabilidade assume um papel relevante no uso responsável de modelos preditivos aplicados à educação. Ao tornar mais visíveis os padrões identificados pelos algoritmos, os métodos de XAI contribuem para qualificar a análise dos resultados produzidos pelos modelos e apoiar a interpretação das estimativas geradas. Por essa razão, estudos recentes destacam que o uso de XAI em educação deve ser compreendido não apenas como um recurso técnico para examinar modelos, mas também como parte de um processo sociotécnico de produção e interpretação de evidências baseadas em dados (KHOSRAVI et al., 2022; FARROW, 2023; SAQR; LÓPEZ-PERNAS, 2024).

2.3.1 XAI em contextos educacionais: requisitos, desafios e limites

A aplicação de técnicas de explicabilidade em contextos educacionais envolve desafios que vão além da interpretação técnica dos modelos. Sistemas de análise educacional operam em ambientes frequentemente caracterizados por dados incompletos ou ruidosos e por decisões que podem impactar diretamente as trajetórias acadêmicas dos estudantes. Nessas condições, a interpretação das previsões não depende apenas dos mecanismos computacionais utilizados pelos modelos, mas também do contexto institucional no qual essas estimativas são produzidas.

Outro aspecto relevante refere-se à diversidade de atores envolvidos no uso dessas explicações. Enquanto cientistas de dados e desenvolvedores tendem a utilizar representações técnicas para compreender o comportamento dos modelos, gestores acadêmicos e docentes geralmente demandam interpretações mais contextualizadas e relacionadas à tomada de decisão. Conforme discutido por [Farrow \(2023\)](#), essa diferença de perspectivas cria desafios adicionais para a comunicação das explicações, uma vez que os resultados precisam ser compreensíveis para públicos com diferentes níveis de familiaridade com métodos analíticos utilizados.

A literatura também aponta limitações técnicas associadas ao uso de métodos de explicabilidade. Uma revisão sistemática conduzida por [Altukhi e Pradhan \(2025\)](#) destaca que algumas técnicas podem apresentar instabilidade, produzindo explicações distintas diante de pequenas variações nos dados ou no modelo. Além disso, determinados métodos explicativos envolvem custos computacionais elevados ou podem introduzir um *trade-off* entre interpretabilidade e desempenho preditivo, o que exige cuidado na escolha das abordagens utilizadas.

Por fim, é importante reconhecer que explicações produzidas por métodos de XAI descrevem associações aprendidas pelos modelos a partir dos dados disponíveis. Conforme discutido por [\(FIOK et al., 2022\)](#), essas explicações não estabelecem relações causais entre variáveis e resultados educacionais. Em contextos institucionais, interpretações que confundem associação com causalidade podem levar a conclusões simplificadas sobre o fenômeno analisado ou orientar intervenções inadequadas.

2.3.2 Explicabilidade e interpretabilidade: distinções conceituais

No campo da Inteligência Artificial Explicável, os termos explicabilidade e interpretabilidade são frequentemente utilizados como sinônimos, embora representem conceitos distintos com implicações relevantes para o desenvolvimento e o uso de modelos de aprendizado de máquina. Em linhas gerais, a explicabilidade está relacionada à capacidade de comunicar, de forma compreensível, as razões associadas às previsões produzidas por um modelo, respondendo à pergunta “por que” uma determinada estimativa foi gerada ([MOLNAR,](#)

2025). Conforme argumenta Miller (2019), explicações eficazes devem considerar aspectos da cognição humana, sendo compreensíveis, relevantes e plausíveis para os usuários que utilizam essas informações no processo de tomada de decisão.

A interpretabilidade, por sua vez, está associada à compreensão do funcionamento de um modelo a partir de sua estrutura interna, incluindo parâmetros, representações e relações entre variáveis (MOLNAR, 2025). De acordo com Lipton (2018), modelos interpretáveis permitem que especialistas examinem como as informações são processadas e como as previsões são produzidas, contribuindo para atividades como auditoria, validação e depuração de sistemas de inteligência artificial.

Embora relacionados, esses conceitos enfatizam dimensões distintas da transparência dos modelos. Enquanto a explicabilidade prioriza a comunicação das decisões do modelo para diferentes públicos, a interpretabilidade está mais diretamente associada à análise técnica de seu funcionamento. Trabalhos como os de Doshi-Velez e Kim (2017) ressaltam que essas duas perspectivas devem ser tratadas de forma complementar, uma vez que sistemas de IA utilizados em contextos sensíveis frequentemente exigem tanto compreensão técnica quanto explicações acessíveis para os usuários finais. Essa distinção também ajuda a compreender por que muitos estudos em XAI recorrem a métodos de explicação pós-hoc para interpretar modelos preditivos complexos.

2.3.3 Métodos de explicabilidade pós-hoc: categorias e adequação

Na área de Inteligência Artificial Explicável, métodos de explicabilidade pós-hoc são amplamente utilizados para interpretar modelos de aprendizado de máquina cujo funcionamento interno não é facilmente compreendido. Esse é o caso de diversos algoritmos modernos, como ensembles e redes neurais, que frequentemente apresentam alto desempenho preditivo, mas oferecem pouca transparência sobre como produzem suas previsões (GUIDOTTI et al., 2018; MOLNAR, 2025). Diferentemente de modelos intrinsecamente interpretáveis, esses métodos são aplicados após o treinamento do modelo e têm como objetivo produzir explicações aproximadas sobre seu comportamento, sem alterar sua estrutura ou seu desempenho.

Uma distinção importante entre esses métodos está relacionada ao nível de análise das explicações. De forma geral, as abordagens de explicabilidade pós-hoc podem ser organizadas em dois níveis complementares: explicações locais e explicações globais. As explicações locais procuram esclarecer predições individuais, indicando por que uma instância específica recebeu determinada classificação. Já as explicações globais buscam descrever o comportamento geral do modelo, identificando padrões de decisão ou estimando a relevância média das variáveis utilizadas nas previsões (MOLNAR, 2025). Muitos desses métodos também são considerados agnósticos ao modelo, pois podem ser aplicados a diferentes algoritmos sem depender diretamente de sua estrutura interna (RIBEIRO;

SINGH; GUESTRIN, 2016).

No contexto desta pesquisa, foram considerados quatro métodos amplamente utilizados na literatura recente sobre explicabilidade aplicada à educação: *LIME*, *SHAP*, *Anchors* e *Permutation Importance*. Essas abordagens apresentam características complementares e permitem analisar o comportamento do modelo sob diferentes perspectivas, combinando explicações individuais e interpretações mais agregadas (SILVA; SANTANA, 2024; NAGY; MOLONTAY, 2024; SILVA et al., 2024; SAARELA; PODGORELEC, 2024).

2.3.3.1 LIME (Local Interpretable Model-Agnostic Explanations)

O método LIME produz explicações locais ao aproximar o comportamento do modelo por meio de um modelo interpretável ajustado na vizinhança de uma instância específica (RIBEIRO; SINGH; GUESTRIN, 2016). A ideia central consiste em gerar pequenas perturbações nos atributos do exemplo analisado e observar como essas variações afetam a predição do modelo. A partir dessas amostras, ajusta-se um modelo linear simples que aproxima o comportamento do modelo original naquele ponto do espaço de dados. O resultado é uma explicação composta por pesos atribuídos às variáveis, indicando quais fatores contribuíram positiva ou negativamente para a predição.

Formalmente, uma explicação local no LIME pode ser definida como a solução do seguinte problema de otimização:

$$\text{Explanation}(x) = \arg \min_{g \in \mathcal{G}} \mathcal{L}(f, g, \pi_x) + \Omega(g)$$

Nessa formulação, f representa o modelo original a ser explicado, enquanto g corresponde a um modelo interpretável pertencente ao conjunto de possíveis explicações \mathcal{G} . A função \mathcal{L} mede o quanto o modelo explicativo g aproxima o comportamento de f na vizinhança da instância analisada, ponderada por π_x , que define a proximidade entre as amostras geradas e o ponto de interesse. O termo $\Omega(g)$ controla a complexidade da explicação, favorecendo modelos mais simples e interpretáveis.

Apesar de produzir explicações intuitivas, o LIME apresenta algumas limitações. Como as explicações dependem de amostragens aleatórias ao redor da instância analisada, pequenas variações na geração das perturbações podem levar a resultados diferentes. Essa característica pode reduzir a estabilidade das explicações, especialmente em cenários com relações altamente não lineares entre variáveis e predição (NAGY; MOLONTAY, 2024; SILVA et al., 2022).

2.3.3.2 SHAP (SHapley Additive exPlanations)

O método SHAP foi proposto por (LUNDBERG; LEE, 2017) e baseia-se na teoria dos jogos cooperativos para quantificar a contribuição de cada variável nas predições

de um modelo. A ideia central consiste em interpretar cada variável como um “jogador” que participa da construção da predição final. Nesse contexto, o valor de cada variável é definido pela contribuição marginal que ela adiciona ao resultado do modelo quando combinada com diferentes subconjuntos das demais variáveis.

O valor SHAP de uma variável i , representado por ϕ_i , é calculado como a média ponderada de sua contribuição marginal em todas as possíveis combinações das demais variáveis:

$$\phi_i = \sum_{S \subseteq F \setminus \{i\}} \frac{|S|!(|F| - |S| - 1)!}{|F|!} [f_{S \cup \{i\}}(x_{S \cup \{i\}}) - f_S(x_S)]$$

Nessa expressão, F representa o conjunto completo de variáveis do modelo e S corresponde a um subconjunto que não contém a variável i . O termo entre colchetes expressa a diferença entre a predição obtida quando a variável está presente e quando está ausente. Ao considerar todas as combinações possíveis, o método produz uma estimativa consistente da contribuição média de cada variável para aquela instância.

Uma vantagem importante do SHAP é a possibilidade de analisar as contribuições das variáveis tanto em nível local quanto global. As explicações locais permitem compreender a predição de um caso específico, enquanto agregações dos valores SHAP possibilitam identificar padrões gerais de influência das variáveis no modelo. Por outro lado, o cálculo exato dos valores de Shapley pode ser computacionalmente custoso em modelos complexos, exigindo aproximações para tornar sua aplicação viável em conjuntos de dados maiores (NAGY; MOLONTAY, 2024; CAMARGOS; SILVEIRA, 2024).

2.3.3.3 Anchors (Explicações Baseadas em Regras)

O método Anchors foi proposto por (RIBEIRO; SINGH; GUESTRIN, 2018) como uma alternativa para produzir explicações locais baseadas em regras lógicas simples. Diferentemente de abordagens que atribuem pesos numéricos às variáveis, o Anchors busca identificar um conjunto mínimo de condições que seja suficiente para “ancorar” a predição do modelo com alta precisão. Em termos práticos, a explicação assume a forma de uma regra do tipo *if-then*, que indica sob quais condições a decisão do modelo tende a permanecer estável.

Formalmente, uma regra \mathcal{A} é considerada uma explicação válida quando a probabilidade de o modelo manter a mesma predição para exemplos que satisfaçam essa condição é maior ou igual a um limiar τ :

$$\text{Prec}(\mathcal{A}) = E_{\mathcal{D}(z|\mathcal{A})} [1\{f(x) = f(z)\}] \geq \tau$$

Nesse caso, f representa o modelo original, enquanto z corresponde a amostras geradas a partir de uma distribuição condicionada à regra \mathcal{A} . A métrica de precisão indica a proporção de exemplos para os quais a predição permanece consistente quando a regra é satisfeita.

A principal vantagem desse método está na clareza das explicações produzidas. Regras condicionais tendem a ser facilmente compreendidas por usuários não técnicos, como gestores ou professores, por se aproximarem de formas tradicionais de raciocínio utilizadas em contextos decisórios. Entretanto, como as regras são construídas para garantir precisão local, elas podem representar apenas uma parte do comportamento do modelo, deixando de considerar outros fatores que também influenciam a predição (MOLNAR, 2025).

2.3.3.4 Permutation Importance (PI)

A Permutation Importance é um método de explicabilidade global que mede a relevância das variáveis a partir do impacto de sua perturbação no desempenho do modelo (MOLNAR, 2025). A ideia consiste em comparar o desempenho do modelo em um conjunto de dados antes e depois de embaralhar aleatoriamente os valores de uma variável específica. Quando a permutação de uma variável provoca uma queda significativa na métrica de avaliação, isso indica que o modelo depende daquela informação para produzir suas previsões.

A importância de uma variável j pode ser expressa pela diferença entre o desempenho do modelo no conjunto original e o desempenho obtido após a permutação dessa variável:

$$PI_j = \mathcal{M}(f, X, y) - \mathcal{M}(f, X_{\pi(j)}, y)$$

Nessa equação, \mathcal{M} representa a métrica de avaliação utilizada, f corresponde ao modelo treinado, X é o conjunto de dados original e $X_{\pi(j)}$ indica o conjunto em que os valores da variável j foram embaralhados. Quanto maior a diferença entre os desempenhos, maior a influência da variável no comportamento do modelo.

Embora seja simples de aplicar e interpretar, esse método apresenta algumas limitações. Em particular, quando variáveis estão fortemente correlacionadas, a permutação de apenas uma delas pode não alterar significativamente o desempenho do modelo, levando à subestimação de sua importância real. Além disso, por se tratar de uma medida agregada, a Permutation Importance não fornece explicações específicas para casos individuais.

A Tabela 1 resume as principais características dos métodos de explicabilidade pós-hoc considerados nesta pesquisa, destacando escopo, forma de apresentação, pontos fortes e limitações. Os métodos apresentam características complementares: *LIME* e *Anchors* são voltados principalmente à interpretação local das predições, enquanto o *SHAP*

permite análises locais e globais, e o *Permutation Importance* oferece uma visão agregada da relevância das variáveis no modelo.

Tabela 1 – Métodos de explicabilidade aplicados neste estudo.

Método	Escopo	Tipo de Explicação	de Forma de Apresentação	Pontos Fortes	Limitações
LIME	Local	Contribuições numéricas	Lista ponderada de atributos	Fácil de aplicar e interpretar em análises individuais	Explicações podem variar conforme as amostras geradas ao redor da instância analisada
SHAP	Local e Global	Contribuições numéricas	Gráficos e valores SHAP	Baseado na teoria dos jogos cooperativos (valores de Shapley), produz explicações consistentes e comparáveis entre variáveis	Alto custo computacional e interpretação mais exigente para usuários não técnicos
anchors	Local	Regras lógicas (if-then)	Texto simbólico	Regras simples que podem ser facilmente interpretadas por humanos	Explicações válidas apenas para regiões específicas do espaço de dados
PI	Global	Relevância geral das variáveis	Ranking de importância baseado na queda de desempenho após permutação	Simples de implementar e interpretar em nível global	Não fornece explicações individuais e pode subestimar a importância de variáveis correlacionadas

Essas abordagens têm sido amplamente utilizadas para investigar como os modelos produzem suas previsões. Em aplicações educacionais, esse tipo de análise é particularmente relevante, ao permitir interpretar as estimativas de risco geradas pelos modelos e compreender melhor os fatores associados às decisões preditivas. Dessa forma, os métodos de explicabilidade pós-hoc oferecem suporte para a análise e a interpretação dos resultados apresentados neste estudo.

3 Trabalhos Relacionados

Este capítulo apresenta e discute os principais trabalhos relacionados à predição de evasão acadêmica no ensino superior, com ênfase em abordagens baseadas em aprendizado de máquina e em métodos da XAI. A revisão é organizada por tópicos, de modo a situar progressivamente o presente estudo na literatura, evidenciando tanto os avanços metodológicos quanto as limitações ainda observadas. São priorizados trabalhos diretamente relacionados ao problema investigado, os quais serão utilizados como base comparativa na análise dos resultados desta dissertação.

3.1 Predição de evasão no ensino superior com aprendizado de máquina

A aplicação de técnicas de aprendizado de máquina à predição de evasão acadêmica constitui uma linha consolidada na literatura, especialmente a partir do final da década de 2000. Estudos partem do pressuposto de que padrões presentes em dados acadêmicos, demográficos ou administrativos podem ser explorados para estimar o risco de desligamento. Ao longo dessa trajetória, observa-se uma evolução tanto na complexidade dos modelos utilizados quanto na natureza dos dados analisados. Esse movimento, por sua vez, torna mais evidente um dilema recorrente na área: o equilíbrio entre desempenho preditivo e transparência das decisões dos modelos.

3.1.1 Estudos clássicos e linhas de base interpretáveis

Os primeiros trabalhos nessa área privilegiaram modelos interpretáveis e dados disponíveis nos estágios iniciais da trajetória acadêmica, como informações de matrícula, perfil educacional prévio e desempenho nos primeiros períodos. Nesse contexto, [Dekker, Pechenizkiy e Vleeshouwers \(2009\)](#) conduzem um dos estudos fundacionais da área ao investigar a predição de evasão em um curso de Engenharia Elétrica, utilizando árvores de decisão como principal modelo. Os resultados indicam acurácia entre 75% e 80%, associada a uma lógica de uso institucional explícita, que inclui aprendizado custo-sensível e análise sistemática de erros de classificação. Os autores argumentam que a utilidade do modelo não se restringe à acurácia global, mas também à sua capacidade de reduzir falhas críticas na identificação de estudantes em risco, o que tem implicações diretas para intervenções precoces.

Em uma perspectiva complementar, [\(KOVAČIĆ, 2010\)](#) exploram a predição de sucesso acadêmico a partir de dados de matrícula em um curso de educação a distância.

Utilizando árvores de decisão e variáveis predominantemente demográficas e administrativas, o estudo alcança desempenho mais modesto, com acurácia entre 59% e 61%. Esses resultados sugerem que informações de perfil inicial, embora relevantes, são insuficientes para capturar a complexidade do fenômeno da evasão, reforçando a necessidade de incorporar dados que representem a trajetória acadêmica ao longo do curso.

No contexto brasileiro, [Manhães et al. \(2011\)](#) analisam a evasão em um curso de Engenharia Civil a partir de dados acadêmicos do primeiro semestre, incluindo notas em disciplinas iniciais e coeficiente de rendimento. Os autores reportam acurácia próxima de 76% a 80%, evidenciando o forte poder discriminativo do desempenho acadêmico inicial. Contudo, o estudo também destaca a existência de padrões contraintuitivos, como estudantes com bom desempenho que evadem e estudantes com baixo desempenho que concluem o curso. Esse achado indica que variáveis acadêmicas, embora centrais, não esgotam a explicação do fenômeno, apontando para a necessidade de análises mais contextualizadas.

3.1.2 Modelos mais complexos e limitações de transparência

Com a ampliação das bases institucionais e o avanço das técnicas de mineração de dados, estudos posteriores passaram a empregar modelos mais complexos, buscando ganhos em desempenho preditivo. Nesse sentido, [Delen \(2011\)](#) comparam regressão logística, árvores de decisão e redes neurais artificiais na predição de evasão de estudantes ingressantes, utilizando uma base longitudinal com mais de 25 mil registros. Os resultados indicam que redes neurais alcançam maior acurácia global, em torno de 81%, enquanto árvores e regressão logística apresentam desempenho inferior.

Apesar desse ganho preditivo, o autor ressalta que modelos mais complexos tendem a apresentar menor transparência, o que pode dificultar sua adoção em ambientes educacionais. Nesse contexto, árvores de decisão são apontadas como uma alternativa mais equilibrada entre desempenho e interpretabilidade, especialmente quando o objetivo é apoiar gestores acadêmicos. Além disso, o estudo evidencia que variáveis educacionais e financeiras são determinantes para a evasão, reforçando achados de trabalhos anteriores.

3.2 XAI aplicada à evasão acadêmica

Nos últimos anos, um número crescente de estudos tem investigado o uso de métodos de Inteligência Artificial Explicável na predição de evasão acadêmica, buscando superar as limitações associadas a modelos preditivos opacos. Esses trabalhos partem da premissa de que a explicabilidade pode ampliar a utilidade prática das previsões ao tornar mais clara a contribuição de variáveis acadêmicas, demográficas ou comportamentais para o risco estimado de evasão.

Entre as abordagens mais recorrentes, destacam-se estudos que aplicam métodos como *SHAP* e *LIME* para interpretar modelos preditivos de evasão. [Melo et al. \(2022\)](#) apresentam um pipeline que combina redes neurais artificiais com métodos de explicabilidade, incluindo *SHAP* e *LIME*, em um estudo de caso com dados educacionais brasileiros. Os autores exploram explicações locais e globais e discutem a relevância de erros críticos, especialmente falsos negativos. Os resultados indicam que a explicabilidade pode auxiliar na compreensão das decisões do modelo, mas também evidenciam desafios relacionados ao desbalanceamento dos dados e à interpretação de métricas agregadas em cenários com forte assimetria entre classes.

