

INSTITUTO FEDERAL DE EDUCAÇÃO CIÊNCIA E TECNOLOGIA DE SANTA  
CATARINA CÂMPUS CAÇADOR

LUIZ ANDRÉ ZEREDO

**IFSC-ALIZA: UMA APLICAÇÃO PARA O ENFRENTAMENTO DA EVASÃO  
ESCOLAR EM INSTITUIÇÕES PÚBLICAS DE ENSINO COM INTELIGÊNCIA  
ARTIFICIAL**

CAÇADOR

2025

LUIZ ANDRÉ ZEREDO

IFSC-ALIZA: Uma aplicação para o enfrentamento da evasão escolar em instituições públicas  
de ensino com inteligência artificial

Monografia apresentada ao Curso de Sistemas de  
Informação do Câmpus Caçador do Instituto Federal de  
Santa Catarina como requisito parcial para a obtenção do  
diploma de Bacharel

Orientador: Prof. Dr. Eduardo Guedes Villar

Coorientador: Prof. Dr. Cristiano Mesquita Garcia

CAÇADOR

2025

Z58i Zeredo, Luiz André.  
IFSC-ALIZA : uma aplicação para o enfrentamento da evasão escolar em instituições públicas de ensino com Inteligência Artificial / Luiz André Zeredo ; orientadores: Eduardo Guedes Villar, Cristiano Mesquita Garcia. -- 2025.  
78 f.

Trabalho de Conclusão de Curso (Graduação)-Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, Caçador, 2025.  
Inclui bibliografias.

1. Predição de evasão. 2. Mineração de Dados Educacionais. 3. Gradient Boosting. 4. Históricos Escolares. I. Villar, Eduardo Guedes. II. Garcia, Cristiano Mesquita.. III. Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina – Graduação em Sistemas de Informação. IV. Título.

CDD 006.3

LUIZ ANDRÉ ZEREDO

IFSC-ALIZA: Uma aplicação para o enfrentamento da evasão escolar em instituições públicas de ensino com inteligência artificial

Este trabalho foi julgado adequado para obtenção parcial do título em Bacharel pelo Instituto Federal de Educação, Ciência e Tecnologia de Santa Catarina, e aprovado na sua forma parcial pela comissão avaliadora abaixo indicada.

Caçador, dia 15, dezembro de 2025.

---

Prof. Eduardo Guedes Villar, Dr.

Orientador

Instituto Federal de Santa Catarina

---

Prof. Cristiano Mesquita Garcia, Dr.

Coorientador

Instituto Federal de Santa Catarina

---

Patrícia Maccarini Moraes, Dr.

Membro

Instituto Federal de Santa Catarina

---

Prof. Adriana Salvador Zanini, Msc

Instituto Federal de Santa Catarina

## **AGRADECIMENTOS**

Dedico um agradecimento especial à minha esposa, Tuanny Caroline Lenz. Seu incentivo foi fundamental para que eu não desistisse nas dificuldades, pela sua paciência com os fins de semana sacrificados e pela força que me deu em todos os momentos de frustração. Este trabalho é a prova do nosso esforço mútuo. Meu diploma é, em grande parte, uma conquista nossa.

Também, meus mais sinceros agradecimentos ao meu orientador, que em todo tempo me incentivou a continuar e não desistir no meio do caminho, sob os mais diversos problemas que enfrentei.



"A sabedoria começa na admissão da própria  
ignorância."

Sócrates

## RESUMO

A evasão no Ensino Superior, particularmente nos Institutos Federais, é um fenômeno complexo e oneroso, representando um desafio para a gestão acadêmica e a alocação de recursos públicos. A identificação precoce de alunos em risco de evasão pode auxiliar à direcionar intervenções pedagógicas. Este trabalho objetiva o desenvolvimento e a validação de um modelo para a predição de evasão no Curso Superior de Sistemas de Informação do IFSC - Câmpus Caçador. A metodologia foi desenhada para superar o desafio dos dados não-estruturados, criando um processo automatizado que "lê" e interpreta históricos escolares em PDF. Este processo extrai não apenas dados brutos, mas cria e dinâmicos indicadores de risco, como o desempenho (notas e frequência) e o número de reprovações dos alunos nos seus últimos três semestres. Múltiplos modelos de *Machine Learning* (aprendizagem de máquina) foram treinados e comparados. O modelo final, um sistema de *Gradient Boosting* (XGBoost) otimizado, demonstrou alta performance: alcançou um *recall* de 84% (capacidade de encontrar os alunos que de fato evadiriam) e uma precisão de 92% (confiabilidade dos alertas gerados), resultando num F1-Score<sup>1</sup> de 0.88. A análise do modelo revelou que ações administrativas ativas, como a quantidade de trancamentos (importância de 0.185), e o comportamento acadêmico recente são fatores preditivos significativamente mais fortes do que métricas consolidadas de longo prazo, como o Coeficiente de Aproveitamento (CAA, importância de 0.054). Concluiu-se que o modelo preditivo é uma ferramenta viável e confiável para o diagnóstico precoce, permitindo uma alocação de recursos de apoio ao discente mais precisa e proativa.

Palavras-Chave: Predição de evasão. Mineração de Dados Educacionais. *Gradient Boosting*. Históricos Escolares.

---

<sup>1</sup> O F1-Score (0.88) é a média harmônica entre a Precisão e o *Recall*, usado para avaliar o equilíbrio do modelo. A Precisão (92%) mede quantos dos alertas de "evasão" estavam corretos, enquanto o *Recall* (84%) mede quantos dos alunos que realmente evadiram foram encontrados pelo modelo.

## ABSTRACT

Student dropout in Higher Education, particularly within Federal Institutes, is a complex and costly phenomenon, representing a challenge for academic management and the allocation of public resources. The early identification of at-risk students can help direct timely pedagogical interventions. This work aims at the development and validation of an Artificial Intelligence (AI) model for dropout prediction at the Superior Course in Information Systems of IFSC – Campus Caçador. The methodology was designed to overcome the challenge of unstructured data by creating an automated process that "reads" and interprets student academic records in PDF format. This process extracts not only raw data but also creates dynamic risk indicators, such as performance (grades and attendance) and the number of failing grades received by students over their last three semesters. Multiple Machine Learning (ML) models were trained and compared. The final model, an optimized Gradient Boosting (XGBoost) system, demonstrated high performance: it achieved a recall of 84% (its ability to find students who would drop out) and a precision of 92% (the reliability of the alerts generated), resulting in an F1-Score<sup>2</sup> of 0.88. Model analysis revealed that active administrative actions, such as the number of course withdrawals (feature importance of 0.185), and recent academic behavior are significantly stronger predictive factors than long-term consolidated metrics, such as the Academic Achievement Coefficient (CAA, importance of 0.054). We conclude that the predictive model is a viable and reliable tool for early diagnosis, enabling a more accurate and proactive allocation of student support resources.

**Keywords:** Dropout Prediction. Educational Data Mining. Gradient Boosting. Academic Records.

---

<sup>2</sup> The F1-Score (0.88) is the harmonic mean between Precision and Recall, used to assess the model's balance. Precision (92%) measures how many of the "dropout" alerts were correct, while Recall (84%) measures how many of the students who actually dropped out were found by the model.

## LISTA DE ILUSTRAÇÕES

<b>Figura 1 – Categorias de fatores de evasão.....</b>	<b>20</b>
<b>Figura 2 – Matriz de confusão (XGBoost otimizado) .....</b>	<b>44</b>
<b>Figura 3 – Importância das Features (Random Forest) .....</b>	<b>45</b>
<b>Figura 4 – Importância das Features (XGBoost) .....</b>	<b>54</b>
<b>Figura 5 – Comparação da Distribuição do CAA entre Grupos .....</b>	<b>55</b>

## LISTA DE TABELAS

<b>Tabela 1 – Eixos do plano estratégico de Acesso, Permanência e Êxito dos estudantes do IFSC .....</b>	<b>29</b>
<b>Tabela 2 – Distribuição e balanceamento de variável alvo .....</b>	<b>35</b>
<b>Tabela 3 – Variáveis Coletadas .....</b>	<b>42</b>
<b>Tabela 4 – Relatório e Classificação (XGBoost Otimizado) .....</b>	<b>43</b>

## LISTA DE ABREVIATURAS E SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
API	Interface de Programação de Aplicações
APLICAÇÃO WEB	Aplicação inserida de forma <i>on-line</i>
CAA	Coefficiente de Aproveitamento Acadêmico
CRISP-DM	<i>Cross-Industry Standard Process for Data Mining</i> (Processo Padrão Interindústrias para Mineração de Dados)
EAD	Educação à Distância
IA	Inteligência Artificial
IBGE	Instituto Brasileiro de Geografia e Estatística
IFSC	Instituto Federal de Santa Catarina
LLM	<i>Large Language Model</i>
MACHINE LEARNING	Aprendizado de máquina para execução posterior de demais atividades
PDF	<i>Portable Document Format</i> (Formato de Documento Portátil)
PEAPEE	Plano Estratégico de Acesso, Permanência e Êxito dos Estudantes
XGBOOST	<i>eXtreme Gradient Boosting</i>

## SUMÁRIO

1	INTRODUÇÃO .....	22
1.1	OBJETIVOS.....	23
1.1.1	Objetivo geral .....	23
1.1.2	Objetivo específico .....	24
1.2	JUSTIFICATIVA.....	24
2	FUNDAMENTAÇÃO TEÓRICA.....	26
2.1	EVASÃO ESCOLAR.....	26
2.1.1	Causas e possibilidades para a superação da Evasão .....	27
2.1.2	Consequências e efeitos da evasão.....	31
2.2	PERMANÊNCIA E ÊXITO .....	35
2.2.1	Perspectiva Tradicional e Crítica sobre a Permanência e o Êxito .....	36
2.2.2	Importância e efeitos de Políticas e Medidas de Permanência e Êxito .....	37
2.3	INTELIGÊNCIA ARTIFICIAL.....	38
2.4	INTELIGÊNCIA ARTIFICIAL E A EDUCAÇÃO.....	39
2.5	MINERAÇÃO DE DADOS EDUCACIONAIS (MDE).....	40
2.5.1	O Processo CRISP-DM na Mineração de Dados Educacionais.....	41
2.6	MODELOS DE <i>MACHINE LEARNING</i> PARA CLASSIFICAÇÃO .....	42
3	METODOLOGIA DA PESQUISA .....	44
3.1	ENTENDIMENTO DO PROBLEMA E DOS DADOS .....	44
3.2	PREPARAÇÃO DOS DADOS .....	44
3.2.1	Extração de Dados - o <i>Pipeline</i> de PDF .....	44
3.2.2	Engenharia de <i>Features</i> .....	45
3.2.3	Limpeza e Definição da Variável-Alvo.....	45
3.2.4	Transformação e Escalonamento.....	46
3.3	MÉTRICAS DE AVALIAÇÃO PARA DADOS DESBALANCEADOS .....	46
3.4	MODELAGEM .....	47
3.4.1	Divisão e amostragem dos dados .....	47
3.4.2	Comparação de algoritmos .....	48
3.4.3	Otimização de hiperparâmetros .....	48
3.5	AVALIAÇÃO .....	48
3.6	AMBIENTE E FERRAMENTAS .....	48
3.7	LÓCUS DA PESQUISA .....	49
3.7.1	Participantes .....	49
3.7.2	Critérios de inclusão .....	49
3.7.3	Critérios de exclusão .....	50
3.7.4	Benefícios da pesquisa para os sujeitos/comunidade.....	50
3.8	INSTRUMENTOS DE COLETA DE DADOS .....	50
4	RESULTADOS E DISCUSSÃO .....	52
4.1	PERFORMANCE PREDITIVA DO MODELO ( <i>XGBOOST</i> OTIMIZADO) .....	52
4.2	ANÁLISE DE IMPORTÂNCIA DAS <i>FEATURES</i> .....	54
4.3	DISCUSSÃO DOS RESULTADOS.....	56
4.3.1	O modelo preditivo é viável e confiável.....	56
4.3.2	Contribuições e repercussões do modelo preditivo.....	57
5	CONSIDERAÇÕES FINAIS .....	59
	REFERÊNCIAS.....	62

APÉNDICE .....	70
----------------	----



## 1. INTRODUÇÃO

A evasão escolar é uma realidade preocupante e presente em diversas instituições de ensino no Brasil. Essa questão tem sido debatida e analisada em diversos estudos e pesquisas nacionais, que buscam compreender as múltiplas razões que levam os alunos a abandonarem a escola antes de concluir suas etapas educacionais. De acordo com o IBGE, “no grupo etário de 14 a 29 anos, 9,0 milhões de pessoas não completaram o ensino médio, seja por terem abandonado a escola antes do término desta etapa ou por nunca a terem frequentado” (IBGE, 2023). Esse dado revela uma preocupação com a formação educacional dos jovens brasileiros, que ainda enfrentam obstáculos significativos para garantir um aprendizado contínuo e de qualidade (IBGE, 2023). Além disso, entre os resultados obtidos neste levantamento do IBGE (2023), sobressaem-se alguns fatores significativos para a explicação do abandono, como: a dificuldade nas matérias, o desejo por uma escola diferente, e a relevância dada à escolha da escola. Em adição, segundo estudo realizado em Minas Gerais, o aluno que apresenta dificuldades em acompanhar todo o conjunto de disciplinas, seguido por dificuldades em acompanhar ciências exatas, eram motivadores para a evasão escolar (Soares et al; 2015).

Para que uma instituição de ensino se mantenha competitiva e eficaz, é necessário que exista uma alta taxa de permanência e êxito dos estudantes. Contudo, muitos cursos, especialmente aqueles voltados para áreas tecnológicas e de desenvolvimento profissional, têm historicamente apresentado taxas elevadas de evasão. Este índice reflete a desistência de uma parte significativa dos estudantes antes de concluírem sua formação, o que acarreta perdas sociais e financeiras para a coletividade. Como exemplo deste desafio, em 2023, o IFSC Câmpus Caçador registrou uma taxa de evasão de 23,73% entre seus setecentos e cinquenta alunos matriculados (IFSC, 2023). Este índice, embora contextualizado por fatores internos e externos, demonstra a dimensão do problema enfrentado pela unidade local e a urgência de ferramentas preditivas.

As razões para o abandono no Ensino Superior são complexas e multifatoriais, sendo exaustivamente discutidas na literatura de Mineração de Dados Educacionais (MDE). Entre as motivações mais apontadas em estudos nacionais (Ramos; Gonçalves Junior, 2020) estão a falta de suporte pedagógico adequado, problemas financeiros que dificultam a continuidade dos estudos, e a necessidade de o aluno conciliar os estudos com a jornada de trabalho. Este último fator é especialmente determinante em cursos noturnos ou de tecnologia (Neri et al., 2009), onde o conflito entre o tempo dedicado ao emprego e o exigido pelo curso acaba por minar a permanência do estudante. O caso dos Institutos Federais (IFs) é ainda mais crítico, visto que

se trata de uma política pública, custeada com recursos públicos, o que torna a evasão uma questão de responsabilidade fiscal e social.

A evasão e a permanência nos Institutos Federais são um foco de estudo aprofundado no contexto catarinense, sendo analisadas sob a ótica dos próprios trabalhadores da educação do IFSC (Zanin; Garcia, 2023). Em paralelo, a aplicação de técnicas de Mineração de Dados Educacionais (MDE) para prever a evasão já foi validada como uma abordagem viável em instituições federais pares, como demonstrado por Silva (2018) na Universidade Federal de Santa Catarina (UFSC). Estes trabalhos, em conjunto, demonstram a relevância de aplicar técnicas de Inteligência Artificial (IA) ao contexto acadêmico local, o que justifica a necessidade de ferramentas preditivas para auxiliar o corpo pedagógico.

Contudo, muitas abordagens de predição focam em métricas de longo prazo, como o Coeficiente de Aproveitamento Acadêmico (CAA), que é um indicador "lento", refletindo o histórico total do aluno, mas não necessariamente o seu risco imediato. A literatura, todavia, aponta para a necessidade de priorizar indicadores dinâmicos de curto prazo (Oliveira et al., 2019). Este estudo, em consonância com esta abordagem, parte da premissa de que o comportamento acadêmico recente (como a queda no desempenho e na frequência nos últimos semestres) e, principalmente, ações administrativas ativas (como o pedido de trancamento de matrícula) são preditores mais fortes e ágeis da evasão iminente do que o histórico consolidado.

Diante do exposto, o objetivo geral deste trabalho é desenvolver e validar um modelo de *Machine Learning*, denominado IFSC-ALIZA, capaz de prever o risco de evasão de discentes do Curso de Sistemas de Informação do IFSC-Caçador, focando na análise de indicadores de comportamento recente.

## 1.1 OBJETIVOS

O presente trabalho possui determinados objetivos, dos quais lista-se a seguir:

### 1.1.1 Objetivo geral

Desenvolver um modelo preditivo, baseado em *Machine Learning* (Gradient Boosting), para identificar riscos de evasão escolar no IFSC – Campus Caçador.

### 1.1.2 Objetivo específico

- a) Desenvolver um *pipeline* para extração e processamento de dados não-estruturados;
- b) Realizar a engenharia de *features* (criação de indicadores) baseada no desempenho recente;
- c) Projetar e implementar um modelo de aprendizado de máquina com capacidade de identificar padrões de risco de evasão;
- d) Avaliar o potencial da aplicação como ferramenta de apoio à gestão acadêmica e modelo replicável.