Em uma perspectiva semelhante, [Krueger, Jr. e Barddal \(2023\)](#) investigam a aplicação de aprendizado de máquina explicável para predição de evasão em instituições educacionais brasileiras. Os autores argumentam que a utilidade das predições depende da capacidade de compreender os fatores associados ao risco estimado, permitindo que equipes educacionais interpretem os resultados e planejem intervenções preventivas. Utilizando modelos como *Random Forest*, *AdaBoost* e *XGBoost*, o estudo emprega métodos de explicabilidade como *SHAP* e *LIME* para analisar a contribuição de diferentes variáveis. Os resultados indicam que fatores associados ao desempenho acadêmico e às condições socioeconômicas apresentam forte influência nas estimativas de evasão, reforçando o papel da explicabilidade na identificação de padrões relevantes para a gestão educacional.

Em linha semelhante, mas com foco mais direto na tomada de decisão, [Nagy e Molontay \(2024\)](#) propõem uma abordagem de predição de evasão orientada à intervenção personalizada. O estudo parte da ideia de que explicações interpretáveis devem servir como base para ações concretas de acompanhamento acadêmico, aproximando o processo de predição das estratégias institucionais de intervenção. Ao enfatizar a necessidade de explicações acionáveis, os autores argumentam que a utilidade da XAI não está apenas na transparência técnica dos modelos, mas também na sua capacidade de apoiar decisões pedagógicas individualizadas.

Mais recentemente, [Bettahi, Belouadha e Harroudi \(2025\)](#) propõem um pipeline modular de aprendizado de máquina para predição de evasão no ensino superior que integra uma camada de explicabilidade baseada em *SHAP*. Diferentemente de abordagens centradas apenas na importância global de variáveis, o sistema permite examinar contribuições individuais e distinguir fatores associados ao aumento ou à redução do risco de evasão. O estudo também incorpora mecanismos de auditoria das predições, identificando casos ambíguos ou próximos ao limite de decisão para revisão humana. Os resultados indicam alto desempenho preditivo, com destaque para o modelo *XGBoost*, e ressaltam a importância de integrar explicabilidade e governança analítica em sistemas de alerta precoce para retenção estudantil.

Estudos também exploram a comparação entre métodos de explicabilidade aplicados

à evasão acadêmica. (SILVA et al., 2024) conduzem uma análise comparativa de técnicas de XAI em um estudo de caso educacional, discutindo diferenças entre modelos e métodos explicativos. Embora o trabalho evidencie o crescente interesse pela explicabilidade no cenário brasileiro, sua análise permanece predominantemente agregada, com menor ênfase na tradução das explicações em subsídios operacionais para gestores e docentes. Esse tipo de limitação é recorrente na literatura e tende a reduzir o potencial impacto prático das abordagens propostas.

Apesar desses avanços, estudos recentes têm problematizado a suficiência das explicações tradicionais para apoiar decisões educacionais complexas. Saqr e López-Pernas (2024) argumentam que a explicabilidade pode ser insuficiente quando não se considera o comportamento do modelo em casos de erro. Ao analisar predições corretas e incorretas em conjunto, os autores mostram que explicações focadas apenas em acertos podem gerar uma falsa sensação de confiabilidade, especialmente nos casos em que intervenções seriam mais necessárias. Esse resultado reforça a importância de analisar *mispredictions* e de utilizar explicações locais e contextualizadas em nível de instância.

3.3 Síntese comparativa e posicionamento do estudo

A Tabela 2 apresenta uma síntese comparativa de estudos empíricos recentes que aplicam métodos de aprendizado de máquina e técnicas de Inteligência Artificial Explicável na predição de evasão acadêmica. A análise conjunta desses trabalhos evidencia que a literatura tem avançado na direção de tornar modelos preditivos mais interpretáveis, buscando compreender não apenas quais estudantes apresentam maior risco de evasão, mas também quais fatores estão associados a essas estimativas.

Os estudos analisados abrangem diferentes contextos educacionais, incluindo ensino técnico, ensino fundamental e médio e ensino superior. Em termos metodológicos, observa-se a predominância de modelos baseados em árvores e métodos de ensemble, frequentemente combinados com técnicas de explicabilidade como SHAP e LIME. Essas abordagens são utilizadas tanto para identificar fatores associados à evasão quanto para interpretar decisões do modelo em nível global ou Local. Parte da literatura também tem explorado aplicações mais específicas da explicabilidade, como a comparação entre métodos explicativos, a análise de erros de predição e o suporte à definição de intervenções educacionais personalizadas.

Apesar desses avanços, a análise comparativa revela algumas limitações recorrentes. Em vários estudos, as análises permanecem concentradas em resultados agregados, com menor atenção às diferenças entre cursos ou perfis acadêmicos específicos. Além disso, embora as técnicas de explicabilidade permitam identificar fatores associados ao risco, nem sempre as interpretações são desenvolvidas de forma a apoiar diretamente a análise institucional ou o planejamento de estratégias de acompanhamento estudantil.

Diante desse cenário, este trabalho propõe uma abordagem analítica orientada à aplicação institucional, estruturada em três eixos complementares. O primeiro consiste na modelagem segmentada por curso, permitindo considerar explicitamente diferenças entre perfis acadêmicos e padrões de evasão. O segundo eixo refere-se à integração de explicabilidade global e local em um mesmo fluxo analítico, possibilitando identificar fatores relevantes para o planejamento institucional e, simultaneamente, interpretar o risco ao nível individual. Por fim, o terceiro eixo enfatiza uma interpretação contextual das explicações, orientada às necessidades de gestores e docentes. Nessa perspectiva, as explicações podem ser utilizadas como subsídios para a compreensão do risco e para o planejamento de ações de acompanhamento acadêmico, ampliando o potencial de uso institucional das análises produzidas.

Tabela 2 – Síntese comparativa de estudos empíricos sobre predição de evasão acadêmica com métodos de XAI

Estudo	Contexto educacional	Modelos utilizados	Métodos XAI	Nível	Foco da análise
Melo et al. (2022)	Ensino técnico (IFRN, Brasil)	MLP	LIME, SHAP	Global, local	Interpretação de fatores associados à evasão
Krueger, Jr. e Barddal (2023)	Ensino fundamental e médio (19 escolas privadas brasileiras)	Decision Tree, Logistic Regression, AdaBoost, XGBoost	SHAP, LIME	Global, local	Identificação de fatores acadêmicos e socioeconômicos
Nagy e Molontay (2024)	Ensino superior (BME, Hungria)	CatBoost, XGBoost, NGBoost, AdaBoost	PI, SHAP, PDP	Global, local	Suporte a intervenções educacionais personalizadas
Saqr e López-Pernas (2024)	Ensino superior (medicina, Finlândia)	Random Forest, XGBoost, SVM, MLP	SHAP	Local	Análise de erros de predição (mispredictions)
Silva et al. (2024)	Ensino técnico (IFMA, Brasil)	Random Forest, Decision Tree, SVM	SHAP, LIME, Anchor	Global	Comparação entre métodos de explicabilidade
Bettahi, Belouadha e Harroud (2025)	Ensino superior (Marrocos)	Random Forest, Naïve Bayes, MLP, XGBoost	SHAP	Global, local	Pipeline explicável para governança e auditoria
Este estudo	Cursos de graduação da UFMA	XGBoost, Random Forest, Logistic Regression, MLP	SHAP, LIME, Anchor, PI	Global, local	Modelagem explicável segmentada por curso para apoio à gestão acadêmica

4 Metodologia

Este capítulo descreve a metodologia adotada neste estudo, apresentando as etapas de processamento e análise utilizadas para a construção e interpretação dos modelos preditivos de evasão acadêmica. O processo metodológico foi estruturado a partir de uma adaptação do *CRISP-DM* (*Cross Industry Standard Process for Data Mining*), amplamente utilizado em pesquisas aplicadas em mineração de dados educacionais (SCHRÖER; KRUSE; GÓMEZ, 2021). Essa adaptação organiza o fluxo analítico desde a preparação dos dados institucionais até a interpretação das previsões geradas pelos modelos.

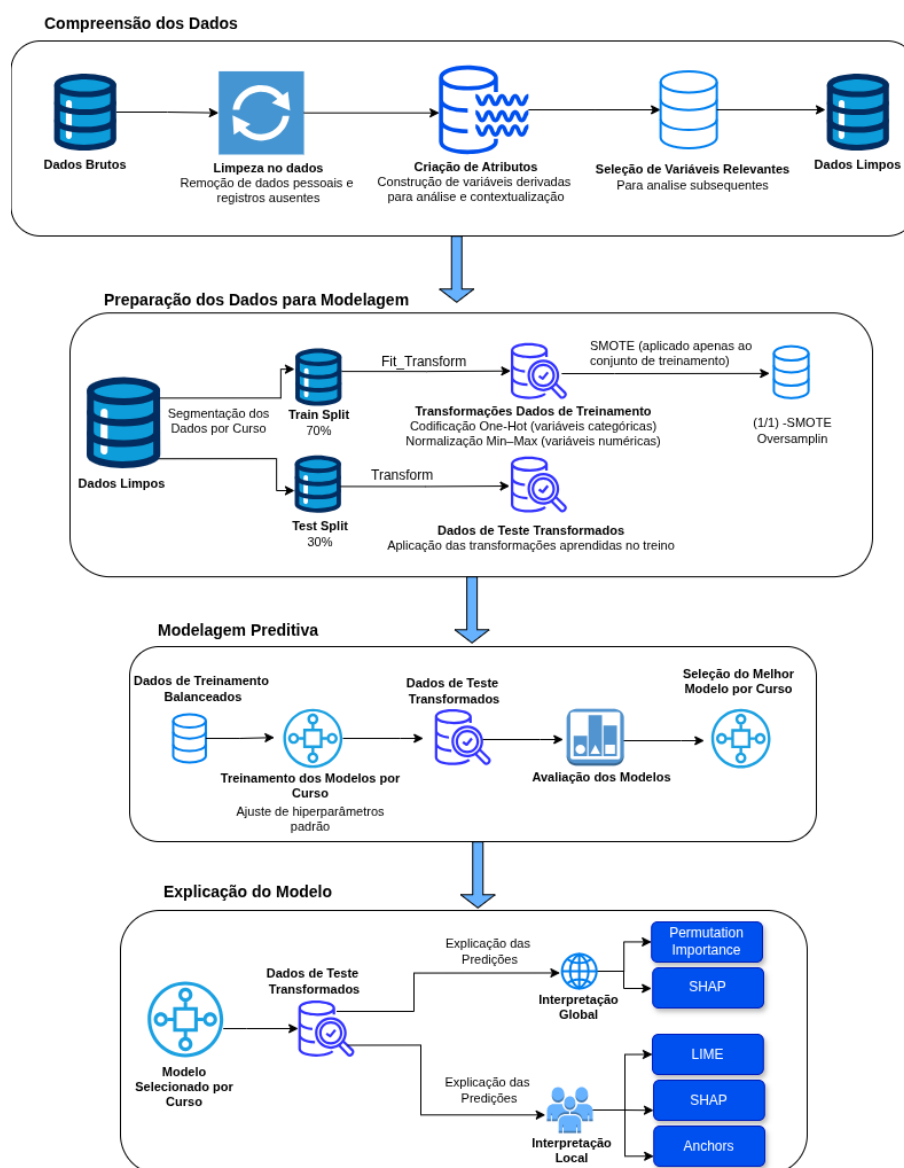
4.1 Visão Geral da Metodologia e Caracterização do Estudo

Este estudo caracteriza-se como uma pesquisa aplicada, de natureza quantitativa e abordagem empírica, baseada na análise de dados institucionais históricos de cursos de graduação da Universidade Federal do Maranhão. O objetivo é investigar padrões associados ao risco de evasão acadêmica por meio da construção e interpretação de modelos preditivos, buscando produzir evidências que possam contribuir para o acompanhamento acadêmico e para a formulação de estratégias institucionais de permanência estudantil. Neste trabalho, a evasão acadêmica é definida como o desligamento formal do estudante antes da conclusão do curso, conforme registrado nos sistemas acadêmicos institucionais.

A metodologia foi organizada de forma modular, com o objetivo de facilitar sua replicação em outros contextos institucionais. O fluxo analítico abrange desde a preparação inicial dos dados até a interpretação das previsões geradas pelos modelos e foi estruturado em quatro etapas principais: (i) *Compreensão dos dados*, que envolve a verificação de integridade dos registros, remoção de inconsistências, criação de atributos e análise exploratória; (ii) *Preparação dos dados*, que contempla procedimentos de transformação, codificação, normalização e balanceamento das classes; (iii) *Modelagem preditiva*, conduzida de forma independente para cada curso analisado; e (iv) *Explicabilidade dos modelos*, na qual métodos da XAI incluindo *SHAP*, *Permutation Importance (PI)*, *LIME* e *Anchors* são utilizados para interpretar o comportamento dos modelos e analisar suas previsões em níveis global e local.

A Figura 1 apresenta uma visão geral desse processo metodológico, ilustrando o encadeamento das etapas de preparação dos dados, modelagem preditiva e análise explicável dos resultados.

Figura 1 – Fluxograma da metodologia proposta.



Elaborado pelo autor

4.2 Compreensão dos Dados

A etapa de *Compreensão dos Dados* tem como objetivo caracterizar o conjunto de informações institucionais utilizado neste estudo e assegurar condições mínimas de qualidade para as análises subsequentes. Nessa fase, são delimitados o escopo institucional, temporal e acadêmico da pesquisa, além de serem verificados a integridade e a consistência dos registros disponíveis. Também são realizadas a criação de atributos derivados e a definição do conjunto de variáveis que utilizados tanto na modelagem quanto na interpretação dos resultados.

Os dados analisados reúnem informações acadêmicas de 5.876 estudantes de quatro cursos presenciais da UFMA, Enfermagem (ENF), Odontologia (ODO), Ciência da

Computação (CCO) e Bacharelado Interdisciplinar em Ciência e Tecnologia (BICT). O conjunto contempla estudantes com ingresso entre os anos de 2011 e 2024. A escolha desses cursos buscou representar áreas de formação distintas e diferentes dinâmicas de progressão acadêmica, permitindo analisar o método proposto em contextos educacionais variados.

A descrição das variáveis disponíveis no conjunto de dados original, em seu formato bruto, é apresentada na Tabela 3.

Foram considerados apenas estudantes com situação acadêmica final definida, classificados como *concluídos* ou *desligados*. Neste estudo, a categoria *desligado* foi adotada como representação operacional da evasão acadêmica. Reconhece-se que essa categoria pode incluir diferentes situações de encerramento do vínculo estudantil, como abandono formal, jubramento ou transferência. Ainda assim, diante das informações disponíveis na base institucional, essa definição foi adotada como alternativa metodológica para viabilizar as análises.

Após a extração, os dados passaram por procedimentos de limpeza e verificação de consistência. Inicialmente, foram removidos identificadores pessoais e demais informações sensíveis, garantindo o tratamento ético e confidencial dos registros em conformidade com a Lei Geral de Proteção de Dados (LGPD). Em seguida, foram excluídos registros com valores ausentes em variáveis consideradas essenciais, como ano de ingresso, situação acadêmica final e indicadores de desempenho. Essa exclusão correspondeu a aproximadamente 4% da amostra. Optou-se por esse procedimento, em vez de imputação, porque a ausência nesses campos compromete diretamente a interpretação do desfecho acadêmico e dos indicadores que estruturam o problema investigado.

O conjunto de dados inclui indicadores institucionais de desempenho acadêmico calculados a partir de fórmulas padronizadas adotadas pela universidade, que sintetizam dimensões complementares da trajetória discente. Esses indicadores possuem natureza acumulada, pois são construídos a partir do histórico acadêmico registrado ao longo do curso. Dessa forma, os valores observados refletem a situação acadêmica consolidada do estudante no momento da extração dos dados, sintetizando informações de desempenho, eficiência e progressão acadêmica registradas nos sistemas institucionais.

A *Média de Conclusão (MC)* representa o desempenho global do estudante por meio de uma média ponderada das notas finais obtidas nas disciplinas aprovadas. O *Índice de Eficiência em Carga Horária (IECH)* expressa a proporção da carga horária matriculada que resultou em aprovação, refletindo a eficiência na conversão de créditos cursados em créditos concluídos. Já o *Índice de Eficiência em Períodos Letivos (IEPL)* relaciona a carga horária acumulada à carga horária esperada para o número de semestres cursados, indicando o ritmo de progressão em relação ao tempo ideal do curso. Por fim, o *Índice de Eficiência Acadêmica (IEA)* é definido como o produto entre *MC*, *IECH* e

Tabela 3 – Variáveis utilizadas no estudo

Nome do campo	Tipo
Curso	Categórico
Curriculo	Numérico
Turno	Categórico
Grau_Academico	Categórico
Campus	Categórico
Modalidade	Categórico
Pais	Categórico
Cidade_Estado	Categórico
Sexo	Categórico
Tipo_Raca	Categórico
Identidade_Genero	Categórico
TNE	Categórico
Ano_Ingresso	Numérico
Periodo_Ingresso	Numérico
Forma_Ingresso	Categórico
Ano_Periodo_Saida	Numérico
Tipo_Saida	Categórico
Status	Categórico
Novo_Status	Categórico
Nivel	Categórico
Tipo	Categórico
Média_de_Conclusão	Numérico
Indice_de_Eficiência_em_Carga_Horária	Numérico
Indice_de_Eficiência_em_Periodos_Letivos	Numérico
Prazo_Integralizacao	Numérico
Prazo_Integralizacao_Ano	Data
Prazo_Integralizacao_Semestre	Numérico
Tipo_Rede_Ensino	Categórico
Estado_Civil	Categórico
Data_Nascimento	Data
Passaporte	Categórico
Tipo_Logradouro	Categórico
Logradouro	Categórico
Numero	Categórico
Complemento	Categórico
CEP	Categórico

IEPL, sintetizando em um único indicador aspectos de desempenho, eficiência e progressão acadêmica.

Além das verificações estruturais, esta etapa incluiu uma exploração descritiva do conjunto de dados, voltada à análise das distribuições, padrões gerais e possíveis diferenças entre cursos e grupos de desfecho (concluintes e desligados). O objetivo dessa exploração foi produzir observações iniciais, que orientassem decisões metodológicas posteriores, especialmente na seleção de variáveis e na definição da estratégia de modelagem. Conforme apontam [Figueiredo e Souza \(2012\)](#), essa etapa é essencial para revelar tendências e discrepâncias que orientam tanto a escolha dos métodos analíticos quanto a interpretação dos resultados.

Por fim, foi realizada a seleção do conjunto de variáveis utilizado nas análises subsequentes. Essa seleção foi guiada por dois critérios principais: (i) evidências discutidas na literatura sobre evasão e permanência no ensino superior e (ii) padrões observados na análise exploratória inicial dos dados. Ao final desta etapa, foi definido o conjunto de dados que utilizado nas fases seguintes de preparação dos dados, modelagem preditiva e interpretação dos resultados.

4.3 Preparação dos Dados para Modelagem

Esta seção descreve os procedimentos de preparação dos dados voltados à modelagem preditiva. O objetivo dessa etapa é estruturar os conjuntos de treinamento e teste e definir, de forma padronizada, as transformações aplicadas às variáveis. As escolhas realizadas buscam evitar vazamento de informação entre os conjuntos, garantir uma avaliação do desempenho modelo e assegurar a reprodutibilidade do processo de modelagem. Todas as operações foram conduzidas separadamente para cada curso, respeitando as particularidades identificadas na etapa de compreensão dos dados.

A primeira decisão desta etapa foi a segmentação dos dados por curso. Estudos sobre predição de evasão indicam que cursos distintos apresentam dinâmicas acadêmicas próprias, relacionadas a diferenças curriculares, perfis estudantis e formas de progressão ao longo da graduação ([KHOSRAVI et al., 2022](#); [SACCARO](#); [FRANÇA](#); [JACINTO, 2019](#)). Diante dessa diversidade, optou-se por conduzir a preparação e a modelagem de forma independente para cada curso. Essa estratégia permite que os modelos aprendam padrões específicos de cada contexto acadêmico e evita que características predominantes de cursos com maior número de estudantes influenciem o aprendizado em cursos com trajetórias distintas.

Em seguida, foi realizado o particionamento dos dados em conjuntos de treinamento e teste, utilizando a proporção de 70% para treinamento e 30% para teste. A divisão foi realizada de forma estratificada, preservando em cada curso, a proporção entre estudantes

concluídos e *desligados*. Esse procedimento contribui para reduzir distorções na avaliação fora da amostra e assegura que o conjunto de teste represente adequadamente a distribuição observada do desfecho em cada curso.

As transformações dos dados foram implementadas por meio de *pipelines* da biblioteca *Scikit-learn*. Todas as operações foram ajustadas exclusivamente no conjunto de treinamento e posteriormente aplicadas ao conjunto de teste. Esse procedimento evita que informações do teste influenciem o ajuste das transformações e organiza as etapas de processamento em um fluxo reprodutível (GÉRON, 2021). As variáveis categóricas foram codificadas por meio de *One-Hot Encoding (OHE)*, mantendo-se no teste as mesmas categorias observadas no treinamento, de modo a evitar a introdução de ordens artificiais em atributos nominais (HARRISON, 2020). As variáveis numéricas foram normalizadas utilizando *Min-Max Scaling*, procedimento que harmoniza as escalas dos atributos e reduz o impacto de diferenças de magnitude no ajuste dos modelos (GÉRON, 2021).

Por fim, foi tratado o desbalanceamento da variável-alvo (*Status*), observado em todos os cursos. Para reduzir o favorecimento da classe majoritária durante o treinamento, aplicou-se o *SMOTE (Synthetic Minority Over-sampling Technique)* apenas no conjunto de treinamento, após a realização das transformações. Essa estratégia permite apoiar o aprendizado de padrões associados à classe minoritária sem alterar a distribuição original do conjunto de teste, preservando a validade da avaliação fora da amostra (DANTAS; ARANHA; SILVA, 2024).

Ao final desta etapa, obtêm-se, para cada curso, os conjuntos de treinamento e teste preparados, acompanhados de um *pipeline* de transformações ajustado no treinamento e de um conjunto de treinamento balanceado. Esses artefatos servem de base para a etapa de modelagem preditiva apresentada na seção seguinte.

4.4 Modelagem Preditiva

Esta seção apresenta a etapa de modelagem preditiva. A análise foi conduzida separadamente para cada curso, de modo a preservar particularidades das trajetórias acadêmicas e dos padrões de desligamento observados em ENF, ODO, CCO e BICT. Com isso, busca-se produzir estimativas de risco mais aderentes ao contexto de cada curso, sem perder a comparabilidade entre os experimentos realizados.

Foram comparados três algoritmos amplamente utilizados em estudos de predição em contextos educacionais, especialmente em investigações sobre evasão e sucesso acadêmico: *XGBoost*, *Random Forest* e *Multilayer Perceptron (MLP)* (SAARELA; PODGORELEC, 2024). Esses métodos têm sido frequentemente empregados nesse tipo de problema por sua capacidade de capturar relações não lineares e interações entre variáveis presentes em dados educacionais. Como referência de base interpretável, foi incluída também a *Regressão*

Logística, por oferecer uma estrutura explicativa mais direta, baseada em coeficientes (MARQUES et al., 2023).

Para preservar a comparabilidade entre cursos e algoritmos, optou-se por não realizar ajuste fino de hiperparâmetros nem otimização dos limiares de decisão. Todos os modelos foram ajustados com as configurações padrão das bibliotecas utilizadas e com limiar de classificação fixado em 0,5. Essa decisão reduz a influência de parametrizações específicas sobre a comparação entre modelos e favorece a replicação do procedimento em outros contextos institucionais. Em aplicações práticas, no entanto, o limiar de decisão pode ser ajustado de acordo com os objetivos institucionais de acompanhamento acadêmico, permitindo priorizar maior sensibilidade na identificação de estudantes em risco.

O desempenho foi avaliado por meio da Área Sob a Curva Precisão–Revocação (*AUC-PR*), métrica adequada a cenários com classes desbalanceadas, por refletir a relação entre precisão e revocação (SAITO; REHMSMEIER, 2015). Para examinar a estabilidade das comparações, foram calculados intervalos de confiança de 95% por reamostragem *bootstrap* com 1.000 iterações, utilizando as previsões obtidas no conjunto de teste. Em cada curso, o modelo final foi definido a partir da maior média de *AUC-PR*, considerando também os respectivos intervalos de confiança.

Após a avaliação dos modelos no conjunto de teste, as probabilidades preditas de desligamento foram utilizadas para organizar os estudantes em faixas de risco, com o objetivo de apoiar a análise de padrões de desempenho acadêmico associados ao desligamento, além de ampliar a aplicabilidade prática das previsões.

Essas faixas não foram tratadas como novas classes preditivas, mas como um recurso analítico para comparar perfis acadêmicos com base nos indicadores *MC*, *IECH* e *IEPL*. Além disso, elas serviram como referência para orientar as análises de explicabilidade apresentadas na seção seguinte.

4.5 Explicabilidade dos Modelos

A etapa de explicabilidade teve como objetivo interpretar as estimativas de risco de desligamento produzidas pelos modelos. Nesta dissertação, a explicabilidade é tratada como uma etapa posterior à modelagem, voltada a examinar como os modelos estruturam suas decisões tanto no nível geral quanto em previsões individuais. As análises foram conduzidas reconhecendo que as explicações refletem padrões aprendidos a partir dos dados disponíveis e, portanto, indicam associações estatísticas, sem implicar relações causais sobre o fenômeno do desligamento acadêmico.

As análises explicativas foram organizadas em duas perspectivas complementares, explicabilidade global e explicabilidade local. Em ambas, as explicações foram examinadas

considerando também as faixas de risco definidas na Seção 4.4, permitindo investigar como diferentes perfis acadêmicos se relacionam com os padrões de decisão aprendidos pelos modelos.

A explicabilidade global teve como objetivo identificar padrões recorrentes de decisão, indicando quais variáveis tendem a influenciar o risco estimado de forma sistemática em cada curso e em cada faixa de risco. Para esse propósito, foram utilizados os métodos de *Permutation Importance (PI)* e *SHAP*. A PI foi empregada para avaliar a dependência do desempenho do modelo em relação a cada atributo, a partir da variação da métrica de avaliação quando seus valores são permutados. O SHAP, por sua vez, permitiu decompor a contribuição média das variáveis para as previsões, oferecendo uma medida interpretável de importância relativa agregada. A análise conjunta desses métodos possibilitou comparar a relevância dos atributos entre cursos, evidenciando fatores associados ao aumento ou à redução do risco estimado de desligamento.

A explicabilidade local foi utilizada para interpretar previsões individuais, aspecto central em aplicações educacionais nas quais o uso das estimativas de risco depende da compreensão de casos concretos e da justificativa de decisões de acompanhamento. Nessa perspectiva, aplicaram-se os métodos *SHAP*, *LIME* e *Anchors*. O SHAP foi utilizado para decompor a probabilidade estimada de um estudante em contribuições positivas e negativas das variáveis, evidenciando fatores associados ao aumento ou à redução do risco predito. O LIME foi empregado para aproximar localmente o comportamento do modelo por meio de um modelo substituto simples, destacando variáveis influentes na vizinhança do exemplo analisado. O Anchors permitiu extrair regras condicionais interpretáveis que sustentam decisões do modelo em casos específicos. Essas explicações locais foram analisadas em articulação com as faixas de risco, permitindo examinar se diferentes perfis acadêmicos apresentam justificativas preditivas distintas.

5 Resultados e Discussão

Este capítulo apresenta e discute os resultados obtidos a partir da metodologia descrita no Capítulo 4, com foco na avaliação do desempenho preditivo dos modelos e na interpretação das estimativas de risco de desligamento acadêmico.

Inicialmente, são apresentados resultados exploratórios que caracterizam o conjunto de dados, incluindo a distribuição das variáveis demográficas, dos indicadores acadêmicos e dos padrões observados entre cursos e situações acadêmicas. Essa análise fornece o contexto necessário para interpretar as etapas posteriores do estudo. Em seguida, examina-se o desempenho preditivo dos modelos ajustados para cada curso, considerando métricas apropriadas para cenários com classes desbalanceadas. Na sequência, são analisadas as explicações produzidas pelos modelos por meio de métodos de explicabilidade global e local, incluindo a investigação de acertos e erros preditivos.

Ao longo do capítulo, os resultados são apresentados e interpretados de forma articulada, sendo relacionados, quando pertinente, à literatura da área, de modo a evidenciar convergências, diferenças e contribuições do estudo para o uso de modelos explicáveis no monitoramento acadêmico.

5.1 Caracterização exploratória dos dados e padrões iniciais

5.1.1 Distribuição dos estudantes por curso e situação acadêmica

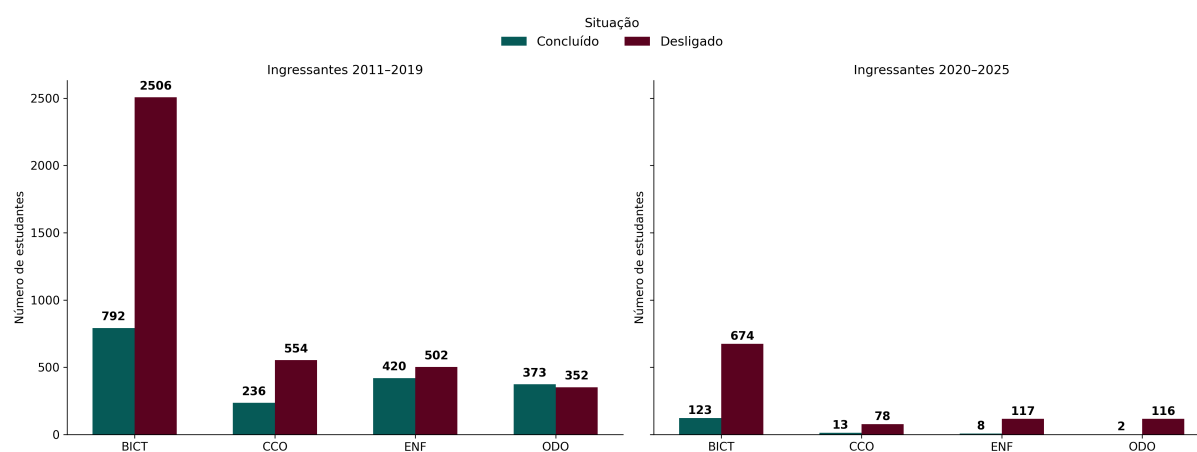
A Figura 2 apresenta a distribuição dos estudantes por curso segundo a situação acadêmica final, classificada como *concluído* ou *desligado*. Para evitar interpretações influenciadas pelo tempo necessário à integralização, a visualização é apresentada em dois recortes temporais: ingressantes de 2011–2019 e ingressantes de 2020–2025. Essa separação permite observar o comportamento das coortes com maior tempo de permanência no curso e, ao mesmo tempo, identificar padrões iniciais entre os ingressantes mais recentes.

No primeiro intervalo (2011–2019), composto majoritariamente por estudantes que já tiveram tempo suficiente para concluir o curso, observa-se maior volume absoluto de desligamentos no BICT e no CCO. No BICT, que também concentra o maior número de estudantes, os desligamentos superam amplamente os concluintes. No CCO, o mesmo padrão aparece em menor escala. Em ENF, a diferença entre desligados e concluintes é mais moderada, enquanto em ODO a distribuição se mostra mais equilibrada, com leve predominância de concluintes. Esses resultados indicam que os cursos apresentam padrões distintos de trajetória acadêmica e permanência estudantil.

No segundo intervalo (2020–2025), os quantitativos são menores e o número de concluintes ainda é reduzido. Esse resultado não deve ser interpretado como evidência direta de piora no desempenho institucional, pois muitas dessas turmas ainda estão em andamento e parte dos estudantes não teve tempo suficiente para completar o ciclo de integralização. Ainda assim, a presença de desligamentos já registrados nesse recorte sugere que parte dos ingressantes recentes interrompe a trajetória acadêmica ainda nos primeiros períodos do curso.

Esse padrão é consistente com evidências reportadas na literatura nacional, como em Saccaro, França e Jacinto (2019), e com levantamentos do Censo da Educação Superior (Inep, 2025), que apontam diferenças nas taxas de evasão entre áreas de formação e perfis institucionais. Para fins de análise e modelagem preditiva, essa heterogeneidade reforça a importância de conduzir as análises de forma segmentada por curso, permitindo capturar padrões específicos de trajetória acadêmica em cada contexto formativo.

Figura 2 – Situação final dos estudantes por curso (2011–2025)



Elaborada pelos autores

5.1.2 Distribuição das variáveis demográficas

A Tabela 4 apresenta a distribuição das principais variáveis demográficas por curso e situação acadêmica final, incluindo sexo, raça, rede de ensino e forma de ingresso. As células mostram o número absoluto de estudantes e a porcentagem correspondente dentro de cada *status* do curso, permitindo comparar a composição demográfica entre concluintes e desligados.

De modo geral, observa-se que a distribuição das categorias demográficas é semelhante entre os dois desfechos. Nos cursos em que determinado grupo é mais frequente na população estudantil, os desligamentos também tendem a se concentrar nesse mesmo grupo. Esse padrão pode ser observado na variável sexo: em BICT e CCO, onde há predominância masculina, a maior parte dos desligamentos ocorre entre homens, enquanto em ENF e

ODO, cursos com maior presença feminina, os desligamentos aparecem majoritariamente entre mulheres. Tendência semelhante é observada nas demais variáveis demográficas, como raça, rede de ensino e forma de ingresso.

Esse comportamento reflete principalmente a composição da população estudantil em cada curso. Como essas categorias representam a maior parcela dos estudantes, é natural que concentrem também maior número absoluto de desligamentos. Embora existam algumas diferenças proporcionais entre categorias, essas variações ocorrem, em geral, em grupos com menor frequência absoluta, o que limita sua utilidade como critério isolado para distinguir concluintes e desligados.

Em conjunto, os resultados indicam que as variáveis demográficas, quando analisadas isoladamente, apresentam capacidade limitada para diferenciar os desfechos acadêmicos observados. As diferenças identificadas tendem a refletir sobretudo a própria composição da população estudantil em cada curso, e não necessariamente padrões claros de associação com o desligamento. Isso não significa que fatores demográficos sejam irrelevantes no fenômeno da evasão, mas sugere que, no conjunto de dados analisado, tais características não se destacam como elementos suficientes para distinguir, por si só, as trajetórias de conclusão e desligamento.

Tabela 4 – Distribuição das variáveis demográficas por curso e situação final

Variável	Categoria	BICT		CCO		ENF		ODO	
		Conc.	Desl.	Conc.	Desl.	Conc.	Desl.	Conc.	Desl.
Sexo	Feminino	295 (32.2%)	741 (23.3%)	34 (13.7%)	63 (10.0%)	350 (81.8%)	448 (72.4%)	244 (65.1%)	238 (50.9%)
	Masculino	620 (67.8%)	2439 (76.7%)	215 (86.3%)	569 (90.0%)	78 (18.2%)	171 (27.6%)	131 (34.9%)	230 (49.1%)
Raça	Pardo	437 (47.8%)	1320 (41.5%)	110 (44.2%)	243 (38.4%)	192 (44.9%)	234 (37.8%)	179 (47.7%)	190 (40.6%)
	Branco	293 (32.0%)	854 (26.9%)	91 (36.5%)	182 (28.8%)	137 (32.0%)	174 (28.1%)	121 (32.3%)	150 (32.1%)
	Preto	153 (16.7%)	616 (19.4%)	39 (15.7%)	121 (19.1%)	86 (20.1%)	100 (16.2%)	51 (13.6%)	60 (12.8%)
	Não informado	0 (0.0%)	207 (6.5%)	1 (0.4%)	55 (8.7%)	0 (0.0%)	63 (10.2%)	0 (0.0%)	46 (9.8%)
	Amarelo	31 (3.4%)	159 (5.0%)	7 (2.8%)	21 (3.3%)	11 (2.6%)	27 (4.4%)	15 (4.0%)	15 (3.2%)
	Indígena	1 (0.1%)	24 (0.8%)	1 (0.4%)	10 (1.6%)	2 (0.5%)	21 (3.4%)	9 (2.4%)	7 (1.5%)
Rede de ensino	Pública	538 (58.8%)	1818 (57.2%)	144 (57.8%)	363 (57.4%)	239 (55.8%)	292 (47.2%)	182 (48.5%)	199 (42.5%)
	Privada	377 (41.2%)	1152 (36.2%)	104 (41.8%)	238 (37.7%)	189 (44.2%)	284 (45.9%)	193 (51.5%)	237 (50.6%)
	Não informado	0 (0.0%)	210 (6.6%)	1 (0.4%)	31 (4.9%)	0 (0.0%)	43 (6.9%)	0 (0.0%)	32 (6.8%)
**Forma de ingresso	SISU – Ampla Concorrência	467 (51.0%)	1617 (50.8%)	136 (54.6%)	326 (51.6%)	195 (45.6%)	337 (54.4%)	160 (42.7%)	270 (57.7%)
	SISU – Esc. Pública – Preto, Indígena ou Pardo	155 (16.9%)	593 (18.6%)	42 (16.9%)	97 (15.3%)	84 (19.6%)	91 (14.7%)	55 (14.7%)	63 (13.5%)
	SISU – Esc. Pública – Preto ou Pardo – Baixa Renda	171 (18.7%)	527 (16.6%)	34 (13.7%)	91 (14.4%)	54 (12.6%)	69 (11.1%)	49 (13.1%)	41 (8.8%)
	SISU – Esc. Pública	45 (4.9%)	154 (4.8%)	22 (8.8%)	46 (7.3%)	17 (4.0%)	46 (7.4%)	18 (4.8%)	37 (7.9%)
	SISU – Esc. Pública – Baixa Renda	46 (5.0%)	164 (5.2%)	8 (3.2%)	21 (3.3%)	27 (6.3%)	20 (3.2%)	13 (3.5%)	17 (3.6%)
	SISU – Esc. Pública – Indígena	4 (0.4%)	33 (1.0%)	3 (1.2%)	18 (2.8%)	9 (2.1%)	15 (2.4%)	15 (4.0%)	12 (2.6%)

Nota. Células exibem n e porcentagem dentro de cada status do respectivo curso. ** Para a variável Forma de ingresso, são apresentadas apenas as seis categorias mais frequentes na base de dados

5.1.3 Variáveis derivadas e caracterização da trajetória acadêmica

Para ampliar a capacidade analítica do conjunto de dados e representar aspectos da trajetória discente não explicitamente disponíveis na base original, foram construídas variáveis derivadas relacionadas ao ingresso, à permanência no curso e à origem territorial

dos estudantes. A partir das informações de data de nascimento e de ingresso foram obtidas as variáveis *Idade_Ingresso* e *Faixa_Etaria*. A variável *Regiao_Origem* foi derivada do município e do estado informados no cadastro discente, enquanto *Tempo_permanencia_periodos* foi calculada a partir dos registros de ingresso e saída. Essas variáveis permitem caracterizar com maior detalhe o perfil de entrada e o percurso acadêmico dos estudantes antes da aplicação dos modelos preditivos.

A variável *Idade_Ingresso* e sua categorização em *Faixa_Etaria*, definida conforme a segmentação adotada pelo Censo da Educação Superior do INEP, permitem examinar como o perfil etário de entrada se relaciona com os desfechos acadêmicos. A Tabela 5 mostra que a maior parte dos ingressantes se concentra nas faixas etárias mais jovens (até 22 anos), padrão compatível com o ingresso logo após a conclusão do ensino médio. Ao comparar os desfechos, observa-se que, de modo geral, as faixas etárias iniciais apresentam maior proporção de conclusão, enquanto faixas acima de 23 anos concentram percentuais mais elevados de desligamento. Esse padrão é mais evidente nos cursos de BICT e CCO. Em ENF e ODO, embora as faixas iniciais também concentrem a maior parte dos estudantes, os percentuais de desligamento tendem a aumentar nas faixas etárias mais altas.