## 1.2 JUSTIFICATIVA

A evasão no Ensino Superior público transcende o impacto individual na carreira do discente, representando um desafio para a sociedade (Ramos; Gonçalves Júnior, 2020). Cada aluno que evade representa um desperdício de investimento público, visto que a vaga foi financiada pelo contribuinte, e uma perda de capital humano qualificado para o país. Além disso, a evasão está frequentemente ligada a frustrações de projeto de vida e à perpetuação de desigualdades sociais. Portanto, a criação de mecanismos que auxiliem na permanência e êxito dos estudantes consiste em uma pauta socialmente relevante e urgente, além de ser um tema de interesse e atenção contínua no próprio IFSC (Zanin; Garcia, 2023).

No âmbito institucional, a gestão pedagógica do IFSC - Câmpus Caçador enfrenta o desafio de identificar o risco de evasão. Pesquisas em Mineração de Dados Educacionais (MDE) (Silva, 2018) apontam que o uso de técnicas preditivas é uma abordagem eficaz para este fim (Bezerra; Barros, 2017). Todavia, as métricas tradicionais, como o Coeficiente de Aproveitamento Acadêmico (CAA), são frequentemente indicadores retroativos, que refletem o desempenho passado, mas não necessariamente o risco imediato de evasão. A principal justificativa prática deste trabalho reside na possibilidade de fornecer uma ferramenta de diagnóstico. Ou seja, ao focar em indicadores de comportamento recente (como médias das notas e percentual de faltas), o modelo de aprendizagem de máquina proposto permite que a coordenadoria identifique o aluno no exato momento em que o seu padrão de desempenho muda, o que possibilita uma intervenção pedagógica antes que o processo de evasão se torne irreversível.

Do ponto de vista técnico e metodológico, este trabalho se justifica por sua abordagem de um problema comum, mas raramente solucionado: a inacessibilidade de dados acadêmicos

críticos. A coleta e preparação de dados é frequentemente citada como um dos maiores desafios na aplicação da MDE no contexto brasileiro (Figueiredo; Paula, 2022). Em muitas realidades institucionais, os dados mais ricos sobre o desempenho do aluno estão "presos" em formatos não-estruturados, como relatórios em PDF. Uma contribuição central desta pesquisa é, portanto, o desenvolvimento de um *pipeline* de extração automatizado (detalhado no Capítulo 3) capaz de ler e interpretar históricos escolares de estudantes. Esta solução viabiliza o modelo preditivo e também oferece uma metodologia replicável para outros cursos ou Campus do IFSC que enfrentam o mesmo desafio, preenchendo uma lacuna entre a teoria da MDE, já aplicada em contextos de dados estruturados (Silva, 2018), e a prática da gestão acadêmica cotidiana.

## 2. FUNDAMENTAÇÃO TEÓRICA

Nesta seção, serão apresentados os três principais elementos teóricos que fundamentam este projeto de trabalho de conclusão de curso, juntamente com suas respectivas contextualizações e explicações sobre o tema.

### 2.1 EVASÃO ESCOLAR

Para Freire (1987), a educação consiste em um instrumento de transformação social, voltado à conscientização crítica dos indivíduos, capacitando-os a interpretar e transformar a realidade que os cerca. Em sua perspectiva, o autor sustenta que a educação não deve ser vista apenas como um meio para atingir objetivos econômicos, mas como um caminho para a emancipação e liberdade capaz de promover uma sociedade mais democrática e participativa.

Entretanto, apesar desse ideal transformador, a evasão escolar permanece como um fenômeno persistente e multifacetado. De acordo com Ceratti (2008), a evasão possui como causas fatores cognitivos, psicoemocionais, socioculturais, institucionais, econômicos e políticos. Mesmo diante dos esforços de professores e equipes pedagógicas para oferecer uma educação mais inclusiva e significativa, ainda prevalece um sistema educativo que reproduz práticas tradicionais, limitadas em sua capacidade de romper com a exclusão social e o fracasso escolar.

A presente pesquisa evidencia que ainda estão presentes crenças que naturalizam a evasão, tratando-a como um processo espontâneo ou individualizado, por vezes desconsiderando-se sua construção histórica e estrutural. O fracasso escolar, nesse contexto, costuma ser atribuído aos próprios estudantes, negligenciando os condicionantes institucionais e sociais que o provocam (Oliveira, 2018). Neste sentido, enfrentar esse desafio requer o fortalecimento da autonomia pedagógica, a valorização de práticas educativas contextualizadas e o reconhecimento do papel da escola como agente de transformação social. A formação dos estudantes deve articular teoria e prática, e promover a conexão entre o conhecimento sistematizado e a realidade vivida pelos alunos, de modo a prepará-los para lidar com as adversidades econômicas e sociais.

Além disso, torna-se fundamental um Projeto Político Pedagógico que vá além do discurso teórico, e articule intenções pedagógicas a práticas concretas. Compreender a evasão escolar em sua complexidade, à luz de estudos empíricos e referenciais teóricos, pode mitigar seus impactos, que vão desde o desamparo social até a perpetuação das desigualdades (Ceratti,

2008).

No contexto do ensino médio, por exemplo, dados do Instituto Mobilidade e Desenvolvimento Social (IMDS, 2021) revelam que o tempo de permanência na escola está fortemente relacionado à escolaridade dos pais. Enquanto 95% dos filhos de pais com ensino superior completo concluem o ensino médio, apenas 25,9% dos filhos de pais sem escolaridade atingem essa mesma etapa. Ainda que essa disparidade tenha diminuído ao longo das gerações, reflexo da redução das taxas de evasão, ela persiste como um marcador de desigualdade.

Este cenário de desigualdade e a importância crescente da qualificação profissional são reflexos de profundas mudanças estruturais nas últimas décadas. A partir de 1996, com a implementação da Lei de Diretrizes e Bases da Educação Nacional (LDB), houve uma valorização crescente da educação básica como pré-requisito para o Ensino Superior. Posteriormente, a década de 2000 foi marcada por políticas públicas de ampliação do acesso, como o Programa Universidade para Todos (ProUni) e o Fundo de Financiamento Estudantil (FIES). Tais políticas fortaleceram a ideia do ensino superior como essencial para a mobilidade social e para o ingresso em ocupações formais e melhor remuneradas (Almeida, 2020). Contudo, essa expansão do acesso trouxe novos desafios para a permanência, pois um perfil de aluno mais heterogêneo passou a integrar as instituições, demandando maior suporte pedagógico e financeiro.

Na década seguinte, a expansão das vagas no ensino superior e as novas exigências do mercado reforçaram a centralidade do ensino médio na trajetória educacional e profissional dos jovens. Tornou-se comum que empresas exigissem, no mínimo, o diploma dessa etapa como critério de contratação. Dessa forma, o ensino médio passou a ser reconhecido como etapa essencial não apenas para o desenvolvimento pessoal e profissional, mas também para a ascensão social, especialmente a partir do processo de modernização econômica e da crescente valorização da educação básica nas últimas décadas do século XX e início do século XXI (Oliveira, 2018).

### **2.1.1 Causas e possibilidades para a superação da Evasão**

A literatura, colocada por Coelho e Garcia (2013) que aborda as causas da evasão no ensino superior identifica várias tipologias, que podem ser classificadas em três grandes categorias: fatores individuais, institucionais e sociais.



Figura 1 – Categorias de fatores de Evasão.

Fonte: Fonte: Elaborado pelo autor (2024), baseado em Tinto (1975); Ramos e Gonçalves Junior (2020); e Zanin e Garcia (2023).

Ramos e Gonçalves Junior (2020) apresentaram uma pesquisa que aborda o direito à educação e diferencia o abandono escolar — entendido como uma saída temporária — da evasão escolar, caracterizada pela desistência definitiva. Entre os fatores que contribuem para o fenômeno geral da interrupção dos estudos, destacam-se a desigualdade social, o atraso escolar e as fragilidades no acesso a serviços de apoio. Os autores defendem a importância de políticas públicas intersetoriais para o enfrentamento dessas vulnerabilidades, com ênfase na valorização da participação de jovens e de suas famílias.

Contudo, é fundamental ressaltar que outras vertentes teóricas problematizam essa distinção. Enquanto a evasão é frequentemente percebida pelo senso comum como um ato "solitário" de desistência do estudante, análises mais críticas, como as de Zanin e Garcia (2023) no contexto dos Institutos Federais, sugerem que o abandono também implica uma responsabilidade do sistema de ensino. Nesta perspectiva, a saída do aluno não é apenas uma escolha individual ou temporária, mas muitas vezes o resultado de processos de exclusão

institucional e da incapacidade da escola em garantir as condições de permanência.

O estudo de 2020 foi realizado com base na percepção de agentes públicos e beneficiários de políticas educacionais em um município paulista, o que revelou as dificuldades enfrentadas por alunos em situação de vulnerabilidade social e, conseqüentemente, a ampliação do risco de abandono escolar. Ramos e Gonçalves Junior (2020) ainda analisam as percepções de educadores, conselheiros tutelares e técnicos da rede de apoio, evidenciando divergências nas causas atribuídas ao problema da evasão, entre fatores individuais, familiares, escolares e socioeconômicos. A pesquisa reforça a necessidade de diálogo entre setores e ações coordenadas, reconhecendo que a interrupção da trajetória escolar transcende o campo educacional e requer práticas integradas e uma compreensão abrangente das diferentes perspectivas envolvidas.

Ferreira (2022) corrobora essa visão ao afirmar que a ausência de políticas públicas eficazes voltadas às necessidades dos estudantes de classes populares contribui significativamente para o abandono escolar. Tais alunos enfrentam desafios adicionais relacionados à adaptação acadêmica, à conciliação entre trabalho e estudo e à carência de suporte institucional. De forma complementar, Anzolin e Kreling (2012), ao analisarem as causas da evasão escolar em cursos de aprendizagem industrial do SENAI/SC, destacaram que a migração para cursos do PRONATEC — incentivada pelo apoio financeiro e pela oferta de diploma técnico — e a necessidade de trabalhar foram fatores determinantes para o abandono e evasão escolar na região. O estudo identificou que a unidade com maior evasão não havia implementado práticas preventivas, ao contrário da unidade comparativa, que adotava o acompanhamento sistemático da frequência estudantil. Com base nesses resultados, os autores recomendam o desenvolvimento de planos de ação com foco em gestão, monitoramento individualizado e estratégias de prevenção.

Além disso, Ramos e Gonçalves (2024) e Leal (2015) destacam a saúde mental como um fator chave para a permanência estudantil. Neste sentido, o estresse acadêmico, a pressão por desempenho e a ausência de suporte psicológico são apontados como elementos que contribuem para o abandono. Instituições que oferecem atendimento psicológico e promovem o bem-estar emocional tendem a apresentar menores taxas de evasão.

Outro aspecto relevante se refere à adequação dos currículos às expectativas dos estudantes e às demandas do mercado. Ferreira (2022) observa que cursos pouco atrativos ou desconectados da realidade laboral tendem a gerar desinteresse, o que pode resultar em evasão. Nesse sentido, é fundamental que as instituições revejam e atualizem constantemente seus programas, a fim de garantir que estejam alinhados às necessidades dos alunos e do mundo do

trabalho.

Zanin e Garcia (2023), ao investigarem a evasão na educação profissional e tecnológica, identificaram que o fenômeno é atravessado por dimensões institucionais, pedagógicas, sociais e subjetivas. A partir da percepção de trabalhadores da educação do IFSC, os autores revelaram olhares distintos e, por vezes, conflitantes quanto às causas do abandono. Enquanto alguns profissionais atribuem a evasão a atores externos à instituição de ensino, como as condições socioeconômicas dos estudantes, outros reconhecem fragilidades nas práticas pedagógicas e na organização institucional. Zanin e Garcia (2023) ainda apontam para a necessidade de políticas de permanência mais integradas, que considerem a escuta qualificada dos estudantes e o fortalecimento dos vínculos com a comunidade, promovendo práticas educativas mais inclusivas e dialógicas.

Em complemento, Cavalcante (2010) discute como as desigualdades sociais e econômicas impactam negativamente o acesso e a permanência no ensino superior. A autora destaca que estudantes de classes sociais mais baixas enfrentam maiores dificuldades para permanecer na universidade, especialmente em virtude da necessidade de conciliar trabalho e estudos, além dos custos da educação superior. Nesse mesmo campo, são identificadas causas externas à universidade, como a atração por carreiras mais valorizadas socialmente, a falta de perspectivas profissionais na área da docência, dificuldades financeiras, distância física da instituição e o nível de escolaridade da família, que influencia diretamente a expectativa de permanência e sucesso acadêmico (Cavalcante, 2010).

Internamente, também se destacam fatores como a escolha equivocada de curso — muitas vezes motivada por falta de informações adequadas —, problemas pedagógicos, como métodos desatualizados e docentes despreparados, e a inadequação dos cursos às exigências do mercado de trabalho. Tais fatores reforçam a necessidade de um acompanhamento mais eficaz da trajetória discente desde o ingresso (Cruz; Feitosa, 2020).

Gomes (1998) critica a redução da evasão à dimensão econômica, argumentando que muitas análises, influenciadas por uma lógica neoliberal, desconsideram os fatores sociais e culturais que influenciam a permanência. Em consonância com essa perspectiva, De Mello e Dos Santos (2012), ao investigarem a evasão no curso de Administração de uma universidade pública no Sul do Brasil, identificaram 28 casos de abandono. As autoras ressaltam que o sentimento de pertencimento possibilita prevenir a evasão, pois estudantes que não se sentem acolhidos tendem a internalizar o fracasso e se afastar silenciosamente. O estudo reforça que a valorização do aluno enquanto sujeito, e não apenas como número ou profissional em formação, é determinante para sua permanência e sucesso acadêmico.

### 2.1.2 Consequências e efeitos da evasão

A evasão no ensino superior no Brasil ocorre por diversos fatores, como abandono, resistência formal, exclusão institucional e transferência (Cunha; Cunha; De Lima, 2020). O primeiro ano letivo costuma ser o período mais crítico, marcado por dificuldades acadêmicas, financeiras e pessoais. Apenas uma parcela reduzida dos estudantes conclui a graduação no tempo previsto. Variáveis como gênero, escolaridade familiar, conciliação entre trabalho e estudo, ausência de apoio emocional e influência pedagógica impactam significativamente na permanência dos discentes em seus cursos (Santos; Andriola, 2019).

Segundo Oliveira (2011), os efeitos da evasão sobre os estudantes manifestam-se de maneira profunda, e podem refletir-se em frustração pessoal, baixa autoestima e dificuldades de reinserção no mercado de trabalho. Além disso, a decisão de abandonar os estudos pode estar associada a pressões familiares, sobretudo em contextos de vulnerabilidade social, nos quais a sobrevivência econômica imediata sobrepõe-se à continuidade educacional (Neri et al., 2009).

Porém, os impactos da evasão extrapolam a esfera individual e afetam também as instituições de ensino. Zanin e Garcia (2023) ressaltam que o abandono escolar compromete a missão social dos Institutos Federais, voltada à democratização do acesso e à promoção da equidade. Segundo esses autores, o fenômeno da evasão fragiliza indicadores institucionais e revela a necessidade de políticas de permanência mais efetivas e contextualizadas. Ainda conforme os autores, a evasão repercute em toda a comunidade escolar, o que exige uma reavaliação das práticas pedagógicas e do compromisso institucional com a permanência e o êxito dos estudantes.

Complementando essa discussão, Ramos e Gonçalves Junior (2020) analisam a evasão a partir da perspectiva dos próprios estudantes evadidos do IFSP. Os autores identificam que, além de questões socioeconômicas, fatores como desmotivação com o curso, ausência de acolhimento institucional, fragilidade nos vínculos interpessoais e dificuldades de adaptação ao ambiente escolar contribuem diretamente para o desligamento. A pesquisa revela que muitos alunos não se sentem pertencentes ao espaço educacional, o que reforça a importância de estratégias de escuta ativa e apoio pedagógico contínuo. Assim, a evasão é entendida como um fenômeno multifatorial e complexo, exigindo abordagens que transcendam análises meramente quantitativas.

No âmbito institucional, a evasão prejudica o orçamento, compromete a qualidade do ensino e afeta a organização das universidades. Para a sociedade, os prejuízos incluem a perda

de talentos e capital humano, o que limita o desenvolvimento social e econômico. Diante dessa problemática, enfatiza-se a necessidade de uma abordagem colaborativa entre as instituições e os estudantes para enfrentar a evasão. As universidades devem desenvolver estratégias de acompanhamento contínuo que promovam a permanência e a conclusão dos estudos, enquanto os estudantes precisam assumir um compromisso com suas escolhas e esforços acadêmicos (Ramos; Gonçalves Junior, 2020).

A evasão, por sua vez, é apresentada como um fenômeno dinâmico e complexo, que demanda atenção específica e constante. Destaca, ainda, que o fenômeno da evasão varia conforme o contexto, sugerindo a realização de estudos específicos, como nos centros de educação e ciências médicas, devido às peculiaridades de cada área acadêmica. Por fim, reafirma-se que o combate à evasão exige atenção contínua por parte da gestão universitária, a qual necessita prevenir abandonos e garantir o aproveitamento do potencial dos estudantes (Fialho, 2014).

A evasão no ensino superior está frequentemente relacionada à insatisfação dos alunos com a instituição, o curso e a profissão escolhida. No caso do curso de Educação Física da UFPI, Silva et. al. (2012) identificaram que um dos principais motivos de desistência do curso consiste na desinformação sobre a profissão e as perspectivas profissionais, além do descontentamento com o *status* e a remuneração esperados. A desvalorização histórica das licenciaturas, especialmente na área de Educação Física, faz com que muitos alunos abandonem o curso em busca de áreas com maior prestígio social e financeiro, como fisioterapia, medicina ou direito. Além dos fatores econômicos e sociais, o preconceito histórico contra profissões como a educação e a educação física contribui para a baixa autoestima e a falta de reconhecimento desses profissionais. Por fim, esses autores ressaltam que a desvalorização do trabalho docente é reflexo das condições do sistema educacional, e a qualidade do ensino só poderá ser melhorada se houver uma valorização da profissão, com políticas públicas que garantam uma formação adequada, melhores condições de trabalho, salários justos e incentivos à carreira dos professores (Silva et al., 2012).