Tabela 5 – Distribuição da variável faixa etária por curso e situação final

	BICT			CCO			ENF			ODO		
	Total	Conc. (%)	Desl. (%)	Total	Conc. (%)	Desl. (%)	Total	Conc. (%)	Desl. (%)	Total	Conc. (%)	Desl. (%)
Até 19	2351	27.4	72.6	545	36.0	64.0	616	42.9	57.1	458	47.2	52.8
20 a 22	657	23.9	76.1	149	22.1	77.9	262	47.3	52.7	233	44.2	55.8
23 a 24	271	14.0	86.0	49	16.3	83.7	51	29.4	70.6	58	41.4	58.6
25 a 29	438	10.0	90.0	79	10.1	89.9	63	27.0	73.0	60	48.3	51.7
30 a 39	296	7.4	92.6	50	8.0	92.0	46	15.2	84.8	32	9.4	90.6
40 a 49	70	10.0	90.0	8	0.0	100.0	9	11.1	88.9	2	0.0	100.0
Maior de 50	12	16.7	83.3	1	0.0	100.0	-	-	-	-	-	-

Nota. Valores expressos em porcentagem dentro de cada faixa etária e curso; a coluna Total indica o número de estudantes em cada grupo.

A variável *Regiao_Origem* acrescenta uma dimensão territorial à caracterização da trajetória acadêmica. Conforme apresentado na Tabela 6, a maior parte dos estudantes ingressa a partir da capital, seguida por alunos provenientes de outros municípios do estado. A participação de estudantes oriundos de outros estados é reduzida em todos os cursos. Ao comparar os desfechos, observa-se que os desligamentos acompanham essa mesma composição, concentrando-se nas categorias mais frequentes da população estudantil. Esse resultado indica que a região de origem, considerada isoladamente, possui capacidade limitada para diferenciar concluintes e desligados, embora contribua para caracterizar o perfil territorial dos estudantes atendidos pela instituição.

A variável *Tempo_permanencia_periodos* permite examinar em que momento da trajetória acadêmica os desligamentos tendem a ocorrer. A Figura 3 apresenta a distribuição do tempo de permanência dos alunos desligados em cada curso. No gráfico, a linha central representa a mediana, enquanto as bordas da caixa correspondem ao primeiro e ao terceiro

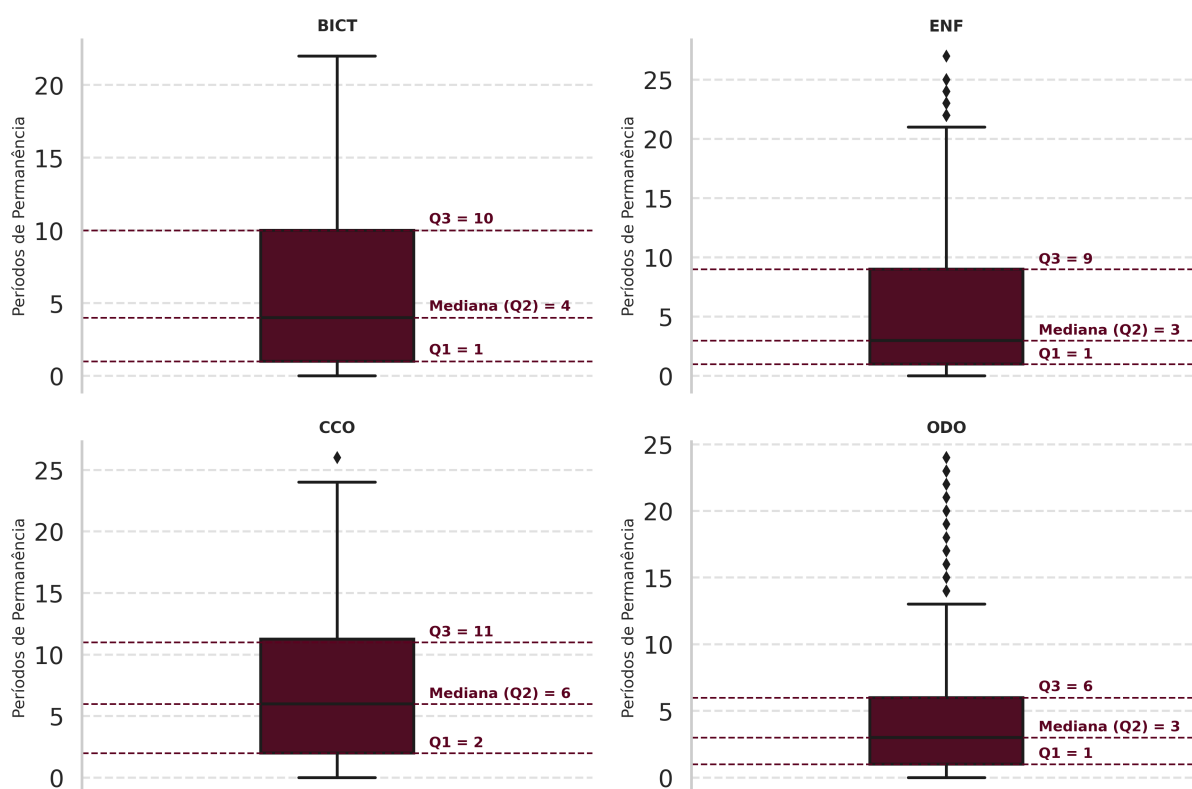
Tabela 6 – Distribuição da variável região de origem por curso e situação final

Variável	Categoria	BICT		CCO		ENF		ODO	
		Conc.	Desl.	Conc.	Desl.	Conc.	Desl.	Conc.	Desl.
Região de origem	Capital	627 (68.5%)	2707 (85.1%)	194 (77.9%)	528 (83.5%)	305 (71.3%)	497 (80.3%)	242 (64.5%)	365 (78.0%)
	Interior	223 (24.4%)	330 (10.4%)	39 (15.7%)	76 (12.0%)	88 (20.6%)	86 (13.9%)	93 (24.8%)	73 (15.6%)
	Outros Estados	65 (7.1%)	143 (4.5%)	16 (6.4%)	28 (4.4%)	35 (8.2%)	36 (5.8%)	40 (10.7%)	30 (6.4%)

Nota. Células exibem n e porcentagem dentro de cada status do respectivo curso.

quartil, delimitando o intervalo interquartil da distribuição. Com base nesses valores, foram definidas três faixas de permanência específicas para cada curso: desligamento precoce (permanência $\leq Q1$), desligamento típico (entre $Q1$ e $Q3$) e desligamento tardio (permanência $> Q3$).

Figura 3 – Distribuição do tempo de permanência (em períodos) dos alunos desligados por curso. As linhas pontilhadas indicam os quartis ($Q1$ e $Q3$) e a mediana ($Q2$), utilizados para definição das faixas de desligamento precoce, típico e tardio.



Elaborada pelos autores

A Tabela 7 apresenta a distribuição dessas faixas considerando apenas estudantes desligados. Em todos os cursos, a maior parcela dos desligamentos ocorre no intervalo típico de permanência: 48,7% no BICT (2–10 períodos), 47,2% em CCO (3–11 períodos), 45,3% em ENF (2–9 períodos) e 44,7% em ODO (2–6 períodos). Os desligamentos precoces também representam uma fração expressiva, variando entre 27,6% e 31,8%, indicando

que parte dos estudantes interrompe a trajetória ainda nos primeiros períodos. Por outro lado, desligamentos tardios aparecem em proporções entre 21,8% e 25,1%, evidenciando situações em que o vínculo com o curso se estende por um período mais longo antes da saída.

Tabela 7 – Distribuição do tempo de permanência dos alunos desligados por curso

Curso	Tempo de permanência (períodos)	% Desligados	Qt. Desligado
BICT	Até 1 período	29.5	938
	Entre 2 a 10 períodos	48.7	1550
	Maior que 10 períodos	21.8	692
CCO	Até 2 períodos	27.6	175
	Entre 3 a 11 períodos	47.2	299
	Maior que 11 períodos	25.1	159
ENF	Até 1 período	31.8	197
	Entre 2 a 9 períodos	45.3	281
	Maior que 9 períodos	22.9	142
ODO	Até 1 período	30.6	143
	Entre 2 a 6 períodos	44.7	209
	Maior que 6 períodos	24.8	116

Nota. Percentuais calculados em relação ao total de alunos desligados em cada curso; *Qt. Desligado* indica o número absoluto de desligamentos. As faixas de períodos são definidas por curso.

Essa distribuição mostra que o desligamento não se concentra apenas no início da trajetória acadêmica. Parte dos casos ocorre nos primeiros períodos, mas uma parcela semelhante aparece ao longo do intervalo típico de permanência e também nos estágios mais avançados do curso. Esse padrão sugere que as dificuldades acadêmicas e institucionais associadas ao desligamento podem surgir em diferentes momentos da formação, o que reforça a importância de acompanhar a trajetória discente ao longo de todo o percurso universitário. Estudos na área de Learning Analytics e predição de evasão também apontam para a importância do acompanhamento contínuo e da identificação precoce de sinais de dificuldade ao longo da trajetória (DEKKER; PECHENIZKIY; VLEESHOUWERS, 2009; NIEROTKA; SALATA; MARTINS, 2023).

5.1.4 Associações entre indicadores de desempenho e implicações para a modelagem

O conjunto de dados inclui indicadores institucionais que descrevem diferentes dimensões do desempenho acadêmico: desempenho global (*MC*), eficiência no aproveitamento da carga horária (*IECH*), progressão em relação ao tempo esperado do curso (*IEPL*) e um indicador composto de eficiência acadêmica (*IEA*), conforme apresentado na metodologia. Nesta subseção, esses indicadores são analisados a partir de duas perspectivas: diferenças descritivas entre concluintes e desligados e associações estatísticas entre os próprios indicadores no grupo de estudantes desligados.

A Tabela 8 apresenta as estatísticas descritivas desses indicadores por curso e situação acadêmica. Em todos os cursos, estudantes concluintes apresentam médias e medianas mais elevadas nos quatro indicadores, além de menor dispersão. Esse padrão reflete trajetórias acadêmicas mais regulares, com maior aproveitamento da carga horária e progressão mais consistente ao longo do curso. Entre os desligados, os valores centrais são mais baixos e a variabilidade tende a ser maior, especialmente em *MC* e *IEA*, indicando maior diversidade de desempenho nesse grupo.

Esse comportamento sugere que o grupo de estudantes desligados não é homogêneo. Parte desses estudantes apresenta níveis intermediários de desempenho, enquanto outros concentram valores muito baixos nos indicadores analisados. Esse padrão é compatível com resultados reportados na literatura, que mostram que o desligamento pode ocorrer em diferentes contextos acadêmicos e não se restringe apenas a situações de baixo rendimento extremo (ORTAGUS et al., 2021; SILVA et al., 2022).

Tabela 8 – Estatísticas descritivas dos indicadores de desempenho por curso e situação acadêmica

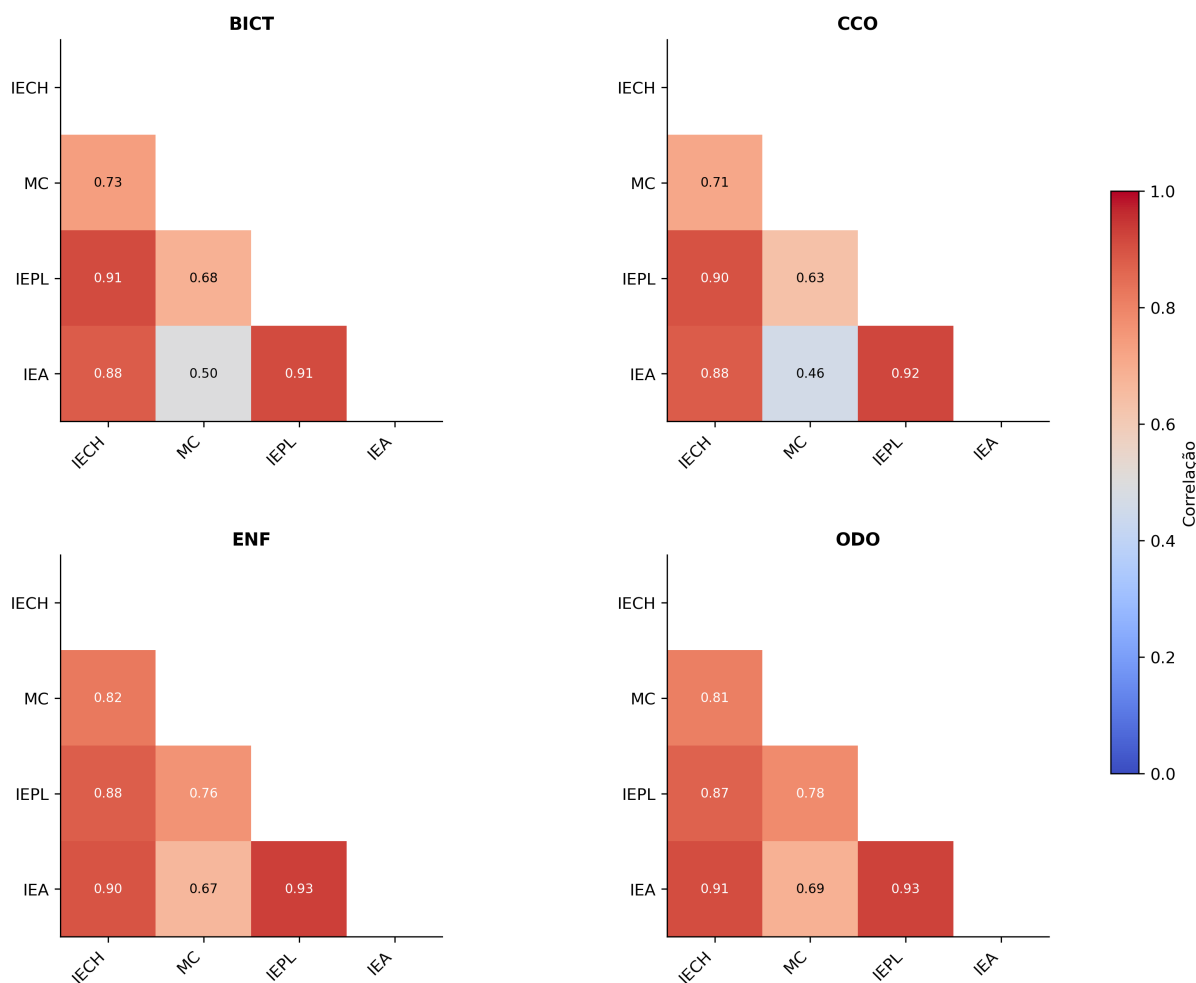
Curso	Status	n	MC			IECH			IEPL			IEA		
			M	Md	DP	M	Md	DP	M	Md	DP	M	Md	DP
BICT	Concluído	915	8.06	8.04	0.35	0.72	0.72	0.14	0.67	0.67	0.18	4.08	3.85	1.70
	Desligado	3180	4.60	7.38	3.88	0.18	0.11	0.21	0.16	0.09	0.20	0.55	0.07	0.95
CCO	Concluído	249	8.40	8.37	0.35	0.84	0.86	0.11	0.77	0.77	0.18	5.57	5.59	1.92
	Desligado	633	5.52	7.59	3.64	0.25	0.23	0.23	0.20	0.15	0.21	0.74	0.28	1.08
ENF	Concluído	428	8.54	8.53	0.31	0.93	0.94	0.06	0.84	0.83	0.13	6.76	6.69	1.41
	Desligado	620	5.21	7.73	3.91	0.38	0.40	0.35	0.33	0.30	0.33	1.88	1.12	2.19
ODO	Concluído	375	8.37	8.37	0.24	0.90	0.90	0.07	0.94	1.00	0.10	7.13	7.40	1.15
	Desligado	468	5.65	8.03	3.95	0.42	0.49	0.37	0.45	0.47	0.41	2.70	1.99	2.80

Nota. n = número de estudantes em cada subgrupo; M = média; Md = mediana; DP = desvio-padrão.

A Figura 4 apresenta as matrizes de correlação de Pearson entre os indicadores no subgrupo de estudantes desligados, separadamente por curso. A análise restrita a esse grupo permite observar como as dimensões de desempenho e eficiência se relacionam em trajetórias que culminaram em desligamento. Em todos os cursos, observa-se correlação elevada entre *IECH* e *IEPL* (superior a 0,87), indicando forte associação entre aproveitamento da carga horária e progressão temporal. O indicador *IEA* também apresenta correlações altas com os demais indicadores, resultado esperado por se tratar de uma medida composta de eficiência acadêmica.

Essas correlações não indicam relações causais, mas sugerem proximidade informacional entre alguns atributos, especialmente entre os índices de eficiência. Em contextos de modelagem preditiva, variáveis fortemente correlacionadas podem compartilhar parte da informação explicativa, o que pode influenciar a interpretação das contribuições individuais das variáveis nos modelos preditivos.

Figura 4 – Matrizes de correlação de Pearson entre os indicadores acadêmicos (MC, IECH, IEPL e IEA) no subgrupo de estudantes desligados, por curso.



Elaborada pelos autores

5.1.5 Síntese integradora e critérios de seleção de variáveis

A análise exploratória apresentada nesta seção permitiu caracterizar o conjunto de dados a partir de diferentes dimensões da trajetória discente, incluindo perfil demográfico, percurso acadêmico e indicadores institucionais de desempenho. Em conjunto, os resultados indicam que as variáveis demográficas apresentam distribuições semelhantes entre concluintes e desligados quando consideradas isoladamente, enquanto os indicadores acadêmicos evidenciam diferenças mais consistentes entre os dois desfechos.

No caso das variáveis demográficas, as variações proporcionais observadas dentro de cada curso foram, em geral, pequenas, e os desligamentos tenderam a acompanhar a composição predominante da população estudantil em cada categoria. Esse comportamento sugere capacidade limitada dessas variáveis para distinguir, isoladamente, estudantes concluintes e desligados. Ainda assim, elas foram mantidas como informações de contexto, pois ajudam a caracterizar o perfil institucional atendido e podem contribuir para a

interpretação dos resultados quando analisadas em conjunto com indicadores de desempenho.

Os indicadores acadêmicos, por sua vez, apresentaram maior capacidade de distinção entre os desfechos. Em todos os cursos, estudantes concluintes exibem valores médios e medianos mais elevados, além de menor dispersão. Entre os desligados, os valores tendem a ser mais baixos e mais heterogêneos, o que sugere a existência de trajetórias acadêmicas distintas dentro desse grupo. Esse padrão indica que o desligamento não está associado a um único perfil de desempenho, mas a diferentes combinações de dificuldades acadêmicas e progressão irregular ao longo do curso.

A análise de correlação revelou associações elevadas entre *IECH* e *IEPL*, além de correlações altas envolvendo o *IEA*. Esse resultado é coerente com a proximidade conceitual entre medidas de eficiência acadêmica e aponta para possível sobreposição parcial de informação. Embora isso não impeça o uso conjunto dessas variáveis, trata-se de um aspecto relevante tanto para a modelagem quanto para a interpretação posterior dos resultados explicáveis.

Com base nesses achados, a seleção de variáveis priorizou os indicadores *MC*, *IECH* e *IEPL*, por representarem dimensões complementares do desempenho acadêmico: rendimento global, eficiência no aproveitamento da carga horária e progressão temporal no curso. O indicador *IEA* foi mantido para fins de comparação e contextualização agregada, mas não foi adotado como variável central em todas as configurações de modelagem, em razão de sua natureza composta e da forte associação com os demais índices.

A análise também evidenciou diversidade relevante entre os cursos em termos de escala, proporção entre concluintes e desligados, perfil etário, tempo de permanência e distribuição dos indicadores acadêmicos. Esse resultado reforça a decisão de conduzir as análises de forma segmentada por curso, evitando generalizações que possam ocultar dinâmicas específicas de cada contexto formativo.

Ao final desta etapa, definiu-se o conjunto de variáveis utilizado na modelagem subsequente. Foram consideradas como variáveis categóricas *Sexo*, *Tipo_Redde_Ensino*, *Tipo_Raca*, *Faixa_Etaria*, *Regiao_Origem* e *Forma_Ingresso*. As variáveis numéricas incluíram *MC*, *IECH* e *IEPL*, enquanto a variável-alvo foi *Status*. Na seção seguinte, esses atributos são utilizados na construção dos modelos preditivos por curso, com avaliação de desempenho e análise explicável das estimativas de risco de desligamento.