De Mello e Dos Santos (2012) afirmam que a resposta a essa situação exige ações mais proativas nos projetos pedagógicos, com ênfase em diálogo, acompanhamento psicopedagógico e um apoio mais contínuo aos estudantes. Como exemplos de potenciais formas de minimizar a evasão no contexto em estudo, são programas de bolsas de permanência, como bolsas de pesquisa e auxílio para alimentação e transporte, os quais podem ser eficazes, mas não são suficientes isoladamente. Para esses autores, a universidade necessita investir em capacitação docente e na implementação de políticas de acompanhamento mais próximas dos estudantes. A

educação, conforme destaca Paulo Freire, precisa ser entendida como um processo transformador, em que a sociedade muda com a educação, e a educação também muda a sociedade (De Mello; Dos Santos, 2012).

Alencar (2020) conduz um estudo com o objetivo entender o fenômeno da evasão sob a ótica dos próprios alunos evadidos, identificando suas características e os motivos do abandono escolar realizado no Curso de Bacharelado em Agronomia da Universidade Federal Rural de Pernambuco. Ao abranger uma análise longitudinal de 2014 a 2018, o autor revelou uma taxa média de evasão de 10,62%, inferior à média institucional da universidade. Segundo Alencar (2020) o perfil dos alunos evadidos inclui majoritariamente homens, com idade média de 23 anos, de origem não branca, solteiros, sem vínculo empregatício, e de famílias com renda de até três salários-mínimos. O autor evidenciou que a evasão ocorre principalmente nos dois primeiros períodos do curso, com desvinculação (47,2%) e desistência (31,9%) sendo as formas mais frequentes de abandono. Em complemento, as entrevistas realizadas por Alencar (2020) apontaram que, no contexto em estudo, os principais motivadores para a evasão incluem baixo envolvimento acadêmico, dificuldades em disciplinas desafiadoras como matemática, problemas de relacionamento com docentes e a falta de orientação vocacional. Esses fatores, somados à falta de identidade com o curso e desmotivação, resultam em um comprometimento reduzido com a conclusão do curso. Alencar (2020) concluiu que esses achados podem ajudar na formulação de estratégias para reduzir a evasão e melhorar a qualidade da educação pública, aumentando os índices de diplomação.

Em relação à pesquisas sobre evasão realizadas na região Sul, especificamente no estado de Santa Catarina, Prim e Fávero (2013) estudaram a evasão escolar no ensino superior em uma faculdade de Blumenau, identificando o perfil dos alunos evadidos e os motivos que influenciaram essa decisão. A taxa de evasão da instituição foi de 12,32% em 2012, significativamente abaixo da média nacional de 22%. No estudo, o perfil dos alunos evadidos é predominantemente masculino, com idade entre 18 e 31 anos, e ingressantes por meio do vestibular. Esses autores evidenciaram que a maior parte da evasão ocorreu nos três primeiros semestres do curso, representando 80,95% dos casos. Os principais motivos para o abandono escolar foram a reprovação (23,81%), problemas financeiros (9,52%), inadimplência (7,14%) e mudança de curso (7,14%), totalizando 47,61%. Os motivos de evasão foram analisados sob a ótica de causas psicológicas, econômicas e sociológicas. Para reduzir a evasão, sugere-se maior atenção às dificuldades de aprendizagem no início do curso, com reforço de atividades complementares. Além disso, recomenda-se maior divulgação de benefícios financeiros para os alunos, a fim de minimizar o impacto da inadimplência. A pesquisa também destaca a

necessidade de testes vocacionais e orientação acadêmica para ajudar os alunos a escolherem o curso adequado, o que poderia prevenir o abandono precoce. Como limitação do estudo, foi mencionada a falta de dados comparativos de períodos anteriores, sugerindo que pesquisas futuras entrevistassem alunos evadidos e investigassem o papel dos coordenadores de curso na retenção dos alunos (Prim; Fávero, 2013).

Uma outra pesquisa na região, realizada por Davok e Bernard (2016), teve em um estudo retrospectivo nos anos de 2008 a 2010, e analisou os índices de evasão nos cursos de graduação da UDESC, identificando tendências comuns na literatura, como os altos índices de evasão nos cursos de licenciatura e nas áreas de Ciências Exatas e da Terra, em contraste com os cursos da área de Ciências da Saúde, que apresentaram menores índices. Em análise geral, Davok e Bernard (2016) identificaram que os cursos da área de Ciências da Saúde foram os que apresentaram os menores índices médios de evasão, enquanto os cursos das áreas de Ciências Exatas e da Terra, Engenharia e Linguística, Letras e Artes ficaram acima da média de evasão da instituição, que foi de 38,2%. Além de avaliar os índices de evasão, o estudo contribuiu para o planejamento institucional e a formulação de políticas de permanência, ao situar o problema da evasão dentro de um corte temporal e incentivar o debate para entender suas causas. O estudo também destacou a importância da análise contextualizada dos dados, que poderia ser realizada pelos Núcleos Docentes Estruturantes (NDEs) dos cursos. Para os autores, os índices de evasão apresentados devem ser vistos como um primeiro passo para entender os fatores que influenciam a permanência dos alunos, sendo estes fatores tanto internos às instituições quanto externos, como questões econômicas, sociais e culturais. Davok e Bernard (2016), ainda ressaltaram que há a necessidade de um debate acadêmico aprofundado sobre o tema, sugerindo o uso de teorias socioculturais e econômicas para ampliar a compreensão da evasão. Embora o estudo tenha sistematizado informações sobre o perfil dos evadidos e aberto caminhos para novas pesquisas, algumas questões ainda carecem de explicações mais detalhadas. Uma medida sugerida para reduzir a evasão é a criação de um serviço de orientação ao estudante, que poderia oferecer apoio e informações aos alunos com dificuldades de adaptação ao ambiente universitário (Davok; Bernard, 2016).

Em complemento, Momm, Momm e Souza (2020) sugeriram que atividades de integração sejam realizadas desde as primeiras semanas de aula, para fortalecer os laços entre os alunos e possibilitar que os colegas identifiquem mudanças no comportamento de alguém que enfrente dificuldades de adaptação ao curso. Segundo os autores, isso permitiria uma intervenção direta ou a comunicação da preocupação ao departamento pedagógico, o que poderia auxiliar na continuidade do aluno no curso. Momm, Momm e Souza (2020) apontaram

também a importância de criar um grupo de trabalho para levantar informações junto aos egressos e às empresas locais, com o intuito de ajustar o curso às demandas profissionais do mercado. Para os autores, essa ação estreitaria os laços entre as empresas e o campus, o que pode aumentar a procura pelos egressos e incentivar os colaboradores das empresas a se qualificarem na instituição. Para alunos que enfrentam dificuldades em conciliar trabalho e estudo, Momm, Momm e Souza, (2020) sugerem ações como a oferta de unidades curriculares em regimes diferenciados, como períodos de férias ou cursos de verão, o que poderia ajudar a diminuir a evasão.

Em relação à reprovação nas disciplinas iniciais, Momm, Momm e Souza (2020) sugerem a inclusão de uma unidade curricular de matemática básica, que revisasse ou apresentasse conceitos fundamentais, uma vez que muitos alunos chegam ao curso com lacunas significativas nesse conteúdo. A criação de programas de monitoria também foi apontada como uma maneira de ajudar os estudantes à superarem dificuldades específicas. Esses autores ainda destacaram o desconhecimento de muitos estudantes sobre os auxílios financeiros disponíveis. Propôs-se, então, uma apresentação clara e objetiva dessas opções logo no início do curso, para que os estudantes possam solicitar o apoio necessário no momento certo. Finalmente, para evitar desistências nas fases mais avançadas do curso, foi sugerido que a equipe pedagógica e os coordenadores elaborassem atividades de divulgação nas escolas locais, informando aos estudantes sobre os cursos oferecidos e as áreas de atuação profissional. Essa divulgação poderia influenciar positivamente a escolha do curso e reduzir as taxas de evasão. Embora o estudo seja inicial, as propostas apresentadas buscam reduzir as taxas de evasão, desde que estudos contínuos sejam realizados para aprimorar as estratégias e garantir que a evasão escolar seja minimizada a níveis aceitáveis (Momm; Momm; Souza, 2020).

## 2.2 PERMANÊNCIA E ÊXITO

A permanência no ensino superior refere-se à continuidade do aluno no curso até a sua conclusão, sendo uma medida essencial para a redução da evasão escolar (Ramos; Gonçalves Junior, 2020). O conceito abrange a capacidade dos estudantes de permanecerem na instituição até a finalização de sua formação acadêmica, superando as dificuldades que surgem ao longo do percurso. Em complemento, já o êxito está intimamente relacionado à conclusão do curso com sucesso, ou seja, ao alcance dos objetivos acadêmicos, tais como aprovação em disciplinas, obtenção de competências e habilidades necessárias para o mercado de trabalho, e conclusão

da graduação dentro do tempo esperado (Tinto, 1993).

Segundo Tinto (1993), a permanência e o êxito dependem da integração do estudante ao ambiente acadêmico e à vida universitária, o que contribui para um melhor desempenho e conclusão do curso.

### **2.2.1 Perspectiva Tradicional e Crítica sobre a Permanência e o Êxito**

Compreender as causas da evasão e retenção implica analisar a dinâmica que envolve o estudante em sua trajetória acadêmica. Segundo Cunha (2023) as causas da evasão podem estar associadas a diversos atores que integram o cotidiano escolar, levantando reflexões sobre o papel e a atuação de cada um. Nesse contexto, o autor destaca o papel: (i) da instituição, ao proporcionar o ambiente educativo; (ii) do corpo docente, com seus objetivos pedagógicos e didáticos; (iii) dos estudantes, em sua busca por conhecimento; e (iv) da sociedade e da família, com suas influências e forças determinantes nas ações e decisões dos estudantes

Entre os comportamentos individuais que podem ser considerados preditores da evasão escolar, Kaufman et al. (2019) destacam as baixas aspirações educacionais e profissionais, a alta frequência de faltas, os problemas de comportamento e os casos de gravidez. Além disso, o autor destaca que um histórico de repetências está fortemente associado a um aumento na probabilidade de abandono, sendo que a frequência dessas experiências eleva significativamente os riscos de desligamento escolar. A decisão de abandonar os estudos emerge, portanto, da interação entre esses diversos fatores (Kaufman et al., 2019).

Em estudo realizado na cidade de Caçador/SC, na mesma instituição a que se destina a presente pesquisa, Karasinski (2019) verificou que é necessário refletir sobre novas formas de organização da gestão escolar. O autor menciona, como parte do processo de adequação profissional às exigências legais, ingresso em um curso formal de especialização na área de EPT. Contudo, Karasinski (2019) ressalta que o sistema educacional, frequentemente centrado na burocracia, prioriza a documentação e formalização processual, relegando a segundo plano os processos não formais de educação. Embora a burocracia desempenhe uma função reguladora, ela não deve prevalecer sobre os objetivos pedagógicos efetivos (Karasinski, 2019). O autor ainda aponta a necessidade de se adotar formas mais flexíveis e inovadoras de organização, incluindo a valorização da formação não formal de docentes. Segundo Karasinski (2019), a liberdade e criatividade na educação demandam alterações no conteúdo e na estrutura curricular, bem como na formação docente, que não deve limitar-se à aquisição de conhecimentos exclusivamente acadêmicos, mas incluir práticas comunicativas e dialógicas.

Por fim, a qualidade das relações entre docentes e estudantes é apontada como elemento central para o desempenho escolar, o que requer um modelo de gestão mais participativo, envolvendo todos os setores escolares e promovendo soluções negociadas publicamente (Karasinski, 2019).

Adicionalmente, Karasinski (2019) sustenta que as escolas são frequentemente administradas com foco em índices que visam otimizar matrículas e estruturar recursos humanos e físicos, o que, sob uma gestão centralizada, gera pressões crescentes sobre a carga horária dos professores. Esse modelo conflita com os paradigmas da educação profissional, os quais precisam ser ajustados para promover uma educação cidadã, integral e transformadora das desigualdades, com perspectiva de humanização do trabalho (Karasinski, 2019).

Por fim, Karasinski (2019) destaca a relação problemática entre ensino, pesquisa e extensão. A escassa carga horária disponível para atividades de extensão e pesquisa nas escolas técnicas prejudica a formação docente, pois sobrecarrega os professores com demandas de ensino, enfraquecendo o processo de formação continuada. A prática docente, segundo Karasinski (2019), deve estar fundamentada na pesquisa, pois a ausência dessa prática transforma o professor em mero reprodutor de metodologias descontextualizadas e ineficazes. Imprescindível autor ressalta o princípio da indissociabilidade entre ensino, pesquisa e extensão, inovando a organização do trabalho docente e a gestão escolar, como forma de combater as ineficiências da EPT e enfrentar os desafios da evasão e retenção escolar (Karasinski, 2019).

### **2.2.2 Importância e efeitos de Políticas e Medidas de Permanência e Êxito**

As políticas de permanência são estratégias adotadas pelas instituições de ensino superior para apoiar os alunos durante sua trajetória acadêmica, com o objetivo de garantir sua conclusão bem-sucedida. Essas políticas incluem medidas como bolsas de estudo, auxílios financeiros, acompanhamento acadêmico e suporte psicológico. Nunes (2018) observa que a criação de um ambiente acolhedor e a disponibilização de serviços de apoio ao aluno são essenciais para que ele se sinta motivado e seguro para continuar seus estudos.

O Instituto Federal de Santa Catarina – Câmpus Caçador implementa o Plano Estratégico de Acesso, Permanência e Êxito dos Estudantes (PEAPEE, 2024), uma iniciativa estruturada em seis eixos. Esse plano visa compreender e enfrentar a evasão escolar, e promover estratégias para garantir que os estudantes ingressem, permaneçam e concluam seus cursos com sucesso.

A seguir detalham-se os seis eixos basilares que compõem o PEAPEE/IFSC:

Tabela 1: Eixos do Plano Estratégico de Acesso, Permanência e êxito dos estudantes do IFSC

Eixo	Detalhamento
Contextualização histórica	Analisa a evasão escolar sob uma perspectiva histórica e social, destacando sua evolução ao longo do tempo. Considerando mais de um século de oferta de ensino técnico, são apresentados fatores que influenciaram a permanência e o abandono dos estudantes em diferentes períodos.
Base legal	Aborda a legislação vigente relacionada ao acesso e à permanência dos alunos na educação profissional e tecnológica. Esse eixo destaca o amparo legal disponível tanto para a instituição quanto para os estudantes, garantindo direitos e deveres dentro do ambiente acadêmico.
Plano de Desenvolvimento Institucional (PDI)	Direcionado para ações concretas, esse eixo propõe mobilizações, projetos e estratégias institucionais voltadas ao enfrentamento da evasão escolar e à ampliação do acesso ao ensino.
Referência conceitual	Reúne os fundamentos teóricos e metodológicos que sustentam as políticas de acesso, permanência e êxito. Esse eixo define os princípios essenciais para orientar as ações institucionais voltadas à redução da evasão.
Aspectos metodológicos	Responsável pela estruturação das métricas utilizadas para avaliar e monitorar a evasão escolar no IFSC. Esse eixo engloba indicadores acadêmicos, sociais e institucionais, além de detalhar a organização das comissões responsáveis pela promoção do acesso, permanência e êxito dos estudantes.
Parâmetros para organização de ações e projetos	Fornece diretrizes para o desenvolvimento de iniciativas que atendam à diversidade do público estudantil. Esse eixo subsidia a elaboração de estratégias voltadas à inclusão e ao sucesso acadêmico dos alunos.

Fonte: Elaborado pelo autor com base no PEAPEE/IFSC (2024)

Com base nesses seis eixos, o PEAPEE (2024) estrutura um conjunto de ações e planejamentos estratégicos voltados à retenção e captação de novos estudantes. Além de contribuir para a permanência dos alunos em sala de aula, o plano fortalece o papel do IFSC como agente de transformação social, promovendo o crescimento educacional e impactando positivamente a qualidade de vida da comunidade.

### 2.3 INTELIGÊNCIA ARTIFICIAL

Em palestra proferida em 1985, Richard Feynman, Prêmio Nobel de 1965 e um dos mais reconhecidos físicos teóricos, debate temas críticos do campo da inteligência artificial. Em um diálogo com o público tem início com uma pergunta-chave: "*Você acha que haverá uma máquina que possa pensar como os humanos e ser mais inteligente do que os humanos.*". Para Feynman, as futuras máquinas não pensarão como os seres humanos, da mesma forma que um avião não voa como os pássaros (Kaufman, 2022). De tempos em tempos surgem novas tecnologias visando atender demandas (dentre elas sociais, mercadológicas, políticas etc.), provocando ansiedade por conhecer, e pelo uso nas mais diversas formas. E neste contexto, a Inteligência Artificial tem sido divulgada como uma novidade no âmbito da educação, suscitando novas possibilidades de uso, e até um certo receio pois há limites éticos a serem considerados.

Segundo Madeira e Neves (2021), as novas tecnologias, algumas delas denominadas de sistemas robóticos inteligentes, estão presentes nos processos produtivos, educacionais e informacionais, visando auxiliar e, até mesmo, substituir a presença humana.

A definição de inteligência artificial como a habilidade de um sistema simular funções cognitivas humanas, como aprendizado e raciocínio, é central para o entendimento de agentes de IA, os quais possuem a capacidade de adaptação e evolução, características fundamentais em ambientes dinâmicos (Witten; Frank, 2011).

Segundo Wooldridge (2009), um modelo de aprendizagem de máquina possui quatro características: autonomia, reatividade, proatividade e socialidade. A autonomia refere-se à capacidade do agente de agir sem a intervenção direta de um ser humano. A reatividade permite ao agente responder rapidamente a estímulos do ambiente, enquanto a proatividade envolve a habilidade de agir com base em objetivos, mesmo sem estímulos externos. A socialidade implica que o agente pode interagir com outros agentes ou seres humanos de maneira cooperativa para alcançar objetivos comuns.

## 2.4 INTELIGÊNCIA ARTIFICIAL E A EDUCAÇÃO

A evasão escolar é uma questão de preocupação global devido aos impactos negativos para a sociedade como um todo, tornando essencial sua investigação para melhor compreensão e atuação preventiva, reduzindo o risco de sua ocorrência. O estudo de Bitencourt, Silva e Xavier (2021) propõe a aplicação da Mineração de Dados Educacionais, utilizando técnicas de Aprendizado de Máquina, para identificar as variáveis mais relevantes na definição do perfil

dos estudantes com risco de evasão. Nessa pesquisa, foram empregadas as técnicas Máquina de Vetores de Suporte, *Gradient Boosting Machine*, *Random Forest* e comitê de máquina em um conjunto de 1.429 registros de alunos de cursos superiores de um campus do IFMG, abrangendo o período de 2013 a 2019. Os resultados indicaram que o comitê de máquina apresentou o melhor desempenho, permitindo determinar a importância das variáveis relacionadas ao fenômeno estudado e traçar o perfil dos estudantes propensos à evasão em diferentes períodos. Com base nesses achados, foi possível propor um processo para identificação e acompanhamento desses alunos.

A análise dos dados do estudo realizado por Pereira (2024) indica que a Inteligência Artificial já está sendo utilizada na educação para personalizar o ensino, identificar alunos em situação de risco e aprimorar a qualidade do aprendizado. Todavia, sua adoção enfrenta desafios, como barreiras técnicas, preocupações éticas e a aceitação por parte dos usuários. As conclusões desse estudo ressaltam a importância de uma incorporação ética da IA no ambiente educacional, o que assegura a proteção dos dados dos estudantes e promove a equidade. Além disso, destaca-se o grande potencial da IA para reduzir a evasão escolar, desde que sua implementação seja conduzida com planejamento e cautela.

Ao analisar a aplicação da Inteligência Artificial (IA) no combate à evasão no ensino superior brasileiro, os resultados apontaram que a IA tem potencial para aumentar a retenção estudantil, o que permite a identificação precoce de alunos em risco e a adoção de medidas de suporte personalizadas. Entretanto, foram identificados desafios éticos e práticos, como a privacidade dos dados e o risco de viés algorítmico, que dificultam sua implementação eficaz. Dessa forma, destaca-se a importância de abordagens colaborativas e multidisciplinares para enfrentar esses desafios, o que garante o uso ético e equitativo da IA na educação superior (Fernandes et. al., 2024).

## 2.5 MINERAÇÃO DE DADOS EDUCACIONAIS (MDE)

Diante da complexidade do fenômeno da evasão, a gestão acadêmica moderna tem recorrido a abordagens quantitativas para extrair conhecimento de grandes volumes de dados institucionais. A Mineração de Dados Educacionais (MDE), ou *Educational Data Mining* (EDM), surge como um campo interdisciplinar focado no desenvolvimento e aplicação de métodos computacionais para explorar dados gerados em ambientes de aprendizagem, com o objetivo de compreender melhor os estudantes e os processos educativos (Bezerra; Barros, 2017).

No contexto específico da evasão, o objetivo central da MDE é a construção de modelos preditivos. Estes modelos são sistemas capazes de analisar o perfil demográfico, o histórico e o comportamento de um aluno para calcular a probabilidade de ele pertencer ao grupo de risco (Silva, 2018). Em vez de depender apenas da percepção pedagógica, a MDE permite uma análise em larga escala, em busca de identificar padrões sutis e interações complexas entre variáveis que, para um ser humano, seriam impossíveis de detectar.

Porém, a aplicação prática da MDE no contexto brasileiro enfrenta um desafio primário e crítico: a coleta e preparação de dados. Como apontado por Figueiredo e Paula (2022), a extração de dados consiste em um dos maiores obstáculos, pois as informações mais ricas frequentemente não estão em bancos de dados estruturados, mas sim "presas" em documentos dispersos, como diários de classe, relatórios pedagógicos ou, como no caso deste trabalho, históricos escolares em formato PDF. O sucesso de qualquer modelo de MDE depende fundamentalmente da superação desta etapa.

Para organizar este processo, a indústria e a academia adotaram metodologias como o *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Este roteiro, amplamente reconhecido no campo da Mineração de Dados (Chapman et al., 2000), divide o projeto em seis fases iterativas e sequenciais, a saber:

- (i) Entendimento do Negócio (o problema da evasão),
- (ii) Entendimento dos Dados (os históricos),
- (iii) Preparação dos Dados (o pipeline de extração em Python),
- (iv) Modelagem (treinamento dos algoritmos),
- (v) Avaliação (análise das métricas) e
- (vi) Deploy (a aplicação web).

### **2.5.1 O Processo CRISP-DM na Mineração de Dados Educacionais**

O Cross-Industry Standard Process for Data Mining (CRISP-DM) não é apenas um roteiro linear, mas um ciclo de vida analítico iterativo que reflete a natureza exploratória da mineração de dados. Desenvolvido para ser neutro em relação à indústria e à ferramenta, ele se tornou o padrão de fato para projetos de Data Science devido à sua flexibilidade e foco na resolução de problemas de negócio — ou, neste caso, problemas educacionais (Chapman et al., 2000). A sua estrutura cíclica permite que o conhecimento adquirido numa fase possa refinar e redirecionar as fases anteriores, garantindo que o modelo final esteja verdadeiramente alinhado

com os objetivos institucionais.

A fase inicial, Entendimento do Negócio, é crítica e frequentemente subestimada. No contexto da MDE, isso traduz-se na tradução do problema pedagógico ("quais alunos estão em risco?") para um problema de mineração de dados ("classificação binária de evasão"). É nesta etapa que se definem os critérios de sucesso, não apenas em termos estatísticos (como acurácia ou *recall*), mas em termos de impacto prático para a gestão escolar, como a capacidade de intervenção precoce (Silva, 2018).

As fases subsequentes de Entendimento e Preparação dos Dados consomem, tradicionalmente, a maior parte do esforço do projeto. Em ambientes educacionais, onde os dados podem estar dispersos em sistemas legados ou documentos não estruturados, a preparação exige técnicas avançadas de limpeza, integração e, crucialmente, engenharia de features. A criação de novos atributos a partir dos dados brutos — como transformar um histórico de notas em indicadores de tendência de desempenho — é onde o conhecimento do domínio pedagógico é incorporado ao modelo matemático (Figueiredo; Paula, 2022).

Finalmente, as fases de Modelagem e Avaliação não são fins em si mesmas, mas meios para validar as hipóteses levantadas. A modelagem envolve a seleção e aplicação de algoritmos adequados à natureza dos dados (como o desbalanceamento de classes comum na evasão), enquanto a avaliação deve ir além das métricas técnicas, questionando se o modelo é robusto, justo e interpretável para os educadores. O ciclo fecha-se com o *Deploy*, que na MDE significa integrar o conhecimento descoberto de volta ao processo educativo, seja através de relatórios, dashboards ou sistemas de alerta, fechando o ciclo entre a tecnologia e a prática pedagógica.

## 2.6 MODELOS DE *MACHINE LEARNING* PARA CLASSIFICAÇÃO

O *Machine Learning* (Aprendizado de Máquina) é o subcampo da IA focado em algoritmos que aprendem padrões a partir de dados, sem serem explicitamente programados para cada regra. O problema da evasão é um problema clássico de classificação supervisionada (Dalla Valle; De Oliveira, 2018): o modelo é supervisionado ao ser treinado com um *dataset* histórico em que o resultado já é conhecido (a variável-alvo `TARGET_EVASAO`, onde 1=Evadiu e 0=Não Evadiu). O objetivo é que o modelo aprenda a função que mapeia os dados de entrada (X) para a saída (Y).

A escolha do algoritmo impacta diretamente a performance e a interpretação do modelo preditivo. Para este trabalho, foi selecionado um conjunto de modelos que representam diferentes abordagens de classificação, o que permite a comparação, uma prática comum na MDE para validar resultados (Gomes; Carneiro; Oliveira, 2020).

Como modelo de *baseline* (ponto de partida), foi utilizada a Regressão Logística (LR). Apesar de ser um modelo estatístico linear e mais simples, ele é rápido e oferece alta interpretação, servindo como uma régua inicial para medir o ganho de performance obtido com modelos mais complexos.

A classe de modelos mais poderosa para este tipo de problema tabular são os baseados em Árvores de Decisão. Uma única árvore, porém, tende a "decorar" os dados de treino (*overfitting*). Para resolver isso, utilizam-se modelos de *Ensemble* que combinam centenas de árvores para formar um modelo final forte. A primeira abordagem testada foi o *Random Forest* (RF), popularizado por Breiman (2001). O RF utiliza a técnica de *bagging* construindo múltiplas árvores em paralelo, cada uma com uma visão aleatória dos dados. A predição final é uma "votação" entre as árvores, tornando o modelo robusto e menos suscetível a ruídos.

A segunda e mais avançada abordagem de ensemble testada foi o *Gradient Boosting Machine* (GBM). Diferente do RF (que constrói árvores em paralelo), o boosting constrói árvores de forma sequencial. A primeira árvore faz uma predição; a segunda árvore é então treinada para corrigir os erros da primeira; a terceira corrige os erros da segunda, e assim por diante. Esta técnica permite que o modelo foque progressivamente nos casos mais difíceis de classificar (como os alunos que 'flertam' com a evasão), geralmente alcançando uma performance superior.

A performance dos modelos selecionados está intrinsecamente ligada à qualidade e à representatividade dos dados de entrada. Neste trabalho, parte-se da premissa metodológica de que a aplicação de Engenharia de *Features* — processo de criação de novos indicadores derivados, como *Media\_Nota\_Ultimo\_Sem* e *Qtd\_Trancamentos* — é essencial para enriquecer o *dataset*. O objetivo desta etapa é fornecer aos algoritmos atributos que potencializem a identificação de padrões complexos de evasão, submetendo estas variáveis criadas à avaliação matemática do modelo para determinar sua real relevância preditiva.

### 3. METODOLOGIA DA PESQUISA

Este capítulo detalha o percurso metodológico adotado para o desenvolvimento do modelo preditivo de evasão denominado IFSC-aliza. A pesquisa é classificada como aplicada quanto à sua natureza, quantitativa quanto à abordagem, e utiliza a Mineração de Dados Educacionais (MDE) como procedimento técnico. Para garantir o rigor e a replicabilidade, o projeto seguiu as fases da metodologia *Cross-Industry Standard Process for Data Mining* (CRISP-DM).

#### 3.1 ENTENDIMENTO DO PROBLEMA E DOS DADOS

A primeira fase consiste no entendimento do problema de negócio (neste caso, institucional), definido em conjunto com a gestão pedagógica como a necessidade de uma ferramenta proativa para identificar alunos em risco de evasão no Curso de Sistemas de Informação do IFSC-Caçador.

A fase de entendimento dos dados revelou o principal desafio metodológico deste trabalho: a inexistência de um banco de dados estruturado contendo o desempenho detalhado dos alunos. Os dados mais granulares e relevantes, como frequência semestral e notas por disciplina, estavam disponíveis apenas em relatórios não-estruturados: os históricos escolares individuais em formato PDF. A amostra de dados foi composta por 454 históricos escolares de alunos com status 'CURSANDO', 'CONCLUINTE', 'CANCELADO' ou 'TRANCADO'.

#### 3.2 PREPARAÇÃO DOS DADOS

A fase de preparação dos dados foi a mais extensa do projeto e corresponde diretamente aos objetivos específicos (a) e (c) desta pesquisa. Essa etapa foi dividida em quatro subetapas, as quais serão descritas a seguir:

##### 3.2.1 Extração de Dados - o *Pipeline* de PDF

Para solucionar o desafio dos dados não-estruturados, foi desenvolvido um *script* em linguagem Python (detalhado no Apêndice A e disponível online em <https://github.com/LuiZeredo/ifsc-aliza>) utilizando a biblioteca *pdfplumber*. Esta biblioteca foi escolhida por sua capacidade superior em extrair texto de PDFs baseando-se no layout visual,

preservando a estrutura de colunas e linhas. O *script* (*extrator\_tcc.py*) foi programado para iterar sobre a pasta de históricos e utilizar Expressões Regulares (regex) para extrair cada um dos campos de interesse (ex: 'Matrícula', 'Status', 'Qtd\_Trancamentos', etc.), bem como a tabela de disciplinas cursadas.

### 3.2.2 Engenharia de *Features*

Conforme a hipótese central deste trabalho, o modelo de aprendizagem de máquina não foi treinado apenas com dados brutos, mas com *features* (indicadores) de alto valor preditivo, criadas manualmente. Este processo, conhecido como Engenharia de *Features* destaca-se como uma contribuição metodológica do estudo pois, foi por meio dele que foi possível mapear todas as informações dos alunos.

O *script* de extração foi programado para calcular e criar as seguintes *features* para cada aluno:"

- Indicadores Demográficos: Idade (calculada a partir da data de nascimento), Cidade e UF de nascimento.
- Indicadores de Longo Prazo: CAA e Período\_Letivo\_Atual.
- Indicadores de Ação Ativa: Qtd\_Trancamentos.
- Indicadores de Comportamento Recente (Último Semestre): Media\_Nota\_Ultimo\_Sem, Media\_Freq\_Ultimo\_Sem, Aprov\_Ultimo\_Sem, Reprov\_Ultimo\_Sem (Repetidos para Penúltimo e Antepenúltimo Semestres).

### 3.2.3 Limpeza e Definição da Variável-Alvo

Os dados extraídos foram carregados em um *notebook Jupyter* (ambiente Google Colab) utilizando a biblioteca *pandas*<sup>3</sup>. A limpeza consistiu em:

- (I) converter valores textuais para numéricos (ex: 'Nenhum' -> 0);
- (II) tratar dados ausentes (N/A), preenchendo o CAA com 0.0 e a idade com a mediana do base de informações.

A variável-alvo (*y*), TARGET\_EVASAO, foi criada a partir da coluna Status. Alunos com status 'CANCELADO', 'TRANCADO' ou 'ABANDONO' foram classificados como Classe

---

<sup>3</sup> Pandas: biblioteca de código aberto para a linguagem de programação Python, utilizada para manipulação e análise de dados.

1 (Evadido), enquanto alunos 'CURSANDO' ou 'CONCLUINTE' foram classificados como Classe 0 (Não Evadido). O estudo cobriu a periodicidade de 2017.1 a 2025.2. A análise do *dataset* revelou a presença de um desbalanceamento de classes, uma característica intrínseca aos problemas de predição de evasão (Garcia; Braga; Carvalho, 2017).. A distribuição das classes é apresentada na Tabela 2 – Distribuição e Desbalanceamento da variável-alvo.

Tabela 2 – Distribuição e Desbalanceamento da variável-alvo.

<b>Categoria</b>	<b>Definição da Classe</b>	<b>Número de Casos (Exemplo)</b>	<b>Proporção</b>
<b>CLASSE 0</b>	Permanência (CURSANDO / CONCLUINTE)	172	37.89%
<b>CLASSE 1</b>	Evasão (CANCELADO / TRANCADO / ABANDONO)	282	62.11%
<b>TOTAL</b>		<b>454</b>	<b>100%</b>

Fonte: Elaborado pelo autor

### 3.2.4 Transformação e Escalonamento

Para que os dados pudessem ser lidos pelos algoritmos de *Machine Learning*, as *features* categóricas (como Nacionalidade, Cidade\_Nascimento e UF\_Nascimento) foram transformadas em vetores numéricos através da técnica de *One-Hot Encoding*<sup>4</sup> (utilizando `pandas.get_dummies`).

Finalmente, todas as *features* predictoras foram submetidas ao escalonamento de dados (normalização) usando o *StandardScaler* da biblioteca *scikit-learn* (Pedregosa et al., 2011). Esse passo coloca todas as *features* (ex: Idade variando de 18-50 e CAA de 0-1) numa escala comum, o que é necessário para rodar modelos como a Regressão Logística.

### 3.3 MÉTRICAS DE AVALIAÇÃO PARA DADOS DESBALANCEADOS

Um dos desafios centrais na avaliação de modelos de evasão é o desbalanceamento dos dados. Em cenários institucionais, a classe 'Evadiu' (minoritária) é quantitativamente inferior à classe 'Não Evadiu' (majoritária). Se um modelo hipotético classificasse 100% dos alunos como

---

<sup>4</sup> One-Hot Encoding: é um método de pré-processamento de dados que converte variáveis categóricas (como texto) em um formato numérico binário.