5.2 Desempenho preditivo e estratificação por risco

5.2.1 Comparação de modelos por curso

O desempenho preditivo dos modelos foi avaliado por meio da métrica *AUC-PR*. Como estimativa central, utilizou-se a média das reamostragens *bootstrap* (1.000 iterações),

enquanto os intervalos de confiança de 95% foram obtidos a partir da distribuição dessas reamostragens, conforme descrito na Seção 4.4. A Tabela 9 apresenta as médias e os respectivos intervalos de confiança por curso e modelo, e a Figura 5 mostra a distribuição das estimativas obtidas no processo de reamostragem, permitindo observar a variabilidade e a estabilidade das comparações.

Tabela 9 – Desempenho dos modelos preditivos por curso (AUC-PR com IC 95%).

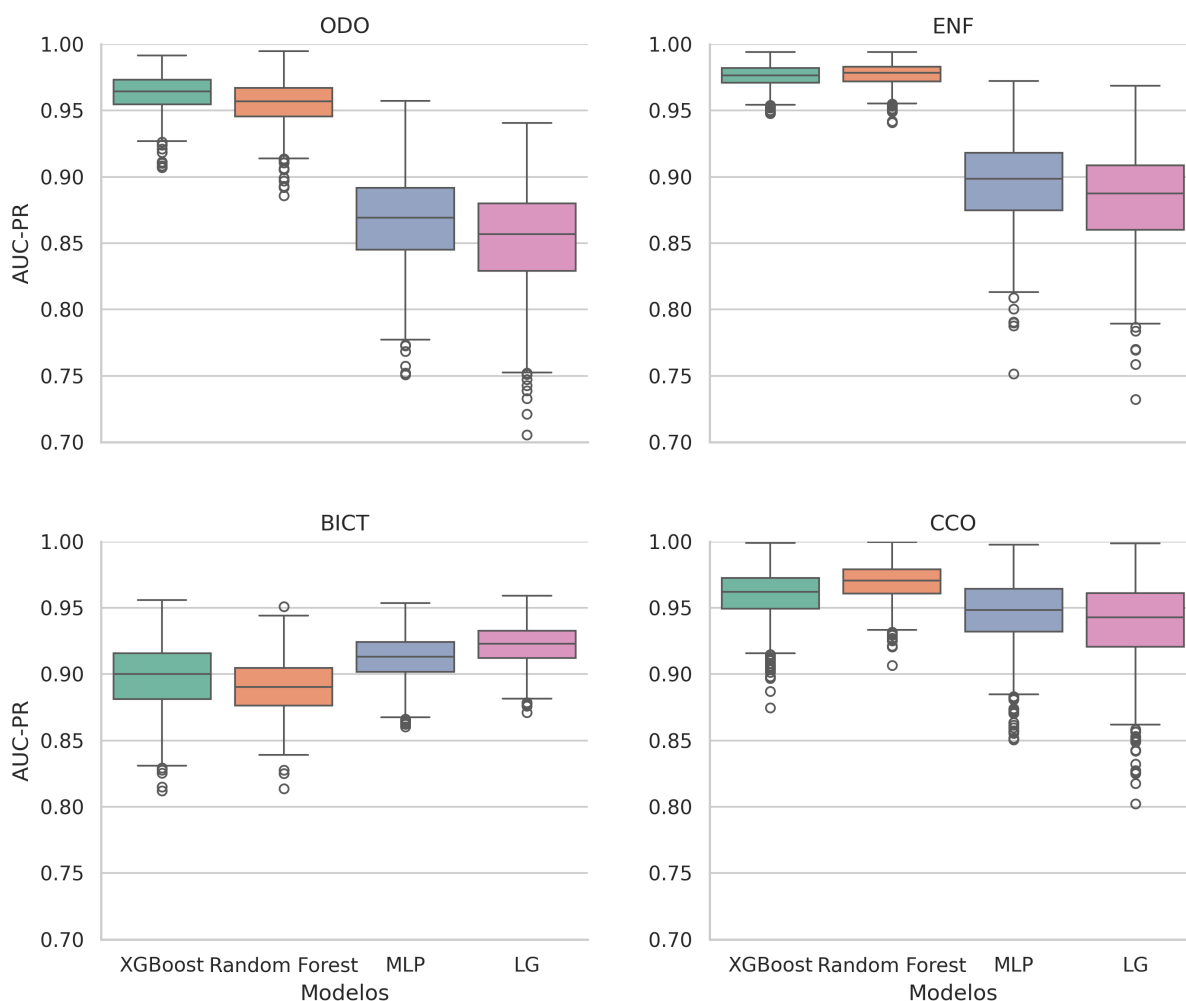
Curso	Modelo	AUC-PR (IC 95%)
ODO	XGBoost	0.9615 [0.9326 – 0.9852]
	Random Forest	0.9538 [0.9164 – 0.9850]
	MLP	0.8635 [0.7886 – 0.9281]
	Logistic Regression	0.8516 [0.7862 – 0.9194]
ENF	Random Forest	0.9770 [0.9584 – 0.9903]
	XGBoost	0.9758 [0.9564 – 0.9899]
	MLP	0.8947 [0.8255 – 0.9490]
	Logistic Regression	0.8831 [0.8500 – 0.9418]
BICT	Logistic Regression	0.9223 [0.8929 – 0.9491]
	MLP	0.9128 [0.8795 – 0.9427]
	XGBoost	0.8983 [0.8445 – 0.9218]
	Random Forest	0.8903 [0.8213 – 0.9297]
CCO	Random Forest	0.9693 [0.9393 – 0.9910]
	XGBoost	0.9597 [0.9184 – 0.9905]
	MLP	0.9462 [0.8868 – 0.9860]
	Logistic Regression	0.9388 [0.8690 – 0.9897]

Em ODO, o *XGBoost* apresentou a maior média de *AUC-PR* (0,9615; IC95% [0,9326–0,9852]), com *Random Forest* muito próximo (0,9538; [0,9164–0,9850]), sugerindo desempenho semelhante entre esses dois modelos. Em ENF, observa-se o mesmo padrão, com *Random Forest* (0,9770; [0,9584–0,9903]) e *XGBoost* (0,9758; [0,9564–0,9899]) apresentando intervalos de confiança amplamente sobrepostos, não havendo evidência clara de superioridade entre eles. Em CCO, os quatro modelos apresentaram desempenho elevado, com leve vantagem do *Random Forest* (0,9693; [0,9393–0,9910]). Em contraste, no BICT, a *Regressão Logística* obteve a maior média de *AUC-PR* (0,9223; [0,8929–0,9491]), sugerindo que, nesse curso, relações aproximadamente lineares entre os indicadores e o desfecho foram suficientes para alcançar desempenho competitivo, sem necessidade de estruturas de decisão mais complexas.

A Figura 5 complementa essa análise ao mostrar a distribuição das estimativas de *AUC-PR* obtidas nas reamostragens. Em ODO e ENF, os modelos com maiores médias também apresentam distribuições mais concentradas, indicando estimativas mais estáveis. No BICT, as distribuições são semelhantes entre os modelos, o que sugere cautela na interpretação de pequenas diferenças nas médias.

Os resultados indicam que o desempenho dos modelos varia entre os cursos. Em ODO e ENF, métodos baseados em árvores apresentam desempenho levemente

Figura 5 – Distribuição de AUC-PR por Modelo e Curso (via Bootstrap)



Elaborada pelos autores

superior, possivelmente por capturar interações e relações não lineares entre os indicadores acadêmicos. Em contraste, no BICT, o melhor desempenho da *Regressão Logística* sugere que a relação entre os indicadores e o desfecho pode ser representada de forma adequada por estruturas mais simples.

Por fim, quando os intervalos de confiança são sobrepostos, diferenças pequenas nas médias de *AUC-PR* não devem ser interpretadas como superioridade absoluta de um modelo sobre outro. Nesses casos, é recomendável priorizar a estabilidade das estimativas e a adequação do modelo ao uso aplicado das probabilidades previstas.

5.2.2 Análise complementar de erros e limiar fixo (0,5)

Conforme definido na Seção 4.4, os modelos foram avaliados com limiar de decisão fixado em 0,5, sem ajuste específico por curso. Essa escolha prioriza a comparabilidade entre algoritmos e contextos acadêmicos, ainda que não corresponda necessariamente ao

ponto de decisão que maximizaria o desempenho em cada curso. A partir desse limiar, esta subseção examina o padrão de erros observado no conjunto de teste, com foco na distribuição de verdadeiros e falsos positivos e negativos para a classe *Desligado*.

A Tabela 10 apresenta a matriz de classificação dos modelos selecionados na subseção anterior, reportando *VP*, *FN*, *FP* e *VN*, além das métricas de *Precisão* e *Recall*. De modo geral, os resultados mostram valores elevados dessas métricas em todos os cursos. O *Recall* permanece acima de 0,90, enquanto a *Precisão* também se mantém alta, indicando que os modelos conseguem identificar a maior parte dos estudantes que efetivamente vieram a se desligar e, ao mesmo tempo, manter controle sobre os alertas indevidos.

Tabela 10 – Matriz de classificação dos modelos selecionados por curso (classe positiva = Desligado, limiar = 0,5).

Curso	Modelo	VP	FN	FP	VN	Precisão	Recall
ODO	XGBoost	123	13	7	105	0.9458	0.9041
ENF	Random Forest	167	13	13	116	0.9285	0.9336
BICT	Logistic Regression	826	62	23	244	0.9737	0.9255
CCO	Random Forest	173	9	4	70	0.9774	0.9507

Nota. VP = Verdadeiro Positivo; FN = Falso Negativo; FP = Falso Positivo; VN = Verdadeiro Negativo. Precisão e Recall referem-se à classe positiva (*Desligado*).

A análise da matriz de confusão permite observar como esses erros se distribuem entre os cursos. Em ODO, ENF e CCO, o número de falsos positivos e falsos negativos permanece relativamente baixo, sugerindo equilíbrio entre identificação de casos de desligamento e controle de classificações incorretas. No BICT, embora o modelo mantenha valores altos de *Recall* e *Precisão*, observa-se maior número absoluto de falsos negativos. Esse comportamento é compatível com o maior volume de estudantes analisados nesse curso e com a maior diversidade das trajetórias acadêmicas observada na análise exploratória.

Do ponto de vista analítico, os falsos negativos merecem atenção especial, pois representam estudantes que vieram a se desligar sem terem sido previamente sinalizados pelo modelo. Já os falsos positivos correspondem a estudantes classificados como em risco que não se desligaram posteriormente. Embora esses casos possam ampliar o esforço institucional de acompanhamento, eles não implicam consequências acadêmicas diretas para o estudante e podem ser interpretados como alertas preventivos.

Ainda que limiares específicos pudessem otimizar métricas em cada curso, a adoção de um valor fixo permite observar de forma mais transparente como o perfil de erros varia entre contextos acadêmicos. Dessa forma, a análise da matriz de confusão complementa a avaliação global baseada na *AUC-PR* apresentada na subseção anterior, ao oferecer uma leitura mais concreta do comportamento dos modelos no processo de classificação.

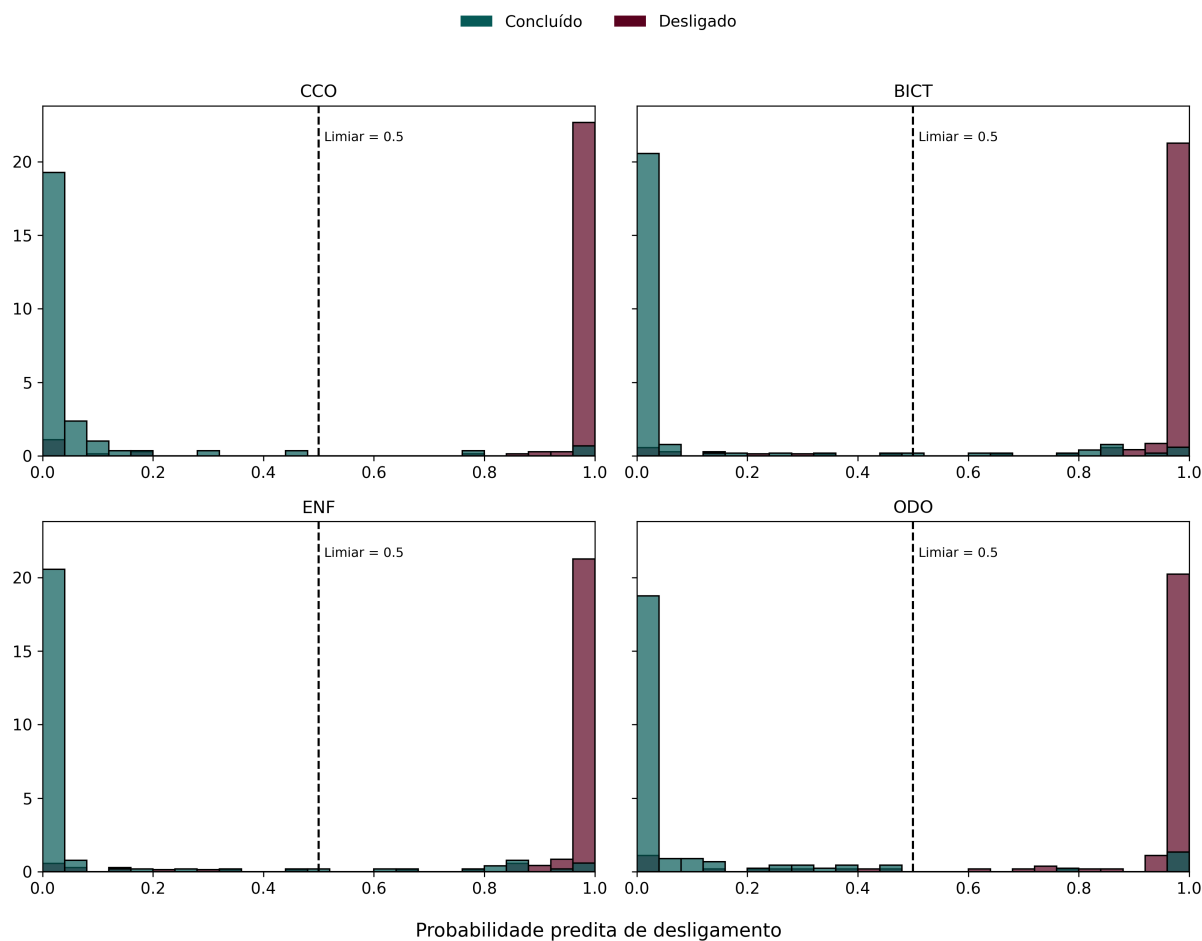
5.2.3 Faixas de risco e leitura aplicada das probabilidades

As faixas de risco foram definidas a partir das probabilidades estimadas pelos modelos com o objetivo de organizar os estudantes em grupos comparáveis quanto ao risco de desligamento. No contexto deste estudo, essa segmentação permite analisar padrões de desempenho acadêmico associados ao risco de desligamento.

A partir das probabilidades previstas, cada estudante foi classificado em duas faixas utilizando o limiar de decisão de 0,5 adotado na modelagem. Estudantes com probabilidade estimada $\geq 0,5$ foram classificados como *Alto Risco*, enquanto aqueles com probabilidade $< 0,5$ foram classificados como *Baixo Risco*.

A Figura 6 apresenta a distribuição das probabilidades estimadas pelos modelos em cada curso. É possível observar forte concentração de escores próximos de 0 e próximos de 1, com poucos casos na região intermediária. Nesse cenário, a divisão em três ou mais níveis produziria grupos pouco diferenciados. Por essa razão, a segmentação binária mostrou-se suficiente para separar os estudantes em dois perfis de risco bem distintos.

Figura 6 – Distribuição das probabilidades previstas de desligamento por curso.



Elaborada pelos autores

A Tabela 11 apresenta, para cada curso, a caracterização dessas faixas em termos dos intervalos percentílicos P10–P90 dos indicadores acadêmicos (*MC*, *IECH*, *IEPL* e *IEA*), da proporção observada de concluintes e desligados, da probabilidade média estimada de desligamento e da taxa de acerto dentro de cada grupo. O indicador *IEA* não foi utilizado na modelagem, mas foi incluído na tabela para contextualizar o perfil acadêmico das faixas, em razão de seu caráter composto. Os intervalos P10–P90 foram adotados por representarem 80% dos casos centrais em cada faixa e por reduzirem a influência de valores extremos, favorecendo uma leitura mais clara dos perfis acadêmicos.

Tabela 11 – Perfis de risco estimados pelo modelo por curso com intervalos 10–90 dos indicadores e métricas agregadas de desempenho.

Curso	Faixa de risco	MC [10–90]	IECH [10–90]	IEPL [10–90]	IEA [10–90]	% Concl.	% Desl.	Prob. Média Desligamento	Taxa de acerto
CCO	Alto	[0.00, 8.36]	[0.00, 0.49]	[0.00, 0.45]	[0.00, 1.57]	2%	98%	0.99	0.98
	Baixo	[7.98, 8.69]	[0.70, 0.97]	[0.58, 0.98]	[3.39, 7.87]	87%	13%	0.03	0.87
BICT	Alto	[0.00, 8.20]	[0.00, 0.45]	[0.00, 0.40]	[0.00, 1.42]	3%	97%	0.98	0.97
	Baixo	[7.64, 8.54]	[0.51, 0.88]	[0.44, 0.86]	[1.77, 6.10]	79%	21%	0.08	0.79
ENF	Alto	[0.00, 8.87]	[0.00, 0.80]	[0.00, 0.82]	[0.00, 4.92]	7%	93%	0.98	0.93
	Baixo	[8.22, 9.0]	[0.86, 1.0]	[0.68, 1.00]	[4.87, 8.48]	91%	9%	0.03	0.91
ODO	Alto	[0.00, 8.89]	[0.00, 0.93]	[0.00, 1.0]	[0.00, 5.92]	5%	95%	0.98	0.95
	Baixo	[8.03, 8.72]	[0.83, 0.98]	[0.84, 1.0]	[5.72, 8.21]	89%	11%	0.06	0.89

Nota. Intervalos correspondem aos percentis 10 e 90 dos indicadores dentro de cada faixa de risco. A classe positiva considerada pelo modelo é *desligado*. Prob. Média refere-se à probabilidade média estimada de desligamento em cada grupo.

De forma consistente entre os cursos, a faixa de *Alto Risco* concentra predominantemente estudantes que efetivamente se desligaram. Em CCO, por exemplo, cerca de 98% dos estudantes classificados nessa faixa pertencem à classe real de desligados, com probabilidade média próxima de 0,99. Padrão semelhante aparece nos demais cursos, com percentuais de desligamento superiores a 93% entre os estudantes classificados como alto risco. Esses resultados indicam que os modelos conseguem isolar um subconjunto de estudantes com probabilidade muito elevada de desligamento.

Na faixa de *Baixo Risco*, predomina o grupo de estudantes concluintes, acompanhado de probabilidades médias reduzidas de desligamento. Ainda assim, uma pequena parcela de estudantes desligados permanece nessa faixa. Esses casos correspondem aos falsos negativos discutidos na subseção anterior, nos quais o modelo atribuiu baixa probabilidade de risco a estudantes que posteriormente se desligaram. Eles delimitam justamente a região em que a distinção entre os perfis se torna menos nítida.

Os intervalos P10–P90 dos indicadores reforçam o contraste entre as faixas. No *Alto Risco*, predominam valores baixos de *MC*, *IECH*, *IEPL* e *IEA*, compatíveis com trajetórias de baixo rendimento e baixa progressão acadêmica. Na faixa de *Baixo Risco*, os indicadores apresentam valores mais elevados e maior estabilidade. Esses intervalos têm

caráter descritivo e não constituem regra determinística de classificação; sua finalidade é explicitar padrões típicos observados em cada faixa e apoiar a interpretação contextual por curso.

A estratificação em duas faixas permite, portanto, organizar a interpretação das probabilidades estimadas pelos modelos em termos de perfis acadêmicos distintos. O grupo de *Alto Risco* delimita estudantes que merecem maior atenção institucional, enquanto o *Baixo Risco* descreve trajetórias acadêmicas mais estáveis. Essas faixas também estruturam as análises interpretativas apresentadas nas seções seguintes, nas quais se examina como os fatores associados ao risco se manifestam em diferentes perfis de estudantes.

5.3 Explicabilidade dos Modelos

Após a avaliação do desempenho preditivo e da definição das faixas de risco, esta seção examina como os modelos estruturam suas estimativas de desligamento. A análise de explicabilidade tem como objetivo interpretar como as variáveis acadêmicas contribuem para as probabilidades estimadas pelos modelos, permitindo compreender quais padrões presentes nos dados estão associados a maiores ou menores níveis de risco.