'Não Evadidos', ele poderia atingir uma Acurácia elevada (ex: 95%), mas seria ineficaz para a gestão escolar ao falhar na identificação dos discentes em risco. Por esta razão, a Acurácia é considerada uma métrica enviesada para este problema (GARCIA; BRAGA; CARVALHO, 2017). Para mitigar este viés e permitir a comparação objetiva entre modelos, adotou-se o F1-Score como métrica principal de decisão, visto que este indicador sintetiza a média harmônica entre a Precisão e a Recall extraídas da Matriz de Confusão.

Para uma avaliação correta, buscou-se analisar a Matriz de Confusão, que foi gerada para os modelos no Colab<sup>5</sup>. A matriz divide as predições em quatro quadrantes: Verdadeiros Positivos (VP - o modelo acertou que o aluno evadiu), Verdadeiros Negativos (VN - o modelo acertou que o aluno permaneceu), Falsos Positivos (FP - o modelo "alarmou" em vão) e Falsos Negativos (FN - o modelo "deixou escapar" um aluno que evadiu).

### 3.4 MODELAGEM

A fase de modelagem consistiu no treino e seleção do modelo de aprendizagem de máquina baseando-se no objetivo de projetar e implementar um modelo capaz de identificar padrões de risco de evasão.

Esta etapa envolveu a divisão do *dataset* em conjuntos de treino e teste, a comparação de desempenho entre múltiplos algoritmos de classificação — incluindo modelos de *baseline* e de *boosting* — e, por fim, a otimização do modelo selecionado, o *XGBoost*. O objetivo foi maximizar a performance preditiva, com foco principal em garantir o equilíbrio entre a capacidade de identificar alunos em risco (*Recall*) e a confiabilidade dos alertas gerados (*Precision*)

#### 3.4.1 Divisão e amostragem dos dados

O *dataset* processado foi dividido em dois subconjuntos: 80% para Treino (para "ensinar" o modelo) e 20% para Teste (para avaliá-lo em dados "nunca vistos"). Foi utilizada a amostragem estratificada (*stratify=y*) para garantir que a proporção de alunos evadidos (classe minoritária) fosse a mesma em ambos os conjuntos, evitando informações tendenciosas na avaliação.

---

### 3.4.2 Comparação de algoritmos

Conforme fundamentado, múltiplos algoritmos foram treinados e comparados para encontrar o de melhor performance para este *dataset*, como o de Regressão Logística (LR) como um modelo de *baseline*, Random Forest (RF) como o modelo de *ensemble (bagging)* e, por fim, XGBoost, LightGBM e CatBoost como modelos de *boosting*. No contexto da classificação de dados tabulares (como históricos escolares), o estado da arte é dominado pelos modelos de *Gradient Boosting* devido à sua capacidade de aprender sequencialmente com os erros das árvores anteriores e alcançar alta performance, como evidenciado em competições internacionais de ciência de dados (Chen; Guestrin, 2016).

### 3.4.3 Otimização de hiperparâmetros

Após a etapa de comparação, o XGBoost foi selecionado como o modelo mais promissor. A escolha não foi baseada na acurácia global, mas sim no seu desempenho superior na métrica *F1-Score* (média harmônica entre precisão e *recall*). Esta métrica foi priorizada por refletir a capacidade do modelo em lidar com o desbalanceamento das classes, garantindo que a identificação dos alunos em situação de evasão (Classe 1) não fosse penalizada em favor da classe majoritária.

## 3.5 AVALIAÇÃO

A avaliação do modelo final (*XGBoost* Otimizado) foi realizada utilizando-se o conjunto de Teste (os 20% de dados "novos"). Conforme justificado na Secção 2.4, a performance não foi medida pela Acurácia, mas sim pela análise da Matriz de Confusão e, especificamente, pelas métricas de Precisão, *Recall* e *F1-Score*.

## 3.6 AMBIENTE E FERRAMENTAS

O *pipeline* de extração foi desenvolvido em linguagem Python (versão 3.12.12) utilizando a biblioteca *pdfplumber*. A análise, limpeza, modelagem e avaliação foram executadas no ambiente Google Colab, utilizando as bibliotecas *pandas* (para manipulação de dados), *scikit-learn* (para pré-processamento, modelagem e avaliação) e *xgboost* (para o modelo final). Os

gráficos e visualizações foram gerados com *matplotlib* e *seaborn*."

### 3.7 LÓCUS DA PESQUISA

A pesquisa será realizada na cidade de Caçador, Santa Catarina.

Caçador é um município brasileiro localizado no Meio-Oeste do estado de Santa Catarina, com uma população estimada em 79.276 habitantes, segundo dados do IBGE (2021). O município conta com uma rede de ensino composta por escolas municipais e Centros Municipais de Educação Infantil (CMEIs), que atendem às modalidades de Creches, Pré-escolas, Ensino Fundamental (séries iniciais) e Educação de Jovens e Adultos (EJA). A estrutura educacional de Caçador reflete o compromisso com a formação de crianças, jovens e adultos, promovendo o acesso à educação em diferentes etapas do aprendizado.

#### 3.7.1 Participantes

Participarão da pesquisa alunos devidamente matriculados e ativos na rede de ensino e do Instituto Federal de Santa Catarina, nos cursos superiores de Sistemas de Informação.

Participaram da pesquisa alunos devidamente matriculados e que possuíam vínculo com o Curso Superior de Sistemas de Informação do IFSC – Câmpus Caçador no recorte temporal de 2017/1 a 2025/2. A definição da amostra final seguiu um processo de filtragem estruturado: inicialmente, foram coletados 467 registros brutos via SIGAA, durante o período de 29 de outubro a 6 de novembro. Deste total, aplicou-se o critério de exclusão para “Alunos Especiais” (13 registros), resultando em uma amostra válida de 454 estudantes.

Para a modelagem preditiva, esses 454 participantes foram segmentados de acordo com sua situação acadêmica final: 266 alunos classificados como evadidos via cancelamento, 16 alunos com curso trancado e 172 alunos que compõem o grupo de permanência (com status “cursando” ou “concluinte”). Esta composição garante que o modelo analise tanto o desfecho definitivo quanto os sinais precoces de interrupção de vínculo. Além disso, a idade média dos alunos é de 26,6 anos e as 5 principais cidades de origem são: Caçador (240 alunos), São Paulo (13 alunos), Videira (12 alunos), Fraiburgo (12 alunos) e Santa Cecília (12 alunos).

#### 3.7.2 Critérios de inclusão

Os critérios de inclusão é ser aluno matriculado e ativo do Instituto Federal de Santa

Catarina, Campus Caçador, nos referidos cursos desde o início do curso de Sistemas de Informação em 2017/1.

### **3.7.3 Critérios de exclusão**

Aluno participante com o status em “Forma de Ingresso” como “Aluno Especial”, tal qual entrou no IFSC para fazer disciplinas parciais, sendo destes, 13 registros.

### **3.7.4 Benefícios da pesquisa para os sujeitos/comunidade**

A pesquisa tem como um de seus principais objetivos, contribuir para a manutenção do quadro de alunos e entendimento das evasões nas instituições de ensino. Limita-se a instituição de ensino do IFSC, mas não se pode deixar de lado a consideração da aplicação em demais instituições. O IFSC Campus Caçador, possui uma série de fatores únicos, como geolocalização, público-alvo, e ofertas, mas sua contribuição para a comunidade com estruturas de qualidade, profissionais qualificados e cursos gratuitos, fazem desta, uma peça-chave para o desenvolvimento local.

Caçador está localizado na região meio-oeste, fazendo desta, uma cidade importante para a ligação de outras localidades. Portanto, possuir centros educacionais são essenciais para a evolução não apenas da cidade, mas de toda a localidade. Neste sentido, entende-se que mapear a evasão e entender o motivo das desistências dos alunos nos cursos de graduação, pode ser um ponto chave para o aumento deste caminho.

Além disso, o modelo preditivo pode ser aplicado diretamente ao ensino, tendo em vista que a avaliação da motivação da evasão pode colaborar com o próprio desempenho dos alunos e professores, mostrando pontos de melhorias para ambos. O Curso Superior de Sistemas de Informação (SI), por sua natureza tecnológica e alta demanda por disciplinas de base (como Lógica, Programação e Banco de Dados), impõe uma carga de estudo que exige alta dedicação e resiliência, fatores que se refletem diretamente nas taxas de permanência (Cruz; Feitosa, 2020). Desta forma, o modelo de inteligência artificial pode ser utilizado pela instituição de ensino como uma ferramenta para análise de melhorias de estrutura física e corpo técnico e docente, ajustando o suporte oferecido à rigidez curricular característica da área de TI.

## **3.8 INSTRUMENTOS DE COLETA DE DADOS**

Os dados foram coletados via histórico escolar de todos os alunos inscritos no curso de

Sistemas de Informação desde sua implementação em 2017/1. A extração dos documentos brutos do Sistema Integrado de Gestão de Atividades Acadêmicas (SIGAA) foi consolidada no dia 06 de novembro de 2025, servindo como base para a estruturação do dataset do projeto com as seguintes colunas da tabela 3:

Tabela 3 – Variáveis coletadas

Matricula	Status	CAA	Nacionalidade
Cidade_Nascimento	UF_Nascimento	Idade	Periodo_Letivo_Atual
Qtd_Trancamentos	Ano_Periodo_Conclusao	Media_Nota_Ultimo_Sem	Media_Freq_Ultimo_Sem
Aprov_Ultimo_Sem	Reprov_Ultimo_Sem	Media_Nota_Penultimo_Sem	Media_Freq_Penultimo_Sem
Aprov_Penultimo_Sem	Reprov_Penultimo_Sem	Media_Nota_Antepenultimo_Sem	Media_Freq_Antepenultimo_Sem
Aprov_Antepenultimo_Sem	Reprov_Antepenultimo_Sem		

Fonte: elaborado pelo autor

## 4. RESULTADOS E DISCUSSÃO

Este capítulo apresenta os resultados obtidos por meio da execução da metodologia descrita no Capítulo 3. Os resultados são apresentados em duas frentes: primeiro, a performance preditiva do modelo final otimizado; e em sequência, a análise de importância das *features* que validam a hipótese central da pesquisa.

### 4.1 PERFORMANCE PREDITIVA DO MODELO (*XGBOOST* OTIMIZADO)

Conforme detalhado na metodologia (Secção 3.3.2), múltiplos algoritmos foram comparados. O *XGBoost* apresentou o melhor equilíbrio inicial entre *Recall* e Precisão, sendo selecionado para a etapa de otimização de hiper parâmetros (*GridSearchCV*). O objetivo da otimização foi maximizar o F1-Score, garantindo um modelo robusto para lidar com o dataset desbalanceado.

O modelo final (*XGBoost* Otimizado) foi então avaliado no conjunto de teste (20% dos dados, "nunca vistos" pelo modelo). O Relatório de Classificação completo é apresentado na Tabela 4.

	<b>Precisão</b>	<b>Recall</b>	<b>F1-score</b>	<b>Suporte</b>
Não evadiu (0)	0.77	0.88	0.82	34
Evadiu (1)	0.92	0.84	0.88	57
Acurácia			0.86	91
Macro <i>AVG</i>	0.85	0.86	0.85	91
Média <i>AVG</i>	0.87	0.86	0.86	91

Tabela 4: Relatório de Classificação (*XGBoost* Otimizado)

Na tabela, a coluna 'Suporte' representa a quantidade absoluta de instâncias reais de cada classe presentes no conjunto de teste. Ou seja, as métricas de desempenho foram calculadas com base em uma amostra de 34 alunos que permaneceram e 57 alunos que evadiram, conferindo representatividade estatística aos percentuais apresentados.

Para uma análise visual da performance do modelo na classe-alvo (Evadiu), a Matriz de Confusão (Figura 2) detalha os acertos e erros.

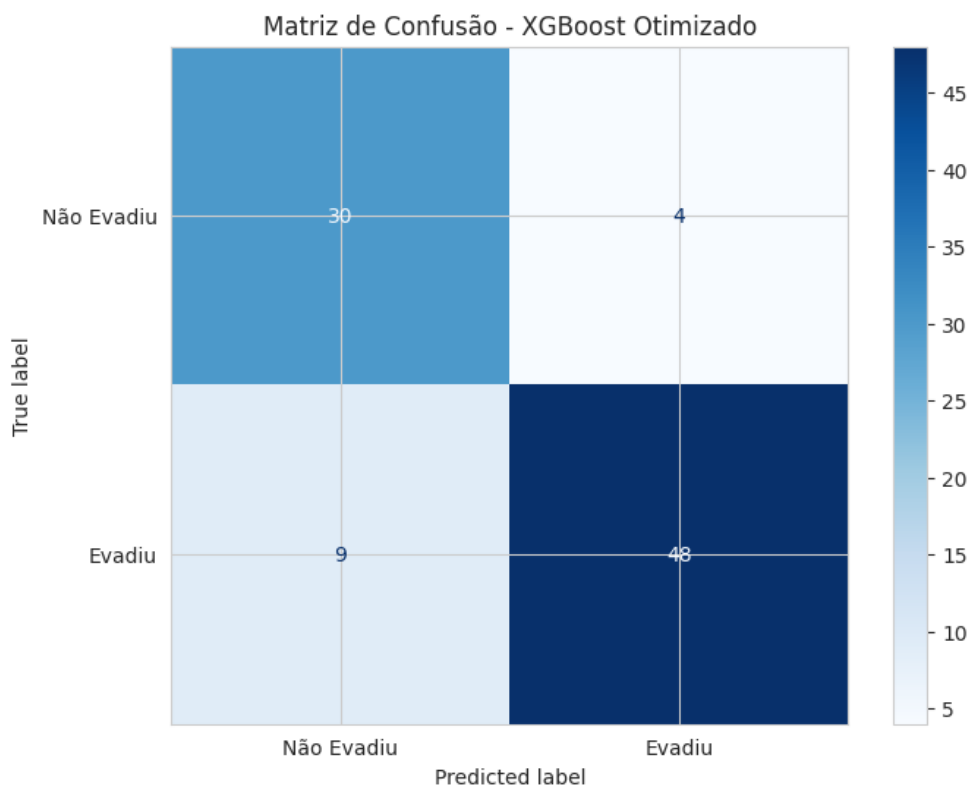


Figura 2: Matriz de Confusão (XGBoost Otimizado)

A análise detalhada da Matriz de Confusão permite compreender as nuances operacionais do modelo:

- i. **Verdadeiros Positivos (VP):** O modelo identificou corretamente a grande maioria dos alunos em situação de risco (48 casos), o que demonstra sua capacidade primária de detecção.
- ii. **Falsos Negativos (FN):** Foram registrados 9 casos em que o modelo falhou ao prever a evasão. Este é o quadrante mais crítico do problema, pois representa alunos que evadiram sem que o sistema emitisse um alerta, impedindo a atuação preventiva da instituição. O fato de o número de Falsos Negativos (9) ser superior ao de Falsos Positivos (4) indica que o modelo atual possui um viés conservador: ele prioriza a certeza do alerta (Precisão) ao invés de uma cobertura total (Recall).
- iii. **Falsos Positivos (FP):** Ocorreram apenas 4 casos de 'alarme falso', onde o modelo sinalizou risco para alunos que permaneceram. No contexto educacional, este erro possui um custo institucional baixo, pois resultaria apenas em uma intervenção pedagógica preventiva em um aluno que não necessariamente precisava, o que é preferível à perda do estudante.

Essa disparidade entre erros (mais FN do que FP) sugere que, embora o modelo seja

confiável quando emite um alerta (Alta Precisão), há uma oportunidade estratégica para trabalhos futuros em ajustar o processo de decisão do algoritmo.

A partir desta matriz, deriva-se duas métricas-chaves. O primeiro é o Recall, calculada por  $(VP / (VP + FN))$ . Esta é a métrica mais importante para o objetivo institucional, pois ela mede a "taxa de captura" do modelo – de todos os alunos que realmente evadiram, quantos o modelo conseguiu identificar - No estudo, alcançamos um Recall de 84%.

A segunda métrica é a Precisão, calculada por  $(VP / (VP + FP))$ . Ela mede a "fidelidade do alerta" – de todos os alunos que o modelo disse que iriam evadir, quantos ele acertou? No estudo, alcançamos uma Precisão de 92%. A F1-Score (0.88 no nosso modelo) é, então, a média harmônica entre estas duas, servindo como a "nota final" de equilíbrio. Foi esta métrica que usamos para otimizar o GridSearchCV <sup>6</sup>, garantindo um modelo que não só encontra a maioria dos alunos em risco (alto *Recall*), mas cujos alertas são, na sua maioria, corretos (alta Precisão).

## 4.2 ANÁLISE DE IMPORTÂNCIA DAS *FEATURES*

Para validar a premissa central do trabalho – de que indicadores de curto prazo e ações ativas são preditores mais fortes que métricas de longo prazo – foi realizada a análise de importância das *features*. Esta análise revela quais *features* (indicadores) os modelos mais utilizaram para tomar as suas decisões.

A Figura 3 apresenta os resultados do modelo *Random Forest* (RF), que atribuiu maior peso a indicadores de desempenho passivo, como a média de notas.

---

<sup>6</sup> GridSearchCV: é uma ferramenta da biblioteca Scikit-learn (SKlearn) de Python utilizada para a otimização sistemática de hiperparâmetros de modelos de machine learning

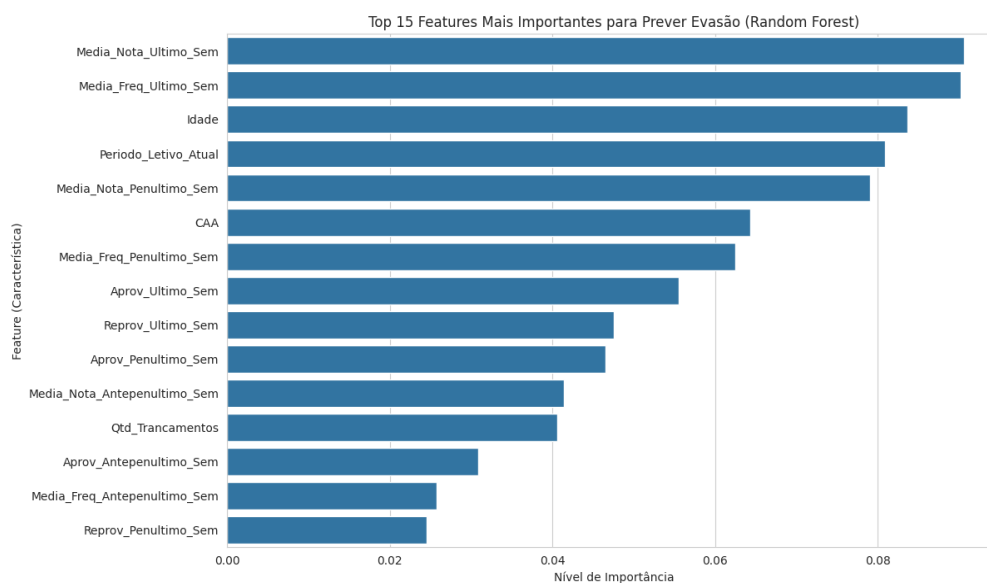


Figura 3: Importância das Features (Random Forest)

A Figura 4, por sua vez, apresenta os resultados do modelo XGBoost.

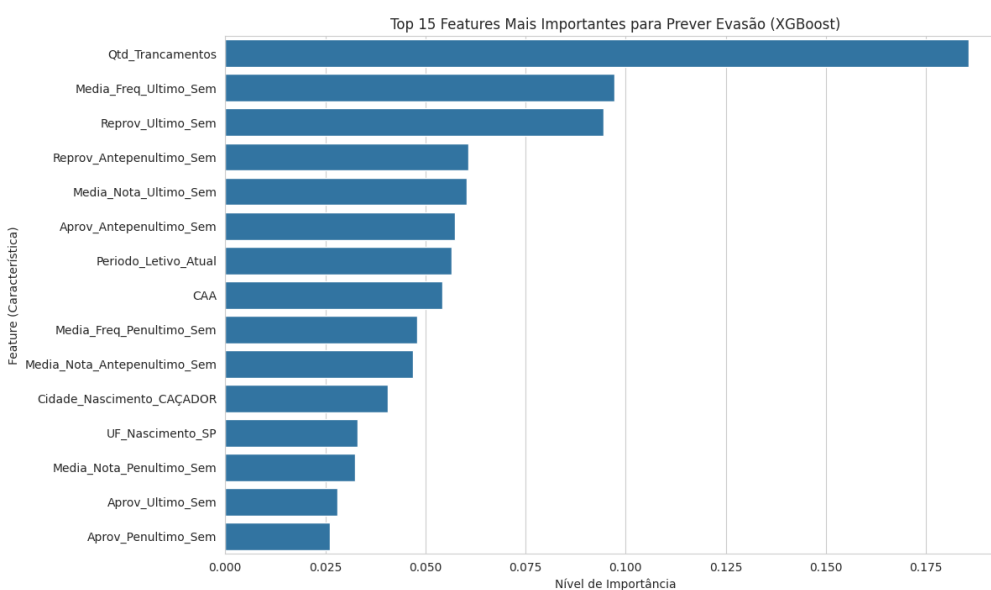


Figura 4: Importância das *Features* (XGBoost)

A diferença na hierarquia de importância das *features* entre os modelos é significativa. Enquanto o Random Forest (Figura 3) atribuiu maior peso a indicadores de desempenho passivo, a análise do modelo otimizado XGBoost (Figura 4) revelou uma hierarquia de risco mais alinhada à gestão institucional.

Os resultados do XGBoost demonstram que os preditores mais dominantes são o indicador Qtd\_Trancamentos (com importância de 0.185), emergindo como o fator mais crítico. Esta descoberta valida a necessidade de intervir *imediatamente* após o primeiro sinal formal de

desengajamento. O modelo atribuiu a este indicador uma relevância 3.4 vezes maior que o indicador tradicional de longo prazo.

Os demais fatores de alta relevância (como *Media\_Freq\_Ultimo\_Sem* e *Reprov\_Ultimo\_Sem*) são indicadores de tendência, comprovando que a desagregação do estudante se manifesta de forma aguda nos últimos períodos, e não através da lentidão do CAA.

O Coeficiente de Aproveitamento Acadêmico (CAA) figura apenas na oitava posição em importância. Este resultado corrobora a tese de que o CAA fornece o contexto histórico do risco, mas são os indicadores de curto prazo e as ações administrativas que fornecem o material necessário para a intervenção eficaz e driblar a evasão.

### 4.3 DISCUSSÃO DOS RESULTADOS

A apresentação dos resultados permite duas conclusões que auxiliam na visualização da validação do presente trabalho:

#### 4.3.1 O modelo preditivo é viável e confiável

A performance do modelo final (Tabela 4) demonstra eficácia para o objetivo proposto. O modelo alcançou uma taxa de *Recall* de 84% para a classe "Evadiu". Isto significa que, de cada 100 alunos que iriam evadir, a ferramenta é capaz de identificar e "levantar o alerta" para 84 deles. Mais importante ainda, a Precisão de 92% indica que os alertas gerados são extremamente confiáveis: de cada 100 alertas disparados pelo sistema, 92 são de fato alunos em risco real. Este equilíbrio, refletido no alto F1-Score (0.88), comprova o potencial da ferramenta, pois ela permite que a gestão pedagógica foque os seus recursos (humanos e financeiros) com o mínimo de desperdício em "falsos positivos".

Os resultados obtidos pelo IFSC-ALIZA confirmam que a evasão no campus é influenciada pela sobreposição de fatores. Enquanto a 'Qtd\_Trancamentos' representa um fator institucional e administrativo de alta relevância, a queda na frequência e o aumento de reprovações refletem fatores pessoais e pedagógicos que se manifestam de forma aguda antes do desligamento definitivo.

### 4.3.2. Contribuições e repercussões do modelo preditivo

A análise da importância das *features* (Figura 4) valida a premissa central de que o risco de evasão está associado a fatores ativos e de curto prazo. Este é o ponto onde o modelo preditivo fornece o seu maior valor analítico e contribui diretamente para a literatura de Mineração de Dados Educacionais.

O XGBoost identificou que o indicador de maior importância é a ação administrativa ativa: a Qtd\_Trancamentos (com importância de \$0.185\$). Esta constatação é o ponto principal da pesquisa, pois o indicador demonstrou ser \$3.4\times\$ mais impactante do que o tradicional Coeficiente de Aproveitamento Acadêmico (CAA, importância de \$0.054\$), que figurou apenas em 8º lugar na hierarquia de relevância do modelo.

Embora o modelo tenha priorizado ações ativas, a análise estatística confirma que o histórico escolar (CAA) ainda é um forte discriminador entre os grupos, conforme visualizado na Figura 5.

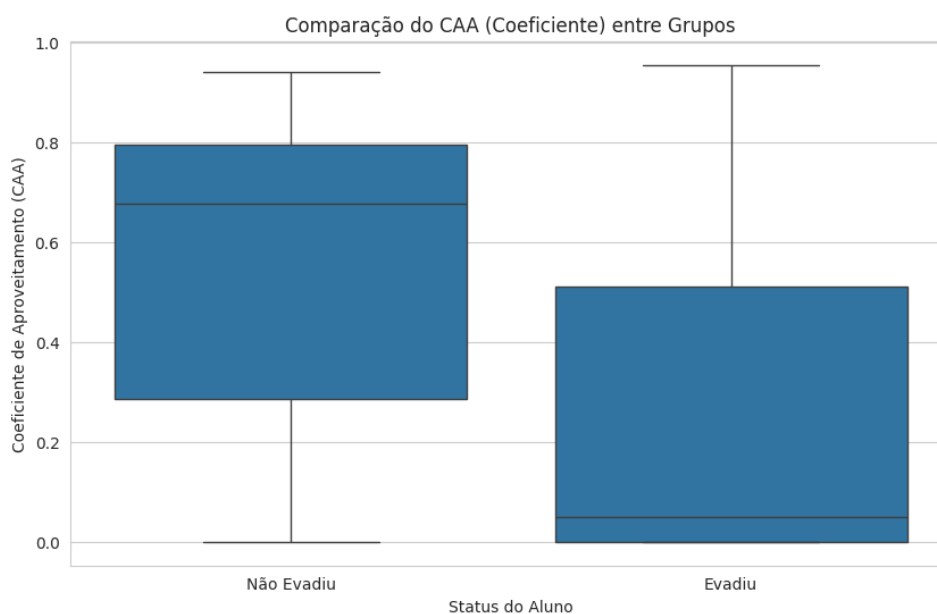


Figura 5 - Comparação da Distribuição do CAA entre Grupos

Observa-se na **Figura 5** que a mediana do CAA para o grupo 'Não Evadiu' (Classe 0) está consolidada em aproximadamente 0,78, demonstrando um desempenho médio satisfatório. Em contraste, a mediana para o grupo 'Evadiu' (Classe 1) é drasticamente inferior, situando-se em cerca de 0,17. Apesar dessa clara distinção estatística, o fato de o modelo XGBoost ter atribuído peso 3,4 vezes maior à variável 'Qtd\_Trancamentos' (Figura 4) do que ao CAA sugere que a *queda abrupta* de vínculo (trancamento) é um sinal de alerta muito mais urgente e preciso para

a intervenção pedagógica do que a simples média de notas acumulada.

Esta diferença quantitativa é a principal validação da hipótese de que métricas consolidadas são inferiores a indicadores de ação pois corrobora a tese de que a intervenção deve ser priorizada no momento da ação (trancamento ou queda de frequência) e não apenas no histórico do aluno. O trancamento (ou a interrupção formal do vínculo) se alinha com o modelo de integração de Tinto (1993), sendo a manifestação final de uma falha de integração acadêmica e social. Estudos mais recentes (Coimbra; Silva; Costa, 2021) corroboram que a interrupção formal reflete o momento da tomada de decisão de evasão, o que exige uma resposta imediata da instituição, transformando o trancamento em um ponto de virada para a intervenção pedagógica.

Adicionalmente, a análise da Figura 4 reforça a força dos Indicadores de Curto Prazo criados pela engenharia de *features*:

- I. Desempenho Recente Agregado: A importância combinada do *Recall* e Reprovação dos últimos três semestres (*Media\_Freq\_Ultimo\_Sem*, *Reprov\_Ultimo\_Sem*, *Media\_Nota\_Ultimo\_Sem* e suas versões anteriores) totaliza 34.4% ( $0.097 + 0.094 + 0.060 + 0.060 + 0.047 + 0.046$ ) da importância total do modelo. Este valor é mais de  $6\times$  superior ao peso atribuído unicamente ao CAA (0.054).
- II. Implicação na Intervenção: O modelo confirmou que a tendência de queda no desempenho (*Media\_Freq\_Ultimo\_Sem* com 0.097 e *Reprov\_Ultimo\_Sem* com 0.094) possui um poder preditivo significativamente maior do que o histórico consolidado do aluno. A intervenção do IFSC, portanto, deve ser orientada para a detecção de mudanças súbitas no comportamento do discente—caracterizadas por baixa frequência e aumento de reprovações—, e não somente na análise de médias históricas. A capacidade da ferramenta em ranquear estas *features* permite à gestão pedagógica concentrar a atenção nos indicadores que realmente sinalizam o risco iminente.

A descoberta constitui a principal contribuição do trabalho, sugerindo que a intervenção do IFSC deve ser priorizada não apenas para alunos com baixo desempenho histórico, mas imediatamente para qualquer aluno que solicite o primeiro trancamento.

## 5. CONSIDERAÇÕES FINAIS

Neste estudo, propôs-se **desenvolver** um modelo preditivo para identificar riscos de evasão escolar no IFSC – Câmpus Caçador, fundamentado em técnicas de *Machine Learning* (*Gradient Boosting*). A pesquisa foi motivada por dois desafios: a necessidade institucional de uma ferramenta de diagnóstico proativa e o obstáculo técnico de dados acadêmicos relevantes estarem disponibilizados exclusivamente em relatórios não-estruturados (PDFs), o que dificultava sua extração e análise. A premissa inicial era que indicadores de comportamento recente e ações administrativas ativas (como trancamentos) seriam preditores mais eficazes do que métricas tradicionais de longo prazo (como o CAA).

Para atingir este objetivo, foi desenvolvida uma metodologia baseada no *pipeline* CRISP-DM (cujas etapas foram detalhadas na Seção 2.5.1). Esta metodologia se iniciou com a criação de um *script* (Apêndice 1 – código para classificação, leitura e manipulação dos históricos escolares) para a extração automatizada de dados dos históricos em PDF. Em seguida, foi realizada uma extensiva etapa de engenharia de *features*, criando indicadores de desempenho dos últimos três semestres. Por fim, múltiplos algoritmos de *Machine Learning* foram treinados, comparados e otimizados. O modelo *XGBoost* foi selecionado para uma otimização profunda (*GridSearchCV*), culminando num modelo preditivo final robusto.

Os resultados validaram a eficácia da abordagem. O modelo preditivo (Apêndice 2 – Aprendizado de máquina para predição) (*XGBoost* Otimizado) demonstrou alta performance no conjunto de dados de teste, alcançando um F1-Score de 0.88, suportado por uma Precisão de 92% (alta confiabilidade dos alertas) e um *Recall* de 84% (alta capacidade de detectar alunos em risco). A análise de importância das *features* validou a premissa: a *Qtd\_Trancamentos* (uma ação ativa) emergiu como o preditor mais importante (0.185), sendo 3.4 vezes mais impactante que o CAA (0.054). Isto comprova que o risco imediato de evasão está mais ligado a ações administrativas e ao desempenho recente do que ao histórico consolidado do discente.

Apesar dos resultados promissores, este estudo possui limitações. A principal é a generalização: o modelo foi treinado com dados de 454 alunos de um único curso - Sistemas de Informação - em um único campus - Caçador. É provável que o modelo necessite de retreino ou ajuste para ser aplicado a outros cursos com perfis de evasão distintos (ex: cursos técnicos ou de ensino médio). Além disso, o *pipeline* de extração de PDFs é sensível a mudanças no layout do documento-fonte, exigindo manutenção caso o IFSC altere o formato dos seus históricos.

Como trabalhos futuros, sugere-se a expansão deste modelo para outros cursos e campus

do IFSC, validando a sua eficácia em populações estudantis maiores e mais diversas. Adicionalmente, a análise comparativa de algoritmos indicou que o modelo LightGBM atingiu uma performance superior (F1-Score 0.94) mesmo sem otimização; devido ao tempo hábil, optou-se pela otimização do XGBoost, mas um trabalho futuro seria aplicar o mesmo processo de GridSearchCV ao LightGBM, que se mostrou um candidato promissor para ganhos adicionais de performance. Por fim, sugere-se a implementação prática deste modelo preditivo como uma ferramenta de software, integrada ao sistema de gestão pedagógica. Tal ferramenta permitiria que os coordenadores de curso analisassem o risco de um aluno em tempo real, transformando esta pesquisa de um estudo preditivo para uma ferramenta de intervenção ativa, auxiliando diretamente na permanência e êxito dos estudantes.

No que tange às variáveis demográficas, este estudo utilizou a idade cronológica dos estudantes coletada no momento da extração dos dados. Reconhece-se, contudo, que a modelagem da idade como uma variável dinâmica, considerando a idade de ingresso ou a idade no exato momento da interrupção do vínculo, pode oferecer perspectivas complementares sobre a maturidade acadêmica. Para esta arquitetura inicial do IFSC-ALIZA, a idade serviu como um indicador de perfil geracional, permitindo identificar as faixas etárias com maior incidência de risco na amostra atual.

Como desdobramento desta pesquisa, identifica-se a oportunidade de desenvolver modelos preditivos especializados por fase do curso. Dado que a literatura aponta uma concentração crítica de evasão nos três primeiros semestres letivos, a desagregação dos dados por semestre permitiria identificar se os gatilhos de risco para alunos ingressantes diferem dos alunos veteranos. Essa evolução do IFSC-ALIZA possibilitaria intervenções ainda mais precoces, focadas especificamente no ciclo básico, onde o suporte pedagógico costuma ser mais demandado.

Com base na performance do modelo XGBoost, propõe-se à instituição a adoção de quatro indicadores-chave para o acompanhamento contínuo dos estudantes:

- i. Indicador de Vínculo: Monitoramento imediato de qualquer solicitação de trancamento de matrícula, visto que esta ação apresentou uma importância preditiva 3,4 vezes superior ao desempenho histórico (CAA).
- ii. Índice de Engajamento Recente: Observação de quedas súbitas na frequência escolar no último semestre letivo, fator que compõe a base de alta relevância do modelo.
- iii. Alerta de Repetência: Acompanhamento de alunos com mais de duas reprovações no semestre anterior, priorizando-os para atendimento pedagógico antes da consolidação do histórico.

- iv. Fator de Localidade: Atenção especial ao suporte de permanência para alunos vindos de cidades fora do eixo principal de Caçador, conforme sugerido pela distribuição geográfica da amostra.