Essa etapa busca identificar quais fatores acadêmicos influenciam as decisões do classificador e como essas contribuições se distribuem entre os diferentes cursos analisados. Ao explicitar a participação relativa das variáveis nas estimativas do modelo, torna-se possível interpretar o comportamento do classificador e compreender como os padrões históricos observados nos dados se refletem nas probabilidades de desligamento atribuídas aos estudantes.

É importante destacar que a explicabilidade é utilizada aqui como ferramenta de interpretação do modelo e não como mecanismo de inferência causal. As contribuições das variáveis indicam associações aprendidas a partir dos dados disponíveis e devem ser compreendidas como evidências sobre o funcionamento do modelo, e não como relações causais diretas com o fenômeno da evasão acadêmica.

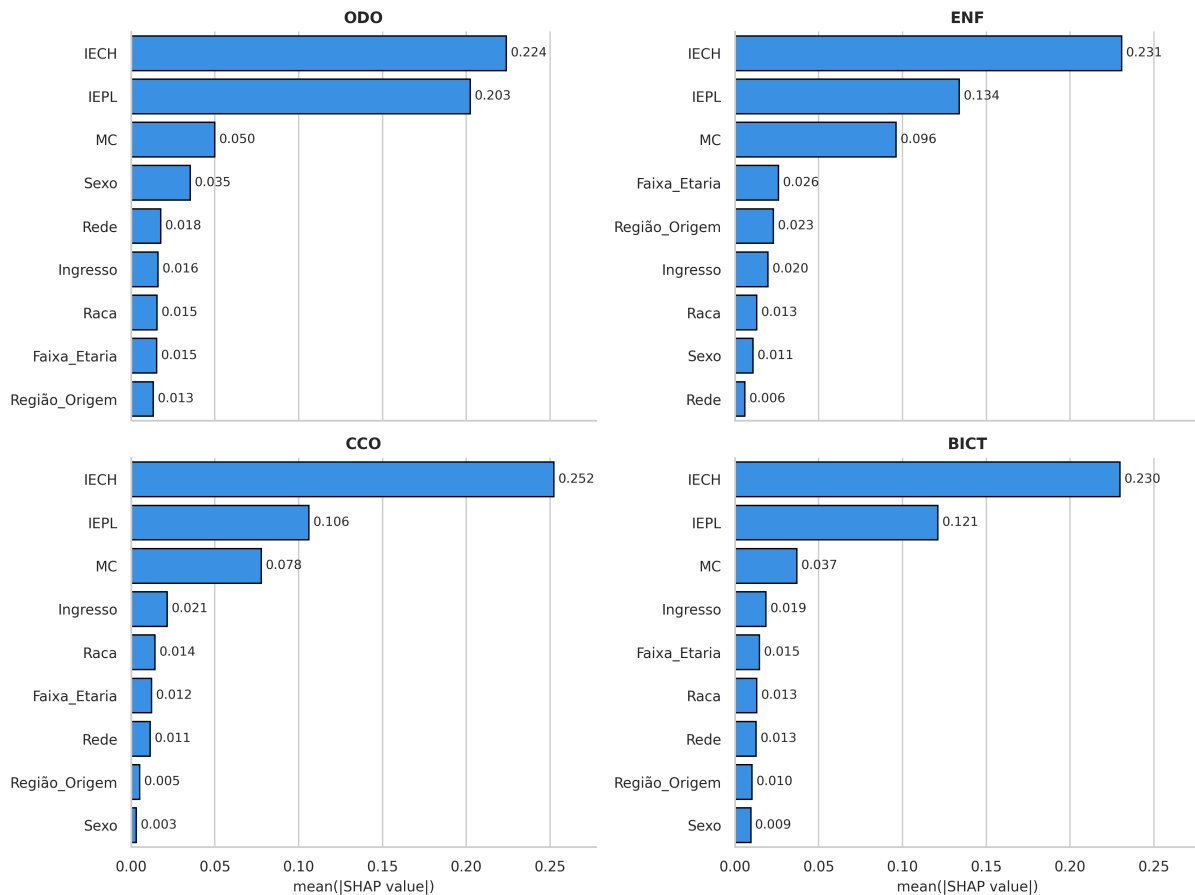
5.3.1 Explicabilidade Global do Risco por Curso

A análise de explicabilidade global teve como objetivo identificar quais variáveis mais influenciam o risco estimado de desligamento em cada curso. Diferentemente das análises locais, que examinam previsões individuais, essa abordagem considera a contribuição média das variáveis ao longo das observações do conjunto de teste, permitindo compreender quais atributos orientam o comportamento geral do modelo.

A Figura 7 apresenta a importância global das variáveis estimada por meio da média dos valores absolutos de *SHAP*. Essa medida indica a relevância média da contribuição de

cada variável para as predições do modelo, quanto maior o valor médio absoluto de SHAP, maior é a influência da variável nas estimativas de risco.

Figura 7 – Importância global das variáveis estimada por meio da média dos valores absolutos de SHAP ($\text{mean}(|\text{SHAP value}|)$) para cada curso



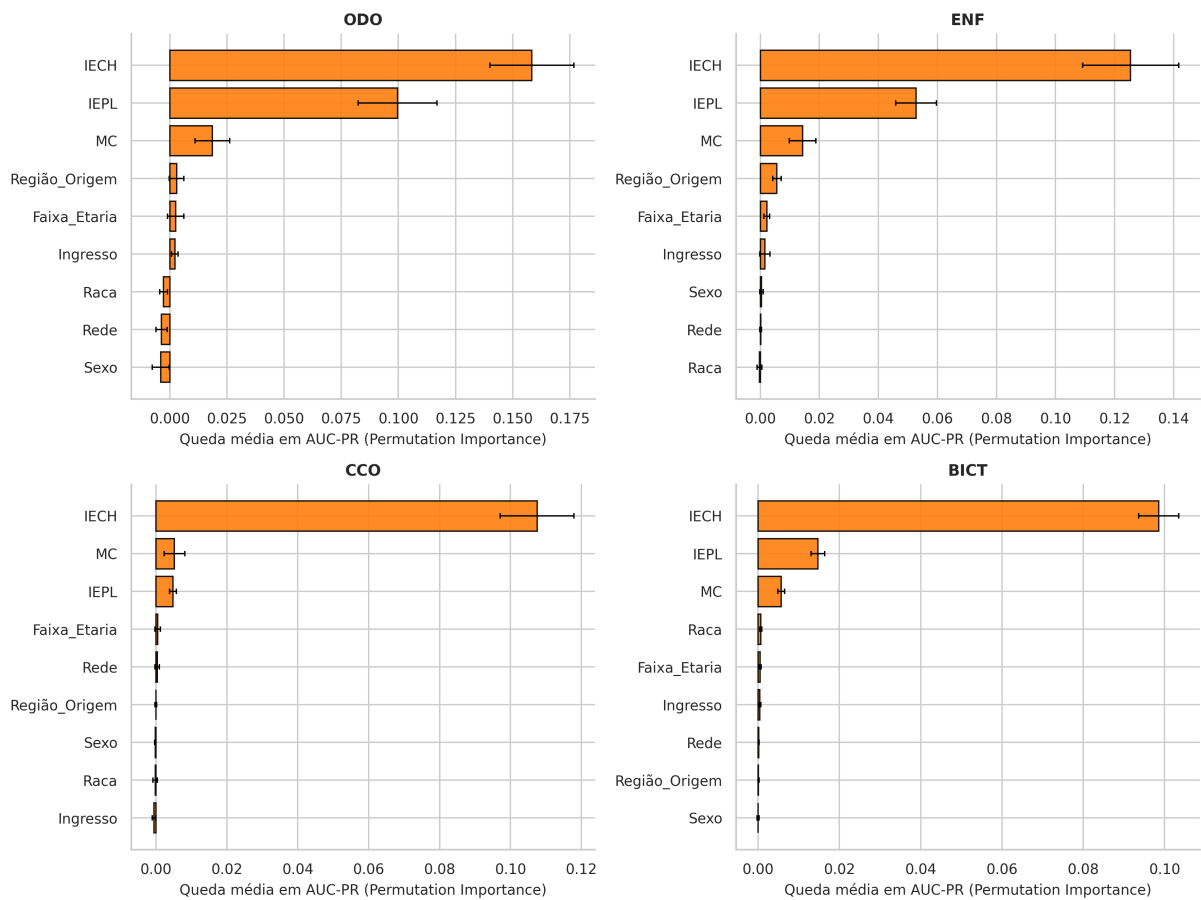
Elaborada pelos autores

Observa-se que os indicadores acadêmicos *IECH*, *IEPL* e *MC* concentram as maiores contribuições para o risco estimado de desligamento. Em todos os cursos analisados essas variáveis ocupam as primeiras posições no ranking de importância, indicando que o modelo organiza suas estimativas principalmente a partir de medidas de desempenho acadêmico.

A Figura 8 complementa essa análise ao apresentar a importância das variáveis, utilizando *Permutation Importance*. Nesse método, a relevância de cada atributo é estimada pela redução média da AUC-PR após a permutação aleatória da variável no conjunto de teste. Quanto maior a queda no desempenho do modelo, maior é a dependência do classificador em relação àquela variável.

O padrão observado é semelhante ao identificado pelo *SHAP*. A permutação de *IECH* produz a maior redução no desempenho preditivo, seguida por *IEPL* e, em menor intensidade, *MC*. Esse resultado indica que a capacidade discriminativa do modelo depende principalmente desses indicadores de desempenho acadêmico.

Figura 8 – Importância das variáveis estimada por Permutation Importance, mensurada pela queda média na AUC-PR após permutação aleatória das variáveis no conjunto de teste



Elaborada pelos autores

Para avaliar a concordância entre os ranqueamentos globais gerados pelo SHAP e pelo PI, calculou-se o coeficiente de Spearman (ρ) para cada curso. A Tabela 12 apresenta a correlação de Spearman entre os rankings derivados de SHAP e Permutation Importance.

Tabela 12 – Correlação de Spearman entre os rankings globais dos indicadores por curso.

Curso	Spearman ρ	P-valor
BICT	0.933	< 0.001
ENF	0.875	< 0.001
CCO	0.567	0.112
ODO	0.433	0.244

Nota. ρ representa o coeficiente de correlação de Spearman. Valores de $p < 0.05$ indicam evidência estatística de associação monotônica significativa.

Em BICT e ENF observa-se alta concordância entre os métodos ($\rho \approx 0,93$), indicando que ambos produzem praticamente a mesma hierarquia de importância das variáveis. Em CCO e ODO a concordância é menor, sugerindo diferenças na forma como cada método distribui a relevância entre os atributos.

Essa divergência pode estar associada à presença de variáveis correlacionadas, como *MC*, *IECH* e *IEPL*. Nesses casos, o SHAP tende a distribuir a contribuição entre variáveis relacionadas, enquanto a Permutation Importance avalia o impacto isolado da permutação de cada variável sobre o desempenho do modelo.

Embora o padrão geral seja semelhante entre cursos, o peso relativo das variáveis apresenta diferenças. Em CCO a distância entre *IECH* e as demais variáveis é mais acentuada, indicando maior centralidade desse indicador na estrutura do risco. Em ENF e BICT, *IECH* e *IEPL* apresentam contribuições mais próximas, sugerindo distribuição mais equilibrada entre eficiência em carga horária e progressão curricular. Em ODO, apesar da predominância dos indicadores acadêmicos, variáveis demográficas apresentam participação relativamente maior quando comparadas aos demais cursos.

Em geral, os resultados sugerem que os indicadores acadêmicos *IECH*, *IEPL* e *MC* constituem os principais fatores associados à estrutura preditiva do modelo. Atributos demográficos apresentam contribuição secundária. Esse padrão é consistente com estudos que identificam o desempenho acadêmico como dimensão central na previsão de desligamentos em cursos de graduação (BARROS et al., 2019; CHICON et al., 2025). Essa associação deve ser interpretada no contexto dos dados analisados e não implica relação causal direta.

5.3.2 Explicabilidade Local Estratificada por Faixa de Risco

A explicabilidade local foi empregada para interpretar como o modelo constrói suas estimativas em casos individuais, considerando as faixas de risco definidas na Subseção 5.2.3. Diferentemente da análise global, que examina padrões médios de contribuição das variáveis, a análise local permite observar como esses fatores se combinam em predições específicas.

Para essa etapa, foram selecionados casos representativos de diferentes regiões da matriz de confusão. A análise foi conduzida de forma estratificada pelas faixas de risco, permitindo examinar o comportamento do modelo tanto em situações de acerto quanto de erro.

Foram analisados quatro tipos de situação: *verdadeiro positivo* e *falso positivo* na faixa de *Alto Risco*, e *verdadeiro negativo* e *falso negativo* na faixa de *Baixo Risco*. Essa organização permite comparar como as contribuições das variáveis se distribuem em diferentes cenários de decisão do classificador.

O objetivo da análise não é generalizar conclusões a partir de exemplos isolados, mas ilustrar como o modelo combina diferentes atributos ao estimar o risco de desligamento em contextos distintos, incluindo situações em que a predição coincide ou diverge do desfecho observado.

5.3.2.1 Explicabilidade Local - Faixa de Alto Risco

Nesta subseção são examinadas previsões individuais classificadas na faixa de *Alto Risco*. A análise utiliza dois casos do curso de CCO: um *Verdadeiro Positivo*, no qual o modelo identificou corretamente o desligamento do estudante, e um *Falso Positivo*, em que o modelo atribuiu alto risco a um estudante que posteriormente concluiu o curso.

5.3.2.1.1 Caso 1: acerto na previsão(*Verdadeiro Positivo*)

A Tabela 13 apresenta os atributos do estudante analisado. Para esse caso, o modelo atribuiu probabilidade de desligamento igual a 0,99, e o desfecho real observado foi de desligamento.

Tabela 13 – Atributos do caso *Verdadeiro Positivo* na faixa de Alto Risco (CCO).

Campo	Valor
Situação real	Desligado
Classe prevista	Desligado
Probabilidade prevista de desligamento	0.99
MC	9.0
IECH	0.1
IEPL	0.050
IEA	0.045
Tempo de permanência (períodos)	6

A Figura 9 mostra as explicações produzidas pelas três técnicas utilizadas. No *SHAP*, valores reduzidos de *IECH* e *IEPL* aparecem como as principais contribuições para o aumento da probabilidade de risco. O *LIME* atribui maior peso local às mesmas variáveis, reforçando esse padrão. Já o método *Anchors* identifica uma regra associada a valores baixos de *IECH*, indicando que, quando essa condição ocorre, o modelo tende a manter a previsão de desligamento.

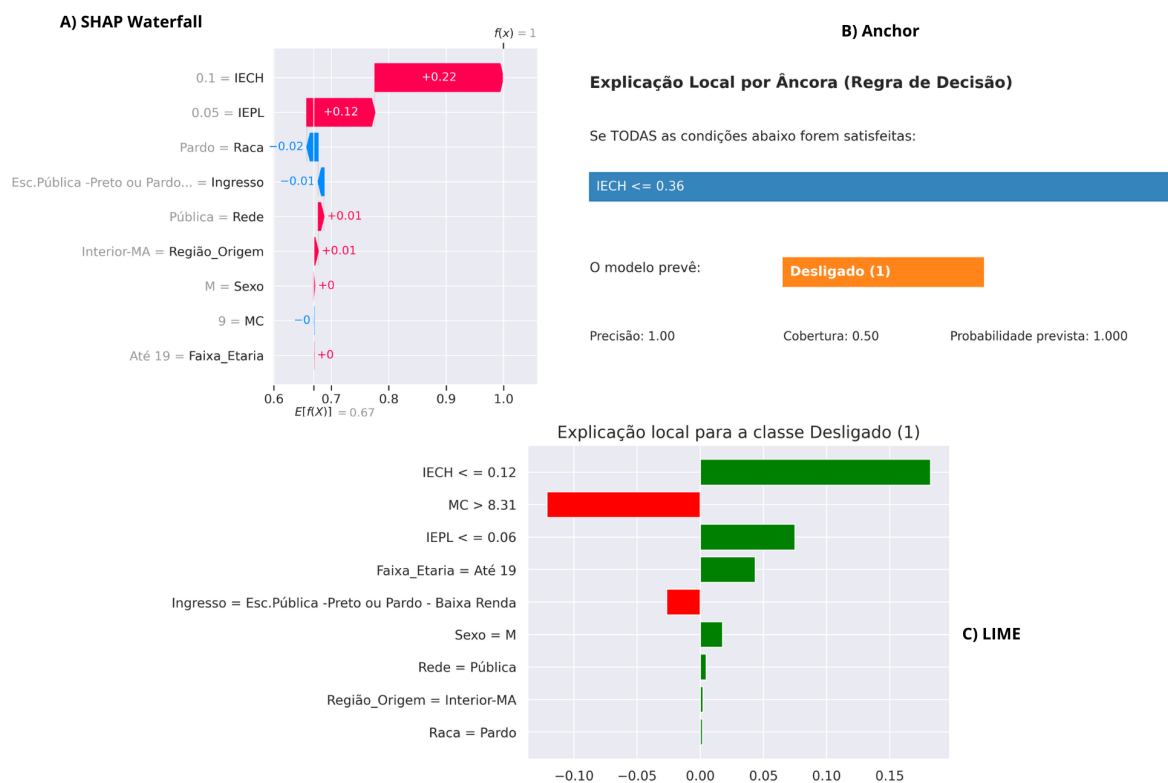
Esse resultado é consistente com o padrão observado na análise global. A decisão do modelo é sustentada principalmente por indicadores de desempenho acadêmico, o que reforça a coerência entre a estrutura geral do modelo e as explicações locais apresentadas para este caso.

5.3.2.1.2 Caso 2: erro na previsão(*Falso Positivo*)

A Tabela 14 e a Figura 10 apresentam um estudante classificado como alto risco (score 0,91), cujo desfecho real foi de *Conclusão*. Esse caso ilustra uma situação em que o modelo atribuiu elevada probabilidade de desligamento, embora o estudante tenha concluído o curso.

No gráfico *SHAP*, observa-se que *IECH* e *IEPL* apresentam as maiores contribuições positivas para a classe *Desligado*, elevando o score de risco. O *LIME* confirma esse padrão ao atribuir maior peso local às mesmas variáveis. Em contraste, o indicador *MC* aparece

Figura 9 – Explicações locais (*SHAP*, *LIME* e *Anchors*) para o caso *Verdadeiro Positivo* na faixa de Alto Risco (CCO).



Elaborada pelos autores

Tabela 14 – Atributos do caso *Falso Positivo* na faixa de Alto Risco (CCO).

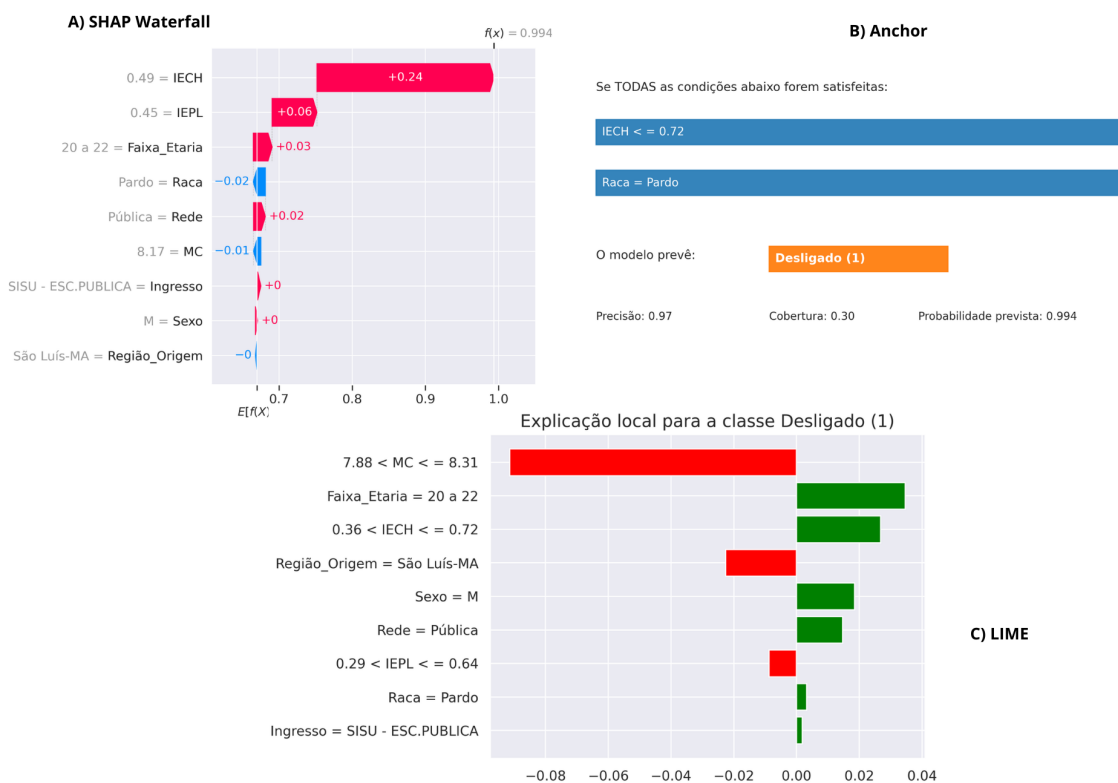
Campo	Valor
Situação real	Concluído
Classe prevista	Desligado
Probabilidade prevista de desligamento	0.91
MC	8.16
IECH	0.49
IEPL	0.45
IEA	1.80
Tempo de permanência (períodos)	15

com contribuição negativa, reduzindo parcialmente o risco estimado. Esse conjunto de contribuições revela um perfil intermediário, no qual diferentes indicadores apontam direções distintas para a decisão do modelo.