Adicionalmente, propõe-se para versões futuras do IFSC-ALIZA a integração de uma análise qualitativa detalhada dos componentes curriculares. Atualmente, o modelo processa dados quantitativos de desempenho; contudo, a inclusão de técnicas de análise por disciplina permitiria identificar quais componentes específicos atuam como 'gargalos' pedagógicos. Essa evolução possibilitaria ao sistema não apenas sinalizar o risco de evasão, mas também correlacioná-lo a disciplinas específicas do núcleo comum ou tecnológico, fornecendo subsídios para que a coordenação do curso realize intervenções direcionadas, como monitorias ou revisão de pré-requisitos em componentes com altos índices de retenção.

## REFERÊNCIAS

ALENCAR, Ana Paula Pereira de Lima. *A evasão escolar no ensino superior pela perspectiva dos discentes: um estudo de caso no Curso de Bacharelado em Agronomia (Sede) da Universidade Federal Rural de Pernambuco*. 2020. Dissertação (Mestrado em Agronomia) – Universidade Federal Rural de Pernambuco, Recife, 2020. Disponível em: [https://bdtd.ibict.br/vufind/Record/UFPE\\_63249f1e2e5b2e006ebf6649643f1aa6](https://bdtd.ibict.br/vufind/Record/UFPE_63249f1e2e5b2e006ebf6649643f1aa6). Acesso em: 17 jan. 2025.

ANZOLIN, Ricardo Maximo; KRELING, Wagner Luiz. Análise das causas de evasão escolar nos cursos de aprendizagem industrial de uma unidade de educação profissional do SENAI/SC no ano de 2012. *Revista e-TECH: Tecnologias para Competitividade Industrial*, Florianópolis, v. 6, n. 1, p. 73-90, 2013.

BEZERRA, C. W. B.; BARROS, F. A. D. M. Uma abordagem baseada em mineração de dados para predição da evasão em cursos de graduação. In: CONGRESSO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO (CBIE), 6., 2017, Brasília. *Anais [...]*. Brasília: SBC, 2017. p. 836-845. Disponível em: <https://sol.sbc.org.br/index.php/cbie/article/view/1842>. Acesso em: 15 nov. 2025.

BITENCOURT, Wanderci Alves; SILVA, Diego Mello; XAVIER, Gláucia do Carmo. Pode a inteligência artificial apoiar ações contra evasão escolar universitária? *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 30, n. 116, p. 669-694, 2021.

BREIMAN, Leo. Random Forests. *Machine Learning*, v. 45, n. 1, p. 5-32, 2001.

CAVALCANTE, J. *Desigualdades sociais e acesso ao ensino superior*. São Paulo: Cortez, 2010.

CERATTI, Marcia Rodrigues Neves. *Evasão escolar: causas e consequências*. 2008. Disponível em: <http://www.diaadiaeducacao.pr.gov.br/portals/pde/arquivos/242-4.pdf>. Acesso em: 17 jan. 2025.

CHAPMAN, Pete *et al.* *CRISP-DM 1.0: Step-by-step data mining guide*. [S. l.]: SPSS, 2000.

CHEN, Tianqi; GUESTRIN, Carlos. XGBoost: A Scalable Tree Boosting System. *In: ACM SIGKDD INTERNATIONAL CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 22., 2016, San Francisco. *Proceedings* [...]. New York: ACM, 2016. p. 785-794.

CIAVATTA, Maria. A formação integrada: a escola e o trabalho como lugares de memória e de identidade. *Trabalho Necessário*, v. 3, n. 3, p. 1-20, 2005.

COELHO, A. J. D. P.; GARCIA, N. M. D. Permanência e evasão escolar em cursos técnicos da área da indústria: um estudo sobre escolas de educação profissional de Joinville/SC. *Revista Tecnologia e Sociedade*, Curitiba, v. 9, n. 18, 2013.

COIMBRA, C. L.; SILVA, L. C.; COSTA, M. A. A evasão na educação superior: definições e trajetórias. *Educação em Revista*, Belo Horizonte, v. 37, 2021.

CRUZ, M. S.; FEITOSA, R. A. Fatores de evasão no ensino superior: uma análise da relação professor-aluno e das práticas pedagógicas. *Revista Brasileira de Educação*, Rio de Janeiro, v. 25, e250005, 2020. Disponível em: <https://www.scielo.br/j/rbedu/a/gSfB7sXz9Q8G8qQ5D8yqQ6k/>. Acesso em: 16 nov. 2025.

CUNHA, J. *Evasão, permanência e êxito na educação profissional: um estudo de caso*. 2023. Disponível em: <https://editorapublicar.com.br/evasao-permanencia-e-exito-na-educacao-profissional-um-estudo-de-caso>. Acesso em: 17 jan. 2025.

CUNHA, M. I.; CUNHA, F. J. P.; DE LIMA, J. E. A evasão na educação superior: conceitos e desafios. *Avaliação*, Campinas, v. 25, n. 2, 2020.

DALLA VALLE, A.; DE OLIVEIRA, M. *Mineração de dados educacionais*. São Paulo: Novatec, 2018.

DAVOK, D. F.; BERNARD, R. P. Avaliação dos índices de evasão nos cursos de graduação da Universidade do Estado de Santa Catarina - UDESC. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas, v. 21, n. 2, p. 503–522, 2016. Disponível em: <https://www.scielo.br/j/aval/a/5VJRg7PrXDTQ5mYXK95rh8r/>. Acesso em: 17 jan. 2025.

DE MELLO, Simone Portella Teixeira; DOS SANTOS, Elaine Garcia. Diagnóstico e alternativas de contenção da evasão no curso de administração em uma universidade pública no sul do Brasil. *Revista Gestão Universitária na América Latina - GUAL*, Florianópolis, v. 5, n. 3, p. 67-80, 2012. Disponível em: <https://www.redalyc.org/pdf/3193/319327516005.pdf>. Acesso em: 17 jan. 2025.

FERNANDES, A. B. *et al.* Inteligência artificial no combate à evasão no ensino superior brasileiro. *Cuadernos de Educación y Desarrollo*, v. 16, n. 3, p. e3792, 2024.

FERREIRA, Sergio Guimarães; RIBEIRO, Giovanna; TAFNER, Paulo. *Abandono e evasão escolar no Brasil*. Brasília: Instituto Mobilidade e Desenvolvimento Social, 2022. Disponível em: <https://imdsbrasil.org/publicacao/abandono-e-evasio-escolar-no-brasil/>. Acesso em: 17 jan. 2025.

FIALHO, Marillia Gabriella Duarte. *A evasão escolar e a gestão universitária: o caso da Universidade Federal da Paraíba*. 2014. Dissertação (Mestrado em Gestão de Organizações Aprendentes) – Universidade Federal da Paraíba, João Pessoa, 2014.

FIGUEIREDO, J. R. P.; PAULA, M. C. de. Desafios e Oportunidades na Coleta de Dados para Mineração de Dados Educacionais no Brasil. *Revista Brasileira de Informática na Educação*, v. 30, p. 1-26, 2022. Disponível em: <https://www.br-ie.org/pub/index.php/rbie/article/view/11263>. Acesso em: 16 nov. 2025.

FREIRE, Paulo. *Pedagogia do oprimido*. 17. ed. Rio de Janeiro: Paz e Terra, 1987.

GARCIA, V. F.; BRAGA, A. P. S.; CARVALHO, A. C. P. L. F. de. Lidando com a Maldição do Desbalanceamento em Classificação. *In: Congresso Brasileiro de Inteligência Computacional (Cbic)*, 13., 2017, Niterói. *Anais [...]*. Niterói: IEEE, 2017. p. 1-8.

GOMES, A. A. *Evasão e evadidos: o discurso dos ex-alunos sobre evasão escolar nos cursos de licenciatura*. 1998. Dissertação (Mestrado em Educação) – Universidade Estadual Paulista, Araraquara, 1998.

GOMES, L. A.; CARNEIRO, T. C. J.; OLIVEIRA, M. C. F. de. Uma Comparação de Algoritmos de Machine Learning para Predição da Evasão na Educação a Distância. In: WORKSHOP DE INFORMÁTICA NA ESCOLA (WIE), 26., 2020. *Anais [...]*. Porto Alegre: SBC, 2020. p. 1-10.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Pesquisa Nacional por Amostra de Domicílios Contínua: Educação 2023*. Rio de Janeiro: IBGE, 2023. Disponível em: <https://www.ibge.gov.br>. Acesso em: 18 dez. 2024.

IBGE – INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. *Educação: Escolas e centros de ensino 2021*. Rio de Janeiro: IBGE, 2021. Disponível em: <https://www.ibge.gov.br>. Acesso em: 20 nov. 2024.

IFSC – INSTITUTO FEDERAL DE SANTA CATARINA. *Relatório de evasão escolar*. Caçador: IFSC, 2023.

IFSC – INSTITUTO FEDERAL DE SANTA CATARINA. *Plano Estratégico de Acesso, Permanência e Êxito dos Estudantes (PEAPEE)*. Florianópolis: IFSC, 2024.

INSTITUTO MOBILIDADE E DESENVOLVIMENTO SOCIAL – IMDS. *Mobilidade Social e Educação*. Rio de Janeiro: IMDS, 2021.

JORDAN, Michael I.; MITCHELL, Tom M. Machine learning: Trends, perspectives, and prospects. *Science*, v. 349, n. 6245, p. 255-260, 2015.

KARASINSKI, E. N. A formação docente e a permanência e êxito na educação profissional. *Revista Brasileira da Educação Profissional e Tecnológica*, v. 2, n. 17, e8603, 2019.

KAUFMAN, Dora. *Desmistificando a inteligência artificial*. Belo Horizonte: Autêntica Editora, 2022.

KAUFMAN, N. O. et al. *Ações para a permanência e êxito escolar na educação profissional e tecnológica*. Santa Maria: UFSM, 2019.

LEAL, E. A. *A permanência na educação superior: desafios e perspectivas*. 2015. Tese (Doutorado em Educação) – Universidade Federal da Bahia, Salvador, 2015.

MADEIRA, A. C. M.; NEVES, B. C. Inteligência Artificial na Tutoria EAD: Perspectivas na Ciência da Informação. *In: Encontro Nacional de Pesquisa em Ciência Da Informação (ENANCIB)*, 21., 2021, Rio de Janeiro. *Anais [...]*. Rio de Janeiro: Ancib, 2021.

MOMM, A. M. P.; MOMM, S. F.; SOUZA, M. L. H. A evasão escolar no curso superior de tecnologia em Jaraguá do Sul. *Repositório Institucional do IFSC*, 2020. Disponível em: <https://repositorio.ifsc.edu.br/handle/123456789/1417>. Acesso em: 17 jan. 2025.

NERI, Marcelo *et al.* *Motivos da evasão escolar*. Brasília: Fundação Getulio Vargas, 2009. Disponível em: [https://www.cps.fgv.br/ibrecps/rede/finais/Etapa3-Pesq\\_MotivacoesEscolares\\_sumario\\_principal\\_anexo-Andre\\_FIM.pdf](https://www.cps.fgv.br/ibrecps/rede/finais/Etapa3-Pesq_MotivacoesEscolares_sumario_principal_anexo-Andre_FIM.pdf). Acesso em: 27 out. 2025.

NUNES, E. A. Evasão no ensino superior: causas e estratégias de mitigação. *Educação e Pesquisa*, São Paulo, v. 44, p. e184407, 2018.

OLIVEIRA, B. *Desafios da permanência universitária*. São Paulo: Editora Universitária, 2011.

OLIVEIRA, R. S. *A evasão escolar no ensino médio: um estudo sobre as causas e consequências*. São Paulo: Cortez, 2018.

OLIVEIRA, T. *et al.* Mineração de dados educacionais aplicada à previsão de evasão. *Revista Brasileira de Informática na Educação*, v. 27, n. 1, 2019.

PEDREGOSA, F. *et al.* Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825-2830, 2011.

PEREIRA, Diego Antônio de Souza *et al.* A contribuição da inteligência artificial para a redução da evasão escolar no ensino superior brasileiro. *Revista Contemporânea*, v. 4, n. 3, p. e3527, 2024.

PRIM, A. L.; FÁVERO, J. D. Motivos da evasão escolar nos cursos de ensino superior de uma faculdade na cidade de Blumenau. *Revista e-TECH: Tecnologias para Competitividade Industrial*, v. 6, n. 1, p. 53–72, 2013.

RAMOS, Ana Carolina; GONÇALVES JUNIOR, Oswaldo. Evasão discente no Instituto Federal de Educação, Ciência e Tecnologia de São Paulo: um estudo de caso sob a perspectiva dos alunos evadidos. *Educação & Pesquisa*, São Paulo, v. 46, e215235, 2020. Disponível em: <https://www.scielo.br/j/ep/a/KtBRcFWvWKBt63LSQCVzdwh/>. Acesso em: 15 jun. 2025.

RAMOS, Ana Carolina; GONÇALVES, Oswaldo. Abandono e evasão escolar sob a ótica dos sujeitos envolvidos. *Educação e Pesquisa*, São Paulo, v. 50, e268037, 2024.

SANTOS, G. G.; ANDRIOLA, W. B. Fatores associados à evasão na educação superior: estudo de caso na Universidade Federal do Ceará. *Ensaio: Avaliação e Políticas Públicas em Educação*, Rio de Janeiro, v. 27, n. 104, p. 614-636, 2019.

SILVA, F. I. C. *et al.* Evasão escolar no curso de educação física da Universidade Federal do Piauí. *Avaliação: Revista da Avaliação da Educação Superior*, Campinas, v. 17, n. 2, p. 391–404, 2012.

SILVA, João Carlos; PEREIRA, Maria Luiza. A utilização de modelos de Inteligência Artificial na análise de dados em pesquisas acadêmicas. *Revista Brasileira de Inteligência Artificial*, v. 12, n. 3, p. 45-60, 2021.

SILVA, R. P. da. *Um modelo de predição de evasão de alunos de graduação da UFSC utilizando Mineração de Dados Educacionais*. 2018. Trabalho de Conclusão de Curso (Graduação em Sistemas de Informação) – Universidade Federal de Santa Catarina, Araranguá, 2018. Disponível em: <https://repositorio.ufsc.br/handle/123456789/191383>. Acesso em: 17 nov. 2025.

SOARES, T. M. *et al.* Fatores associados ao abandono escolar no ensino médio público de Minas Gerais. *Educação e Pesquisa*, São Paulo, v. 41, n. 3, p. 757-772, 2015.

TINTO, Vincent. Dropout from Higher Education: A Theoretical Synthesis of Recent Research. *Review of Educational Research*, v. 45, n. 1, p. 89-125, 1975.

TINTO, Vincent. *Leaving college: rethinking the causes and cures of student attrition*. 2. ed. Chicago: University of Chicago Press, 1993.

WITTEN, Ian H.; FRANK, Eibe. *Data mining: practical machine learning tools and techniques*. 3. ed. San Francisco: Morgan Kaufmann, 2011.

WOOLDRIDGE, Michael. *An introduction to MultiAgent Systems*. 2. ed. Chichester: John Wiley & Sons, 2009.

## GLOSSÁRIO

**MACHINE LEARNING (Aprendizado de Máquina):** Subcampo da Inteligência Artificial que se dedica ao estudo e construção de algoritmos que podem aprender e fazer previsões a partir de dados

## APÊNDICE

Apêndice 1 – Código para classificação, leitura e manipulação dos históricos escolares.

```
import os
import re
import pdfplumber
import pandas as pd
import logging
from pathlib import Path
from collections import defaultdict
from datetime import datetime

# --- Configs Iniciais ---
PASTA_PDFS = r'C:\Users\luizeredo\OneDrive\Documentos\Históricos\historicos'

ARQUIVO_SAIDA_EXCEL = 'dados_extrator_tcc.xlsx'
ARQUIVO_LOG = 'processamento_tcc.log'

# --- Configuração do Log ---
logging.basicConfig(
    filename=ARQUIVO_LOG,
    level=logging.INFO,
    format='%(asctime)s - %(levelname)s - %(message)s',
    encoding='utf-8'
)

# --- REGRAS DE EXTRAÇÃO (REGEX) ---
# Otimizadas para layout IFSC. O '^\\s*' lida com espaços em branco no início da linha.
REGEX_DADOS = {
```

```

'Matricula': r'Matrícula\s+(\d+)',
'Status': r'Status:\s+([A-Z ]+?)\s+',
'CAA': r'\*CAA:\s+([0-9.]++)',
'Cidade_Nascimento': r'Local de Nascimento:\s+([A-ZÁÉÍÓÚÇÃÕ ]+?)\s+UF:',
'UF_Nascimento': r'UF:\s+([A-Z]{2})',
'Nacionalidade': r'Nacionalidade:\s+([A-ZÁÉÍÓÚÇÃÕ ]+?)\s+',
'Data_Nascimento': r'Data de Nascimento:\s+(\d{2}/\d{2}/\d{4})',
'Qtd_Trancamentos': r'Trancamentos:\s+(\w+)',
'Periodo_Letivo_Atual': r'Período Letivo Atual:\s+(\d+)'
}

# Regex para a tabela de disciplinas. Captura (Semestre, Num1, Num2, Situação).
# DOTALL permite que a busca ignore quebras de linha dentro das células da tabela.
REGEX_DISCIPLINAS = re.compile(
    r'(\d{4}\.\d)\s+.*?\s+(\[d.,\])\s+(\[d.,\])\s+(APROVADO|REPROVADO|REP\.\s+
    FALTA)',
    re.IGNORECASE | re.DOTALL
)

def extrair_texto_pdf(caminho_pdf):
    """Extrai texto do PDF de forma limpa usando pdfplumber."""
    texto_completo = ""
    try:
        with pdfplumber.open(caminho_pdf) as pdf:
            for page in pdf.pages:
                texto_completo += page.extract_text() + "\n"
    return texto_completo
except Exception as e:
    logging.error(f'Erro ao ler o PDF: {caminho_pdf} -> {e}')

```

```
return None

def extrair_dado(texto, regra, padrao_nao_encontrado='N/A'):
    #Busca um único dado no texto usando Regex.
    match = re.search(regra, texto)
    if match:
        return match.group(1).strip()
    return padrao_nao_encontrado

def calcular_idade(data_nasc_str):
    #Calcula idade a partir da data de nascimento (dd/mm/aaaa).
    try:
        data_nasc = datetime.strptime(data_nasc_str, '%d/%m/%Y')
        hoje = datetime.now()
        # Cálculo preciso (Ano atual - Ano nasc - (Se o aniversário ainda não passou))
        idade = hoje.year - data_nasc.year - ((hoje.month, hoje.day) < (data_nasc.month,
data_nasc.day))
        return idade
    except:
        return 'N/A'

def _calcular_metricas_semestre(lista_disciplinas):
    """(Func Auxiliar) Calcula médias e contagens para um semestre."""
    total_notas, total_freq, aprovadas, reprovadas = 0, 0, 0, 0

    if not lista_disciplinas:
        return {'media_nota': 0, 'media_freq': 0, 'aprovadas': 0, 'reprovadas': 0}

    for d in lista_disciplinas:
```

```

total_notas += d['nota']
total_freq += d['freq']
if d['status'] == 'APROVADO':
    aprovadas += 1
elif d['status'].startswith('REP'):
    reprovadas += 1

count = len(lista_disciplinas)

return {
    'media_nota': round(total_notas / count, 2) if count > 0 else 0,
    'media_freq': round(total_freq / count, 2) if count > 0 else 0,
    'aprovadas': aprovadas,
    'reprovadas': reprovadas
}

def calcular_metricas_recentes(disciplinas):
    """Calcula métricas separadas para o último, penúltimo e antepenúltimo semestre."""

    metricas_finais = {}
    semestres_unicos = sorted(list(disciplinas.keys()))

    for i, nome_sem in enumerate(['Ultimo', 'Penultimo', 'Antepenultimo']):
        if len(semestres_unicos) >= (i + 1):
            sem_key = semestres_unicos[-(i + 1)]
            metricas = _calcular_metricas_semestre(disciplinas[sem_key])
        else:
            metricas = _calcular_metricas_semestre([]) # Semestre vazio = zeros

```

```
metricas_finais.update({
    f'Media_Nota_{nome_sem}_Sem': metricas['media_nota'],
    f'Media_Freq_{nome_sem}_Sem': metricas['media_freq'],
    f'Aprov_{nome_sem}_Sem': metricas['aprovadas'],
    f'Reprov_{nome_sem}_Sem': metricas['reprovadas']
})

return metricas_finais

def processar_historico_pdf(caminho_pdf):
    """Extrai todos os dados de UM histórico."""

    texto_pdf = extrair_texto_pdf(caminho_pdf)
    if not texto_pdf:
        return None

    # 1. Extrai dados estáticos
    dados_aluno = {}
    for chave, regra in REGEX_DADOS.items():
        dados_aluno[chave] = extrair_dado(texto_pdf, regra)

    # NOVO: Calcula Idade (Substitui 'Data_Nascimento')
    dados_aluno['Idade'] = calcular_idade(dados_aluno.get('Data_Nascimento', 'N/A'))

    # 2. Extrai dados das disciplinas (tabela)
    disciplinas_por_semestre = defaultdict(list)
    matches = REGEX_DISCIPLINAS.findall(texto_pdf)

    for match in matches:
        try:
```

```

semestre = match[0]

# Converte números capturados, lidando com vírgula (,)
num1, num2 = float(match[1].replace(',', '.')), float(match[2].replace(',', '.'))

# Lógica de separação: Nota <= 10.0, Frequência > 10.0
if num1 <= 10.0 and num2 > 10.0:
    nota, freq = num1, num2
elif num2 <= 10.0 and num1 > 10.0:
    nota, freq = num2, num1
elif num1 >= 70.0 and num2 <= 10.0: # Fallback p/ (7.0 e 80.0) ou (10.0 e 100.0)
    freq, nota = num1, num2
elif num2 >= 70.0 and num1 <= 10.0:
    freq, nota = num2, num1
else:
    logging.warning(f'Não deu p/ determinar nota/freq (ambiguidade) em
{caminho_pdf.name}')
    continue

disciplinas_por_semestre[semestre].append({
    'freq': freq, 'nota': nota, 'status': match[3].upper()
})
except Exception as e:
    logging.warning(f'Erro ao processar linha disciplina em {caminho_pdf.name}: {e}')

# 3. Calcula métricas recentes (Engenharia de Features)
metricas_recentes = calcular_metricas_recentes(disciplinas_por_semestre)
dados_aluno.update(metricas_recentes)

return dados_aluno

```

```
def main():
    """Lê todos os PDFs e salva os dados no Excel."""

    pasta = Path(PASTA_PDFS)
    if not pasta.is_dir():
        logging.critical(f'Pasta de PDFs não encontrada: {PASTA_PDFS}')
        return

    logging.info(f'--- INÍCIO DA EXTRAÇÃO (TCC) ---')
    todos_os_dados = []

    # Ordem das colunas para o Excel final
    colunas_excel = [
        'Matricula', 'Status', 'CAA', 'Nacionalidade', 'Cidade_Nascimento', 'UF_Nascimento',
        'Idade', 'Periodo_Letivo_Atual', 'Qtd_Trancamentos', 'Ano_Periodo_Conclusao',
        'Media_Nota_Ultimo_Sem', 'Media_Freq_Ultimo_Sem', 'Aprov_Ultimo_Sem',
'Reprov_Ultimo_Sem',
        'Media_Nota_Penultimo_Sem', 'Media_Freq_Penultimo_Sem',
'Aprov_Penultimo_Sem', 'Reprov_Penultimo_Sem',
        'Media_Nota_Antepenultimo_Sem', 'Media_Freq_Antepenultimo_Sem',
'Aprov_Antepenultimo_Sem', 'Reprov_Antepenultimo_Sem',
        'Arquivo_Origem'
    ]

    arquivos_pdf = list(pasta.glob('*.pdf'))

    if not arquivos_pdf:
        logging.warning(f'Nenhum PDF encontrado em {PASTA_PDFS}')
        return
```

```

for caminho_pdf in arquivos_pdf:
    dados_aluno = processar_historico_pdf(caminho_pdf)

    if dados_aluno:
        dados_aluno['Arquivo_Origem'] = caminho_pdf.name
        todos_os_dados.append(dados_aluno)
        logging.info(f'Sucesso ao processar {caminho_pdf.name} (Matrícula:
{dados_aluno.get('Matricula')})")
    else:
        logging.warning(f'Falha ao extrair dados de {caminho_pdf.name}")

# 3. Salva no Excel usando pandas
if not todos_os_dados:
    logging.warning("Processamento concluído, mas nenhum dado foi extraído.")
    return

try:
    df = pd.DataFrame(todos_os_dados)

    # Filtra e reordena colunas
    df_final = pd.DataFrame(columns=colunas_excel)
    df_final = pd.concat([df_final, df[df.columns.intersection(colunas_excel)]],
ignore_index=True)
    df_final = df_final.reindex(columns=colunas_excel, fill_value='N/A')

    df_final.to_excel(ARQUIVO_SAIDA_EXCEL, index=False, engine='openpyxl')
    logging.info(f'Sucesso! {len(todos_os_dados)} históricos salvos em
'{{ARQUIVO_SAIDA_EXCEL}}'.")

```

```
except Exception as e:
    logging.error(f"Erro ao salvar arquivo Excel: {e}")

logging.info("--- FIM DA EXTRAÇÃO ---")

if __name__ == "__main__":
    main()
```

## Apêndice 2 – Código do aprendizado de máquina para predição

```
import os
import re
import pdfplumber
import pandas as pd
import logging
from pathlib import Path
from collections import defaultdict
from datetime import datetime

# --- Configs Iniciais ---
# Caminho da pasta que contém os PDFs
PASTA_PDFS = r'C:\Caminho\Para\Seus\PDFs'

ARQUIVO_SAIDA_EXCEL = 'dados_extrator_tcc.xlsx'
ARQUIVO_LOG = 'processamento_tcc.log'

# --- Configuração do Log ---
logging.basicConfig(
    filename=ARQUIVO_LOG,
```

```

level=logging.INFO,
format='%(asctime)s - %(levelname)s - %(message)s',
encoding='utf-8'
)

# --- REGRAS DE EXTRAÇÃO (REGEX) ---
# Regras otimizadas para o layout do IFSC
REGEX_DADOS = {
    'Matricula': r'Matrícula\s+(\d+)',
    'Status': r'Status:\s+([A-Z ]+?)\s+',
    'CAA': r'\*CAA:\s+([0-9.]+)',
    'Cidade_Nascimento': r'Local de Nascimento:\s+([A-ZÁÉÍÓÚÇÃÕ ]+?)\s+UF:',
    'UF_Nascimento': r'UF:\s+([A-Z]{2})',
    'Nacionalidade': r'Nacionalidade:\s+([A-ZÁÉÍÓÚÇÃÕ ]+?)\s+',
    'Data_Nascimento': r'Data de Nascimento:\s+(\d{2}/\d{2}/\d{4})',
    'Qtd_Trancamentos': r'Trancamentos:\s+(\w+)',
    'Periodo_Letivo_Atual': r'Período Letivo Atual:\s+(\d+)'
}

# Regex p/ tabela: Captura (Semestre, Num1, Num2, Situação)
REGEX_DISCIPLINAS = re.compile(
    r'(\d{4}\.\d)\s+.*?\s+([\d.,]+)\s+([\d.,]+)\s+(APROVADO|REPROVADO|REP.\s+
FALTA)',
    re.IGNORECASE | re.DOTALL
)

def extrair_texto_pdf(caminho_pdf):
    """Extrai texto do PDF de forma limpa usando pdfplumber."""
    texto_completo = ""

```

```

try:
    with pdfplumber.open(caminho_pdf) as pdf:
        for page in pdf.pages:
            texto_completo += page.extract_text() + "\n"
        return texto_completo
except Exception as e:
    logging.error(f'Erro ao ler o PDF: {caminho_pdf} -> {e}')
    return None

def extrair_dado(texto, regra, padrao_nao_encontrado='N/A'):
    """Busca um único dado no texto usando Regex."""
    match = re.search(regra, texto)
    if match:
        return match.group(1).strip()
    return padrao_nao_encontrado

def calcular_idade(data_nasc_str):
    """Calcula idade a partir da data de nascimento (dd/mm/aaaa)."""
    try:
        data_nasc = datetime.strptime(data_nasc_str, '%d/%m/%Y')
        hoje = datetime.now()
        # Cálculo preciso
        idade = hoje.year - data_nasc.year - ((hoje.month, hoje.day) < (data_nasc.month,
data_nasc.day))
        return idade
    except:
        return 'N/A'

def _calcular_metricas_semestre(lista_disciplinas):

```

```

"""(Func Auxiliar) Calcula médias e contagens para um semestre."""
total_notas, total_freq, aprovadas, reprovadas = 0, 0, 0, 0

if not lista_disciplinas:
    return {'media_nota': 0, 'media_freq': 0, 'aprovadas': 0, 'reprovadas': 0}

for d in lista_disciplinas:
    total_notas += d['nota']
    total_freq += d['freq']
    if d['status'] == 'APROVADO':
        aprovadas += 1
    elif d['status'].startswith('REP'):
        reprovadas += 1

count = len(lista_disciplinas)

return {
    'media_nota': round(total_notas / count, 2) if count > 0 else 0,
    'media_freq': round(total_freq / count, 2) if count > 0 else 0,
    'aprovadas': aprovadas,
    'reprovadas': reprovadas
}

def calcular_metricas_recetes(disciplinas):
    """Calcula métricas separadas para o último, penúltimo e antepenúltimo semestre."""

    metricas_finais = {}
    semestres_unicos = sorted(list(disciplinas.keys()))

```

```

for i, nome_sem in enumerate(['Ultimo', 'Penultimo', 'Antepenultimo']):
    if len(semestres_unicos) >= (i + 1):
        sem_chave = semestres_unicos[-(i + 1)]
        metricas = _calcular_metricas_semestre(disciplinas[sem_chave])
    else:
        metricas = _calcular_metricas_semestre([]) # Semestre vazio = zeros

    metricas_finais.update({
        f'Media_Nota_{nome_sem}_Sem': metricas['media_nota'],
        f'Media_Freq_{nome_sem}_Sem': metricas['media_freq'],
        f'Aprov_{nome_sem}_Sem': metricas['aprovadas'],
        f'Reprov_{nome_sem}_Sem': metricas['reprovadas']
    })

return metricas_finais

def processar_historico_pdf(caminho_pdf):
    """Extrai todos os dados de UM histórico."""

    texto_pdf = extrair_texto_pdf(caminho_pdf)
    if not texto_pdf:
        return None

    # 1. Extrai dados estáticos
    dados_aluno = {}

    for chave, regra in REGEX_DADOS.items():
        dados_aluno[chave] = extrair_dado(texto_pdf, regra)

    # Calcula Idade
    dados_aluno['Idade'] = calcular_idade(dados_aluno.get('Data_Nascimento', 'N/A'))

```

```

# 2. Extrai dados das disciplinas (tabela)
disciplinas_por_semestre = defaultdict(list)
matches = REGEX_DISCIPLINAS.findall(texto_pdf)

for match in matches:
    try:
        semestre = match[0]

        # Converte números capturados, lidando com vírgula (,)
        num1, num2 = float(match[1].replace(',', '.')), float(match[2].replace(',', '.'))

        # Lógica de separação: Nota <= 10.0, Frequência > 10.0
        if num1 <= 10.0 and num2 > 10.0:
            nota, freq = num1, num2
        elif num2 <= 10.0 and num1 > 10.0:
            nota, freq = num2, num1
        elif num1 >= 70.0 and num2 <= 10.0: # Fallback p/ (7.0 e 80.0)
            freq, nota = num1, num2
        elif num2 >= 70.0 and num1 <= 10.0:
            freq, nota = num2, num1
        else:
            logging.warning(f"Não deu p/ determinar nota/freq (ambiguidade) em
{caminho_pdf.name}")
            continue

        disciplinas_por_semestre[semestre].append({
            'freq': freq, 'nota': nota, 'status': match[3].upper()
        })
    except Exception as e:

```

```

logging.warning(f"Erro ao processar linha disciplina em {caminho_pdf.name}: {e}")

# 3. Calcula métricas recentes
metricas_recentes = calcular_metricas_recentes(disciplinas_por_semestre)
dados_aluno.update(metricas_recentes)

return dados_aluno

def main():
    """Lê todos os PDFs e salva os dados no Excel."""

    pasta = Path(PASTA_PDFS)
    if not pasta.is_dir():
        logging.critical(f"Pasta de PDFs não encontrada: {PASTA_PDFS}")
        return

    logging.info(f"--- INÍCIO DA EXTRAÇÃO (TCC) ---")
    todos_os_dados = []

    # Ordem das colunas para o Excel final
    colunas_excel = [
        'Matricula', 'Status', 'CAA', 'Nacionalidade', 'Cidade_Nascimento', 'UF_Nascimento',
        'Idade', 'Periodo_Letivo_Atual', 'Qtd_Trancamentos', 'Ano_Periodo_Conclusao',
        'Media_Nota_Ultimo_Sem', 'Media_Freq_Ultimo_Sem', 'Aprov_Ultimo_Sem',
'Reprov_Ultimo_Sem',
        'Media_Nota_Penultimo_Sem', 'Media_Freq_Penultimo_Sem',
'Aprov_Penultimo_Sem', 'Reprov_Penultimo_Sem',
        'Media_Nota_Antepenultimo_Sem', 'Media_Freq_Antepenultimo_Sem',
'Aprov_Antepenultimo_Sem', 'Reprov_Antepenultimo_Sem',
        'Arquivo_Origem'

```

```
]

arquivos_pdf = list(pasta.glob('*pdf'))

if not arquivos_pdf:
    logging.warning(f'Nenhum PDF encontrado em {PASTA_PDFS}')
    return

for caminho_pdf in arquivos_pdf:
    dados_aluno = processar_historico_pdf(caminho_pdf)

    if dados_aluno:
        dados_aluno['Arquivo_Origem'] = caminho_pdf.name
        todos_os_dados.append(dados_aluno)
        logging.info(f'Sucesso ao processar {caminho_pdf.name} (Matrícula:
{dados_aluno.get('Matricula')})')
    else:
        logging.warning(f'Falha ao extrair dados de {caminho_pdf.name}')

# 3. Salva no Excel usando pandas
if not todos_os_dados:
    logging.warning("Processamento concluído, mas nenhum dado foi extraído.")
    return

try:
    df = pd.DataFrame(todos_os_dados)

    # Filtra e reordena colunas
    df_final = pd.DataFrame(columns=colunas_excel)
    df_final = pd.concat([df_final, df[df.columns.intersection(colunas_excel)]],
```

```
ignore_index=True)

df_final = df_final.reindex(columns=colunas_excel, fill_value='N/A')

df_final.to_excel(ARQUIVO_SAIDA_EXCEL, index=False, engine='openpyxl')
    logging.info(f'Sucesso!  {len(todos_os_dados)}  históricos  salvos  em
'{{ARQUIVO_SAIDA_EXCEL}}'.')

except Exception as e:
    logging.error(f'Erro ao salvar arquivo Excel: {e}')

logging.info("--- FIM DA EXTRAÇÃO ---")

if __name__ == "__main__":
    main()
```