A regra identificada pelo método *Anchors* também reflete essa condição. A explicação é baseada principalmente em valores intermediários de *IECH*, resultando em uma regra mais específica e com menor cobertura quando comparada ao caso anterior. Esse padrão indica uma região do espaço de atributos em que as decisões do modelo tendem a apresentar maior incerteza.

Nesse caso, os indicadores acadêmicos se aproximam do limite superior observado na caracterização da faixa de alto risco (Subseção 5.2.3). O erro, portanto, é coerente com a

Figura 10 – Explicações locais (*SHAP*, *LIME* e *Anchors*) para o caso *Falso Positivo* na faixa de Alto Risco (CCO).



Elaborada pelos autores

lógica do classificador. Como o modelo aprende padrões históricos agregados, combinações de atributos semelhantes ao perfil típico de desligamento tendem a receber escores elevados de risco, mesmo quando a trajetória individual culmina em conclusão.

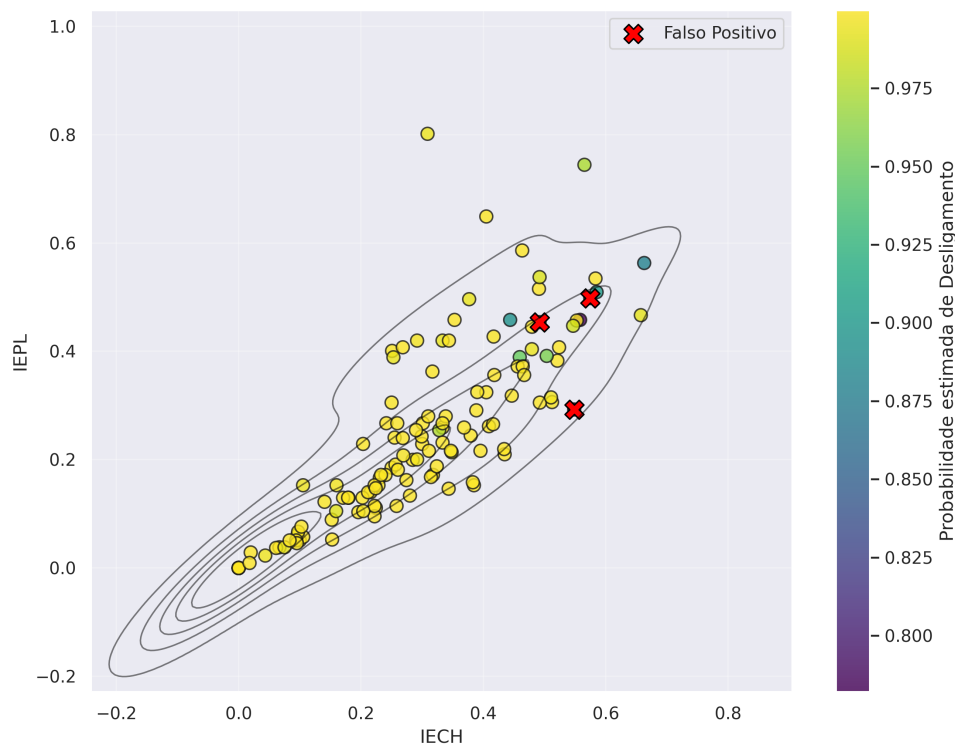
5.3.2.1.3 Síntese da faixa de alto risco

A comparação entre os dois casos analisados mostra que, na faixa de *Alto Risco*, as decisões do modelo são fortemente influenciadas pelos indicadores *IECH* e *IEPL*. No caso de acerto (*Verdadeiro Positivo*), valores muito baixos desses indicadores produzem explicações convergentes entre os métodos e resultam em escores de risco próximos de 1. Já em perfis com valores mais intermediários, como observado no *Falso Positivo*, a contribuição dessas variáveis permanece associada ao padrão de risco, porém com menor intensidade, o que torna a decisão do modelo mais sensível a pequenas variações nos atributos.

A Figura 11 ajuda a contextualizar esse comportamento no conjunto de teste do CCO. Observa-se uma concentração de estudantes classificados como alto risco ao longo de uma região diagonal no plano *IECH-IEPL*, delimitada pelos contornos de densidade. As cores representam a probabilidade estimada de desligamento e indicam que os escores

permanecem elevados dentro dessa região. Esse padrão mostra que combinações de baixa eficiência em carga horária e baixa progressão curricular estão associadas a maiores probabilidades de desligamento.

Figura 11 – Regiões empíricas de risco – Faixa Alto Risco (CCO)



Elaborada pelos autores

Os casos de *Falso Positivo* tendem a aparecer próximos ao limite superior dessa região. Esses estudantes apresentam níveis de *IECH* e *IEPL* mais altos do que os perfis mais críticos dentro da faixa, aproximando-se de níveis de desempenho compatíveis com trajetórias de conclusão. Nesse contexto, o erro não decorre de incoerência do modelo, mas da proximidade desses casos com a transição entre perfis de maior risco e trajetórias acadêmicas mais estáveis.

Do ponto de vista do monitoramento acadêmico, esses resultados indicam que a classificação de um estudante na faixa de *Alto Risco* deve ser interpretada como um sinal de prioridade para acompanhamento, e não como um rótulo definitivo de desligamento. A faixa concentra estudantes com probabilidade elevada de evasão, mas inclui perfis com diferentes níveis de fragilidade acadêmica.

5.3.2.2 Explicabilidade Local – Faixa de Baixo Risco

Nesta subseção, a explicabilidade local é utilizada para interpretar previsões individuais classificadas na faixa de *Baixo Risco*, conforme definido na Subseção 5.2.3.

Mantendo a comparabilidade analítica com a faixa de *Alto Risco*, são examinados dois casos do curso de CCO: um *Verdadeiro Negativo* (baixo risco e conclusão confirmada) e um *Falso Negativo* (baixo risco, mas desligamento).

5.3.2.2.1 Caso 1: acerto na predição (*Verdadeiro Negativo*)

A Tabela 15 apresenta os atributos do estudante classificado como *Baixo Risco*, com escore estimado de desligamento igual a 0,01 e situação real de *Conclusão*. Os indicadores acadêmicos apresentam valores elevados de *IECH* (0,96) e *IEPL* (0,91). O *IEA* (7,8) e o tempo de permanência (9 períodos) indicam uma trajetória acadêmica regular.

Tabela 15 – Atributos do caso *Verdadeiro Negativo* na faixa de Baixo Risco (CCO).

Campo	Valor
Situação real	Concluído
Classe prevista	Concluído
Probabilidade prevista de desligamento	0.01
MC	8.9
IECH	0.96
IEPL	0.91
IEA	7.8
Tempo de permanência (períodos)	9

Na Figura 12, as três técnicas apresentam padrão convergente de explicação. No *SHAP*, os valores elevados de *IECH* e *IEPL* contribuem negativamente para a classe *Desligado*, reduzindo o escore de risco para próximo de zero. O *LIME* atribui maior peso local às mesmas variáveis como fatores que diminuem o risco estimado. O método *Anchors* identifica uma regra baseada em limiares superiores desses indicadores, indicando que, quando essas condições são atendidas, o modelo tende a classificar o estudante como *Concluído*.

Esse caso evidencia uma região decisória estável do modelo. Valores elevados de eficiência acadêmica reduzem sistematicamente a probabilidade estimada de desligamento, produzindo explicações consistentes entre os métodos.

5.3.2.2.2 Caso 2: erro na predição (*Falso Negativo*)

A Tabela 16 apresenta um estudante classificado como *Baixo Risco*, com escore estimado de desligamento igual a 0,06, cujo desfecho real foi de *Desligado*. Os indicadores acadêmicos apresentam valores relativamente elevados de *IECH* (0,82) e *IEPL* (0,72), o que explica a classificação inicial como baixo risco. No entanto, o *IEA* (4,9) e o tempo de permanência (4 períodos) sugerem uma trajetória menos consolidada.

Na Figura 13, observa-se que as três técnicas mantêm coerência interna na decisão do modelo. No *SHAP*, os valores de *IECH* e *IEPL* contribuem negativamente para a classe *Desligado*, reduzindo o escore de risco. O *LIME* atribui maior peso às mesmas

Figura 12 – Explicações locais (*SHAP*, *LIME* e *Anchors*) para o caso *Verdadeiro Negativo* na faixa de Alto Risco (CCO).



Elaborada pelos autores

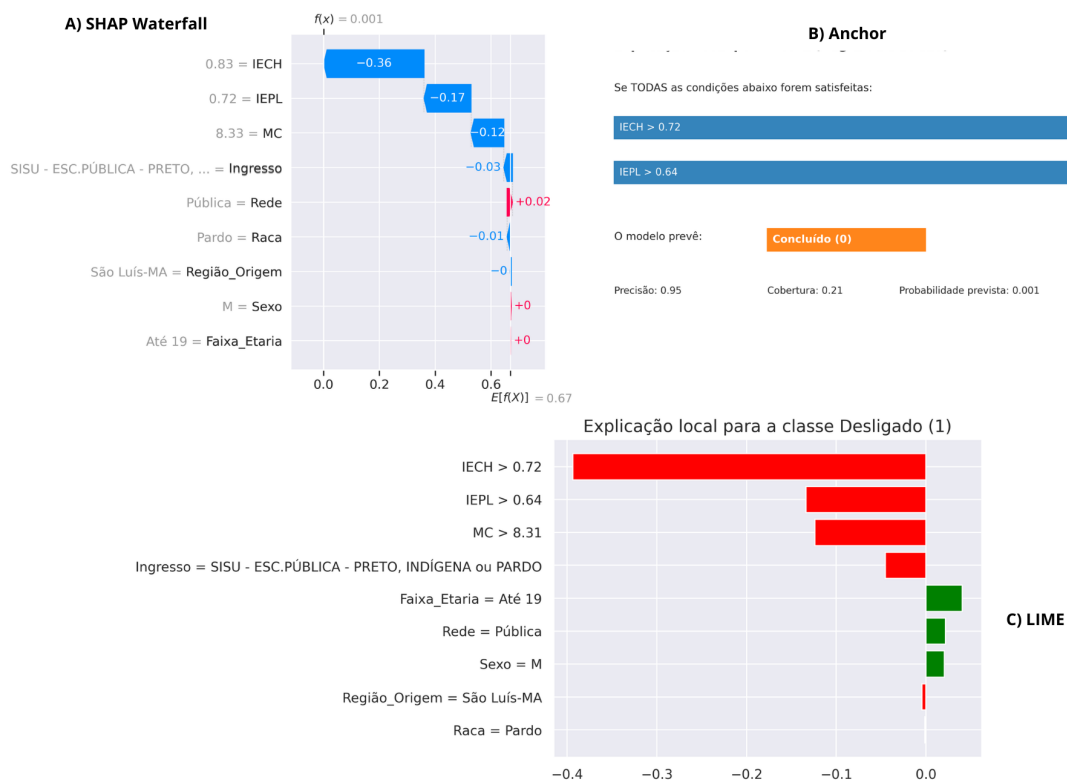
Tabela 16 – Atributos do caso *Falso Negativo* na faixa de Baixo Risco (CCO).

Campo	Valor
Situação real	Desligado
Classe prevista	Concluído
Probabilidade prevista de desligamento	0.06
MC	8.3
IECH	0.82
IEPL	0.72
IEA	4.9
Tempo de permanência (períodos)	4

variáveis como fatores que diminuem a probabilidade estimada de desligamento. O método *Anchors* identifica regra baseada em limiares superiores desses indicadores, sustentando a classificação como *Concluído*. As explicações são consistentes com a lógica aprendida pelo modelo.

Esse caso evidencia que a subestimação do risco tende a ocorrer quando o perfil observado se aproxima do padrão típico de estabilidade acadêmica, mas a trajetória real apresenta descontinuidade. O modelo identifica corretamente a estrutura histórica predominante, porém não capta fatores adicionais ou eventos posteriores que influenciam o desligamento. Trata-se, portanto, de um limite associado à natureza dos dados administrativos utilizados na modelagem.

Figura 13 – Explicações locais (*SHAP*, *LIME* e *Anchors*) para o caso *Falso Negativo* na faixa de Alto Risco (CCO).



Elaborada pelos autores

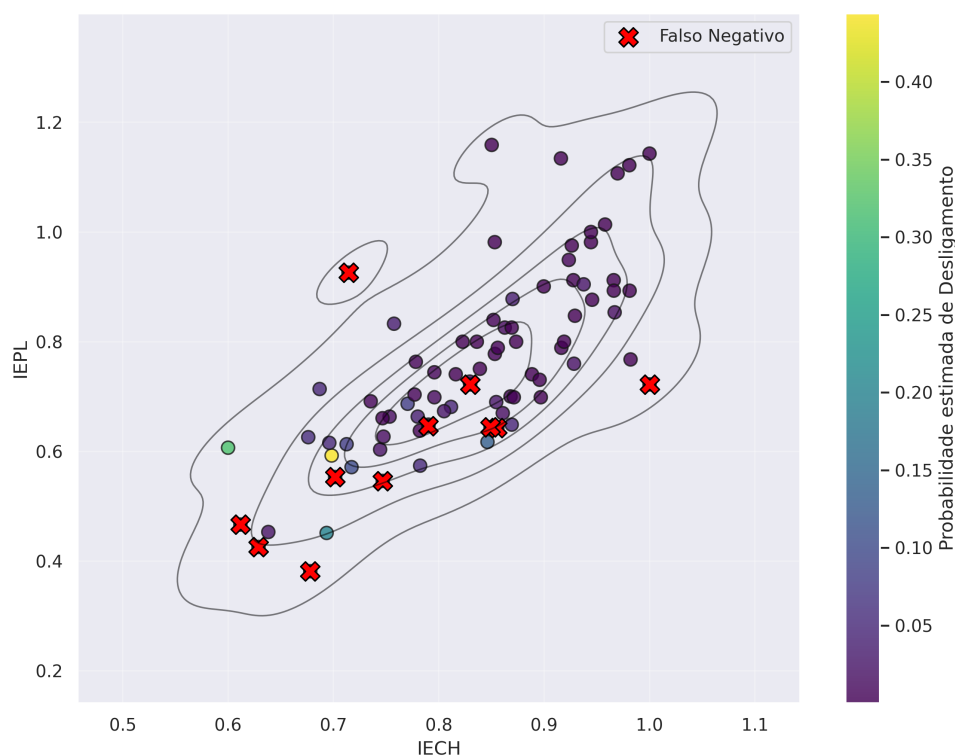
5.3.2.2.3 Síntese da faixa de baixo risco

A análise dos casos indica a existência de uma região relativamente estável no comportamento do modelo, caracterizada por valores elevados de *IECH* e *IEPL*. Nesses perfis, as três técnicas de explicabilidade convergem ao indicar contribuições que reduzem o escore estimado de desligamento, o que sustenta a classificação do estudante como *Concluído*.

A visualização conjunta de *IECH* e *IEPL* no conjunto de teste (Figura 14) reforça essa interpretação. Observa-se uma concentração de estudantes com baixa probabilidade estimada de desligamento em uma faixa contínua associada a níveis elevados de eficiência acadêmica. Essa região também apresenta menor presença de desligamentos históricos, o que contribui para a estabilidade das predições nesse intervalo.

O caso de *Falso Negativo* evidencia, contudo, um limite do modelo. Embora os indicadores acumulados posicionem o estudante na região de maior estabilidade, o desfecho real foi de desligamento. Esse resultado sugere que o erro pode estar associado a mudanças ocorridas ao longo da trajetória acadêmica ou a fatores que não estão representados nos dados administrativos utilizados. Como o modelo opera com indicadores consolidados,

Figura 14 – Regiões empíricas de risco – Faixa Baixo Risco (CCO)



Elaborada pelos autores

ele não captura variações tardias ou eventos externos capazes de alterar o percurso do estudante.

Esse padrão permite diferenciar dois tipos de erro observados nas análises locais. Na faixa de alto risco, o *Falso Positivo* corresponde a uma superestimação do risco em perfis próximos à região de transição. Já na faixa de baixo risco, o *Falso Negativo* reflete uma subestimação do risco em trajetórias que aparentam estabilidade acadêmica. Em ambos os casos, os erros não indicam incoerência nas explicações do modelo, mas refletem limites associados às informações disponíveis nos dados.

Em conjunto, esses resultados sugerem que a incorporação futura de variáveis dinâmicas ou indicadores que capturem a evolução temporal da trajetória acadêmica pode reduzir situações de subestimação do risco. O modelo identifica com consistência padrões estruturais de estabilidade, mas sua capacidade preditiva permanece condicionada à natureza acumulada dos dados utilizados.

5.4 Síntese dos resultados

Os resultados apresentados neste capítulo indicam que a análise exploratória, a modelagem preditiva e as técnicas de explicabilidade convergem para um mesmo padrão

geral. Na caracterização inicial dos dados, observou-se diversidade entre os cursos quanto à distribuição dos estudantes, ao tempo de permanência e ao comportamento dos indicadores acadêmicos. As variáveis demográficas, quando consideradas isoladamente, apresentaram capacidade limitada para diferenciar concluintes e desligados. Em contraste, os indicadores acadêmicos revelaram diferenças mais consistentes entre os dois desfechos.

Esse padrão também se manifesta na etapa de modelagem preditiva. As probabilidades estimadas pelos modelos e sua organização em faixas de risco indicam que o risco de desligamento se estrutura principalmente em torno de três dimensões do desempenho estudantil: a eficiência na conversão da carga horária matriculada em aprovações (*IECH*), o ritmo de progressão ao longo do curso (*IEPL*) e o desempenho acadêmico global (*MC*). Estudantes com valores mais elevados nesses indicadores tendem a se concentrar na faixa de *Baixo Risco*, associada a trajetórias acadêmicas mais estáveis. Por outro lado, valores reduzidos aparecem com maior frequência entre estudantes classificados na faixa de *Alto Risco*, refletindo padrões de menor aproveitamento e progressão mais lenta ao longo do curso.

As análises de explicabilidade reforçam essa interpretação. Na explicabilidade global, métodos como *SHAP* e *Permutation Importance* posicionaram *IECH*, *IEPL* e *MC* entre os atributos mais influentes na estrutura das estimativas de risco. Nas análises locais, *SHAP*, *LIME* e *Anchors* mostraram que esses mesmos indicadores também organizam as predições em nível individual, tanto nos casos de acerto quanto nos casos de erro. Nos perfis mais extremos, as explicações tendem a ser mais estáveis e convergentes entre os métodos. Já nas regiões intermediárias, pequenas variações nos indicadores acadêmicos tornam a decisão do modelo mais sensível, concentrando parte dos falsos positivos e falsos negativos observados.

Em conjunto, esses resultados sugerem que o desligamento acadêmico, tal como representado pelos dados analisados, está mais fortemente associado a padrões de desempenho e progressão ao longo do curso do que a características demográficas isoladas. Ao mesmo tempo, a análise local evidenciou que algumas trajetórias escapam ao padrão predominante aprendido pelo modelo, especialmente em situações de transição ou em casos influenciados por fatores não capturados nos dados administrativos. Isso reforça a necessidade de interpretar as predições como apoio à análise institucional, e não como diagnóstico definitivo.

Por fim, a explicabilidade mostrou-se útil por tornar mais transparente a forma como o modelo combina as variáveis disponíveis para estimar o risco de desligamento. Nesse sentido, os métodos explicativos não prescrevem decisões, mas oferecem suporte para que gestores e docentes contextualizem melhor os casos analisados e compreendam os limites e possibilidades do uso de modelos preditivos no acompanhamento acadêmico (SAARELA; PODGORELEC, 2024; ALTUKHI; PRADHAN, 2025).

6 Conclusão

Esta dissertação investigou como técnicas de Inteligência Artificial Explicável podem ser aplicadas a modelos preditivos com o objetivo de interpretar estimativas de risco de evasão acadêmica em cursos de graduação da Universidade Federal do Maranhão. O estudo partiu do reconhecimento de que modelos de aprendizado de máquina podem apresentar bom desempenho preditivo, mas frequentemente oferecem pouca transparência quanto aos fatores que influenciam suas decisões, o que limita sua incorporação mais ampla às práticas de análise e acompanhamento acadêmico.

Diante desse problema, este trabalho propôs, como contribuição central, um processo analítico reutilizável que integra análise exploratória para compreensão dos dados institucionais, modelagem preditiva segmentada por curso para estimar o risco de desligamento e técnicas de explicabilidade voltadas à interpretação das previsões. A articulação dessas etapas em uma mesma abordagem mostra como estimativas probabilísticas podem ser convertidas em evidências interpretáveis sobre trajetórias acadêmicas e padrões de desempenho associados a estudantes em risco de desligamento.

Além disso, o estudo contribui, em termos empíricos e metodológicos, para ampliar o uso do aprendizado de máquina em análises educacionais e aproximar a ciência de dados das demandas existentes da gestão universitária. A estrutura proposta também pode ser adaptada a outras instituições, desde que sejam identificados indicadores locais equivalentes às dimensões de eficiência e progressão acadêmica e ajustados os limiares de risco às políticas e às características curriculares de cada curso.

Do ponto de vista empírico, a análise indicou que a eficiência na conversão da carga horária matriculada em aprovações (*IECH*), o ritmo de progressão ao longo do curso (*IEPL*) e a média de conclusão (*MC*) constituem as dimensões que mais estruturam as previsões dos modelos, enquanto os atributos demográficos apresentaram influência secundária. Esses resultados sugerem que trajetórias acadêmicas marcadas por progressão abaixo do esperado e menor aproveitamento das disciplinas tendem a concentrar estimativas mais elevadas de risco. A identificação desses padrões reforça a importância de acompanhar estudantes que apresentam sinais persistentes de baixo desempenho e de considerar estratégias institucionais de apoio acadêmico, como ações de nivelamento, orientação e tutoria.

A análise também evidenciou diferenças entre cursos quanto à distribuição do risco acadêmico. Em programas com maiores taxas históricas de desligamento, como Ciência da Computação e BICT, as faixas de risco identificadas pelos modelos ajudam a localizar grupos que demandam acompanhamento prioritário. Em cursos com menor evasão, como Enfermagem e Odontologia, os mesmos indicadores podem apoiar a identificação

mais precoce de mudanças na trajetória acadêmica. Nesse contexto, o processo analítico proposto não apenas torna mais visíveis os padrões associados ao risco de desligamento, mas também oferece uma base objetiva para apoiar práticas institucionais de acompanhamento estudantil e decisões pedagógicas voltadas à permanência e à conclusão dos cursos.

Apesar dos resultados alcançados, algumas limitações devem ser consideradas. O conjunto de dados utilizado inclui principalmente indicadores acadêmicos e características demográficas presentes nos registros administrativos da universidade. Outros fatores frequentemente associados à evasão, como condições socioeconômicas, aspectos psicossociais e níveis de engajamento estudantil, não estavam disponíveis na base analisada. Dessa forma, os padrões identificados refletem relações estatísticas observadas nas variáveis disponíveis e não devem ser interpretados como evidências causais do fenômeno da evasão acadêmica.

Como continuidade desta pesquisa, estudos futuros podem ampliar o escopo das análises por meio da incorporação de novas fontes de dados institucionais e da utilização de abordagens que considerem a evolução temporal das trajetórias acadêmicas. A análise de dados longitudinais pode contribuir para capturar mudanças ao longo do percurso dos estudantes e melhorar a sensibilidade dos modelos a variações na progressão acadêmica. Também se mostra relevante investigar a aplicação do processo analítico proposto em outras instituições de ensino superior, de modo a avaliar sua consistência em diferentes contextos educacionais.

Referências

- ALTUKHI, Z. M.; PRADHAN, S. Systematic literature review: Explainable AI definitions and challenges in education. *arXiv*, 2025. Citado 3 vezes nas páginas 23, 24 e 69.
- AMBIEL, R. A. M.; CORTEZ, P. A.; SALVADOR, A. P. Predição da potencial evasão acadêmica entre estudantes trabalhadores e não trabalhadores. *Psicologia: Teoria e Pesquisa*, v. 37, p. e37305, 2021. Citado na página 20.
- BARROS, T. M.; PACHECO, R. A. C.; OLIVEIRA, A. C. de; CORREA, D. D.; SANTOS, J. A. dos. Determination of dropout student profile based on correspondence analysis technique. *IEEE Latin America Transactions*, v. 17, n. 9, p. 1517–1523, 2019. Citado na página 60.
- BARROSO, P. C. F.; OLIVEIRA, Í. M.; NORONHA-SOUSA, D.; NORONHA, A.; MATEUS, C. C.; VÁZQUEZ-JUSTO, E.; COSTA-LOBO, C. Fatores de evasão no ensino superior: uma revisão de literatura. *Psicologia Escolar e Educacional*, v. 26, p. e228736, 2022. Citado na página 20.
- 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. Citado 2 vezes nas páginas 32 e 34.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. New York: Springer, 2006. Citado na página 21.
- BONIFRO, F. D.; GABBRIELLI, M.; LISANTI, G.; ZINGARO, S. P. Student dropout prediction. In: *Proceedings of the International Conference on Artificial Intelligence in Education*. [S.l.: s.n.], 2020. p. 129–140. Citado 2 vezes nas páginas 21 e 22.
- Brasil. Ministério da Educação. Comissão especial de estudos sobre a evasão nas universidades públicas brasileiras. *Ministério da Educação*, 1997. Citado na página 20.
- CAMARGOS, R. C.; SILVEIRA, I. F. Técnicas de aprendizado de máquina interpretáveis na predição de evasão escolar: uma revisão. *Revista Ibérica de Sistemas e Tecnologias de Informação*, p. 636–646, 2024. Citado na página 27.
- CHICON, P. M. M.; PASCHOAL, L. N.; SAWICKI, S.; ROOS-FRANTZ, F.; FRANTZ, R. Z. A predictive model for the early identification of student dropout using data classification, clustering, and association methods. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, v. 20, p. 12–21, 2025. Citado na página 60.
- DANTAS, J. C. da S.; ARANHA, E. H. da S.; SILVA, T. R. da. Predição de alunos em risco de reprovação: Uma comparação do impacto de diferentes técnicas de amostragem. *RENOTE*, v. 22, n. 3, p. 230–239, 2024. Citado na página 40.
- DEKKER, G. W.; PECHENIZKIY, M.; VLEESHOUWERS, J. M. Predicting students drop out: A case study. In: *Proceedings of the 2nd International Conference on Educational Data Mining (EDM 2009)*. Córdoba, Spain: [s.n.], 2009. p. 41–50. Citado 2 vezes nas páginas 30 e 48.

- DELEN, D. Predicting student attrition with data mining methods. *Journal of College Student Retention: Research, Theory & Practice*, v. 13, n. 1, p. 17–35, 2011. Citado na página 31.
- DOSHI-VELEZ, F.; KIM, B. Towards a rigorous science of interpretable machine learning. *arXiv*, 2017. Citado na página 25.
- FARROW, R. The possibilities and limits of explicable artificial intelligence (XAI) in education: A socio-technical perspective. *Learning, Media and Technology*, v. 48, n. 2, p. 266–279, 2023. Citado 2 vezes nas páginas 23 e 24.
- FIGUEIREDO, D. J.; SOUZA, R. R. *Métodos de Análise de Dados*. Rio de Janeiro: LTC, 2012. ISBN 978-8521618330. Citado na página 39.
- FILHO, R. L. L. S.; MOTEJUNAS, P. R.; HIPÓLITO, O.; LOBO, M. B. C. M. A evasão no ensino superior brasileiro. *Cadernos de Pesquisa*, v. 37, n. 132, p. 641–659, 2007. Citado na página 20.
- FIOK, K.; FARAHANI, F. V.; KARWOWSKI, W.; AHRAM, T. Explainable artificial intelligence for education and training. *Journal of Defense Modeling and Simulation*, v. 19, n. 2, p. 133–144, 2022. Citado 2 vezes nas páginas 23 e 24.
- GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. 2nd. ed. [S.l.]: O'Reilly Media, 2021. Citado 2 vezes nas páginas 21 e 40.
- GOMES, M.; HIRATA, G. Determinantes da evasão no ensino superior: uma abordagem de riscos competitivos. *Pesquisa e Planejamento Econômico*, v. 52, n. 2, p. 1–32, 2022. Citado na página 20.
- GUIDOTTI, R.; MONREALE, A.; RUGGIERI, S.; TURINI, F.; GIANNOTTI, F.; PEDRESCHI, D. A survey of methods for explaining black box models. *ACM Computing Surveys*, v. 51, n. 5, p. 1–42, 2018. Citado na página 25.
- HARRISON, M. *Machine learning pocket reference: Working with structured data in Python*. [S.l.]: O'Reilly Media, 2020. Citado na página 40.
- HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. 2. ed. New York: Springer, 2009. Citado na página 21.
- Inep. *Painel Estatístico do Censo da Educação Superior*. 2025. Acesso em: 2025. Disponível em: <<https://www.gov.br/inep/pt-br/aceso-a-informacao/dados-abertos/inep-data>>. Citado 2 vezes nas páginas 15 e 44.
- KAADOUD, I. C.; FAHED, L.; LENCA, P. Explainable AI: A narrative review at the crossroad of knowledge discovery, knowledge representation and representation learning. In: *Proceedings of the Twelfth International Workshop on Modelling and Reasoning in Context (MRC 2021)*. [S.l.: s.n.], 2021. v. 2995, p. 28–40. Citado na página 23.
- KHOSRAVI, H.; SHUM, S. B.; CHEN, G.; CONATI, C.; TSAI, Y.-S.; KAY, J.; KNIGHT, S.; MARTINEZ-MALDONADO, R.; SADIQ, S.; GAŠEVIĆ, D. Explainable artificial intelligence in education. *Computers and Education: Artificial Intelligence*, v. 3, p. 100074, 2022. Citado 2 vezes nas páginas 23 e 39.

- KOCSIS, Á.; MOLNÁR, G. Factors influencing academic performance and dropout rates in higher education. *Oxford Review of Education*, v. 51, n. 3, p. 414–432, 2025. Citado 4 vezes nas páginas 15, 19, 20 e 21.
- KOVAČIĆ, Z. J. Early prediction of student success: Mining students' enrolment data. *Journal of Education Technology in Higher Education*, 2010. Citado na página 30.
- KRUEGER, J. G. C.; JR., A. S. B.; BARDDAL, J. P. An explainable machine learning approach for student dropout prediction. *Expert Systems with Applications*, v. 233, p. 120933, 2023. Citado 2 vezes nas páginas 32 e 34.
- LIPTON, Z. C. The mythos of model interpretability. *Queue*, v. 16, n. 3, p. 31–57, 2018. Citado na página 25.
- LUNDBERG, S. M.; LEE, S.-I. A unified approach to interpreting model predictions. In: *Proc. NeurIPS*. [S.l.: s.n.], 2017. Citado na página 26.
- MANHÃES, L. M. B.; CRUZ, S. M. S. da; COSTA, R. J. M.; ZAVALA, J.; ZIMBRÃO, G. Previsão de estudantes com risco de evasão utilizando técnicas de mineração de dados. *Revista Brasileira de Informática na Educação*, 2011. Citado na página 31.
- MARQUES, J.; GONÇALVES, C.; VIEIRA, P. A.; BORGES, A.; DIAS, V.; SANTOS, W.; SILVA, R. Explorando a explicabilidade da inteligência artificial: Técnicas para compreender e interpretar modelos de aprendizado de máquina. In: *Título da coletânea*. [S.l.]: SBC, 2023. p. 25–43. ISBN 9788576695592. Citado na página 41.
- MELO, E.; SILVA, I.; COSTA, D. G.; VIEGAS, C. M. D.; BARROS, T. M. On the use of explainable artificial intelligence to evaluate school dropout. *Education Sciences*, v. 12, n. 12, p. 845, 2022. Citado 3 vezes nas páginas 23, 32 e 34.
- MILLER, T. Explanation in artificial intelligence: Insights from the social sciences. *Artificial Intelligence*, v. 267, p. 1–38, 2019. Citado na página 25.
- MOLNAR, C. *Interpretable machine learning: A guide for making black box models explainable*. 3rd. ed. [S.l.: s.n.], 2025. Citado 2 vezes nas páginas 25 e 28.
- NAGY, M.; MOLONTAY, R. Interpretable dropout prediction: Towards XAI-based personalized intervention. *International Journal of Artificial Intelligence in Education*, v. 34, n. 2, p. 274–300, 2024. Citado 6 vezes nas páginas 15, 16, 26, 27, 32 e 34.
- NEUGEBAUER, M.; BECKER, M.; BIHLER, L.-M.; WAGNER, J. Well-being scarring effects of college non-completion. *Social Science Research*, v. 127, p. 103138, 2025. Citado na página 15.
- NIEROTKA, R. L.; SALATA, A.; MARTINS, M. K. Fatores associados à evasão no ensino superior: Um estudo longitudinal. *Cadernos de Pesquisa*, v. 53, p. e09961, 2023. Citado 2 vezes nas páginas 20 e 48.
- NIYOGISUBIZO, J.; LIAO, L.; NZIYUMVA, E.; MURWANASHYAKA, E.; NSHIMYUMUKIZA, P. C. Predicting student dropout using a two-layer ensemble machine learning approach: A novel stacked generalization. *Computers and Education: Artificial Intelligence*, v. 3, p. 100066, 2022. Citado 2 vezes nas páginas 15 e 21.

- OECD. *Education at a glance 2023: OECD indicators*. [S.l.]: OECD Publishing, 2023. Citado na página 15.
- ORTAGUS, J. C.; TANNER, M.; HU, X.; KRAMER, D. A. Investigating why academically successful community college students leave college without a degree. *AERA Open*, v. 7, p. 1–14, 2021. Citado na página 49.
- PÉREZ, B.; CASTELLANOS, C.; CORREAL, D. Predicting student drop-out rates using data mining techniques: A case study. In: *Proc. IEEE Colombian Conf. Applications in Computational Intelligence*. [S.l.: s.n.], 2018. p. 111–125. Citado na página 15.
- QUIMIZ-MOREIRA, M.; DELGADILLO, R.; PARRAGA-ALAVA, J.; MACULAN, N.; MAURICIO, D. Factors, prediction, explainability, and simulating university dropout through machine learning: A systematic review, 2012–2024. *Computation*, v. 13, n. 8, p. 198, 2025. Citado na página 22.
- RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Why should i trust you? explaining the predictions of any classifier. In: *Proc. KDD*. [S.l.: s.n.], 2016. p. 1135–1144. Citado na página 26.
- RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Anchors: High-precision model-agnostic explanations. In: *Proc. AAAI*. [S.l.: s.n.], 2018. v. 32. Citado na página 27.
- RODRIGUES, H. S.; MARTINS, A. C.; SILVA, J. P.; ROCHA, M. L. Investigating AI algorithms to predict college students' academic performance: A systematic mapping study. In: *Proc. ICAART*. [S.l.: s.n.], 2024. Citado na página 15.
- SAARELA, M.; PODGORELEC, V. Recent applications of explainable AI (XAI): A systematic literature review. *Applied Sciences*, v. 14, n. 19, p. 8884, 2024. Citado 3 vezes nas páginas 26, 40 e 69.
- SACCARO, A.; FRANÇA, M. T. A.; JACINTO, P. d. A. Fatores associados à evasão no ensino superior brasileiro. *Estudos Econômicos*, v. 49, n. 2, p. 337–373, 2019. Citado 2 vezes nas páginas 39 e 44.
- 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, p. e0118432, 2015. Citado na página 41.
- SANTOS, C. O.; PILATTI, L. A.; BONDARIK, R. Evasão no ensino superior brasileiro: conceito, mensuração, causas e consequências. *Debates em Educação*, v. 14, n. 35, p. 294–314, 2022. Citado na página 20.
- SAQR, M.; LÓPEZ-PERNAS, S. Why explainable AI may not be enough: Predictions and mispredictions in decision making in education. *Smart Learning Environments*, v. 11, n. 1, p. 52, 2024. Citado 3 vezes nas páginas 23, 33 e 34.
- SCHRÖER, C.; KRUSE, F.; GÓMEZ, J. M. A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, v. 181, p. 526–534, 2021. Citado na página 35.
- SEMESP, I. *Mapa do ensino superior no Brasil 2025*. 2025. Relatório técnico. Citado na página 15.

SILVA, D. B.; FERRE, A. A. O.; GUIMARÃES, P. S.; LIMA, R.; ESPÍNDOLA, I. B. Evasão no ensino superior público do brasil: estudo de caso da universidade de são paulo. *Avaliação: Revista da Avaliação da Educação Superior*, v. 27, n. 2, p. 248–259, 2022. Citado 2 vezes nas páginas 20 e 49.

SILVA, F. C. da; FEITOSA, R. M.; BATISTA, L. A.; SANTANA, A. M. Análise comparativa de métodos de explicabilidade da inteligência artificial no cenário educacional: Um estudo de caso sobre evasão. In: *Proc. SBIE*. [S.l.: s.n.], 2024. p. 2968–2977. Citado 3 vezes nas páginas 26, 33 e 34.

SILVA, F. da C.; SANTANA, A. M. Explicabilidade dos modelos de aprendizado de máquina no cenário educacional: uma revisão sistemática. *Revista Novas Tecnologias na Educação*, v. 22, n. 1, p. 477–486, 2024. Citado na página 26.

SILVA, L. B. E.; MARIANO, A. S. Definición de evasión y sus implicaciones (límites) para las políticas de educación superior. *Educação em Revista*, v. 37, p. e26524, 2021. Citado na página 19.

SILVA, W. S. da; OLIVEIRA, V. T. de; ARAÚJO, S. S.; VIEIRA, D.; CASTRO, M. F. de. Explainability e auditability: Interpretando e validando modelos de machine learning. In: *Proc. SBSI*. [S.l.: s.n.], 2022. p. 456–467. Citado na página 26.

TEODORO, L. d. A.; KAPPEL, M. A. A. Aplicação de técnicas de aprendizado de máquina para predição de risco de evasão escolar em instituições públicas de ensino superior no brasil. *Revista Brasileira de Informática na Educação*, v. 28, p. 838–863, 2020. Citado na página 22.

TETE, M. F.; SOUSA, M. M.; SANTANA, T. S. de; SILVA, S. F. da. Predictive models for higher education dropout: A systematic literature review. *Education Policy Analysis Archives*, v. 30, n. 61, 2022. Citado na página 15.

TINTO, V. Dropout from higher education: A theoretical synthesis of recent research. *Review of Educational Research*, Sage Publications, v. 45, n. 1, p. 89–125, 1975. Citado na página 20.

ULLOA-CÁZAREZ, R. L.; GARCÍA-DÍAZ, N.; SORIANO-EQUIGUA, L. Multi-layer adaptive fuzzy inference system for predicting student performance in online higher education. *IEEE Latin America Transactions*, v. 19, n. 1, p. 98–106, 2021. Citado na página 15.

Universidade Pública do Brasil no Sul. Pesquisando causas e possíveis soluções para a problemática da evasão em um curso de administração numa universidade. In: *Anais do Colóquio Internacional sobre Gestão Universitária na América do Sul*. Florianópolis, Brasil: [s.n.], 2011. Citado na página 19.