

## Inteligência de Dados e Modelagem Preditiva na Gestão Universitária: O Observatório de Evasão da UFPR como Ferramenta de Apoio à Decisão

Lineu Alberto Cavazani de Freitas<sup>1</sup> , Leonardo Gomes de Melo<sup>1</sup> ,  
Edicleia Furlanetto<sup>1</sup> , Eliane Felisbino<sup>1</sup> ,  
Vitor Hiei Noriduki Fujimoto<sup>1</sup> 

<sup>1</sup>Universidade Federal do Paraná (UFPR)  
Rua XV de Novembro, 1299 – Centro – Curitiba – PR – Brasil  
CEP 80060-000

lineu.alberto@ufpr.br

**Abstract.** *Student attrition in Higher Education is a challenge that affects public investment and the social mission of universities. This paper presents the Dropout and Low Demand Data Observatory at Federal University of Paraná, an initiative designed to institutionalize proactive and evidence-based management. The methodology is organized into four complementary pillars: (i) monitoring dashboards for student flow diagnosis; (ii) a risk calculator that allows for anticipating dropout; (iii) surveys with dropouts and alumni to understand the subjective causes of the phenomenon; and (iv) a portal that functions as a hub for scientific outreach and knowledge dissemination on the subject. The results demonstrate that integrating these fronts allows the university to move beyond simply recording withdrawals to acting preventively in student support. The project stands out for transforming administrative records into strategic decisions and shared knowledge, leveraging open-source technologies such as R and Python to ensure data sovereignty and strengthen retention policies.*

**Keywords:** *Evidence-Based Management, Digital Governance, Student Attrition, Machine Learning, Knowledge Dissemination.*

**Resumo.** *A evasão no Ensino Superior é um desafio que compromete o investimento público e a função social das universidades. Este artigo apresenta o Observatório de Dados de Evasão e Baixa Procura da Universidade Federal do Paraná (UFPR), uma iniciativa desenhada para institucionalizar uma gestão proativa e baseada em evidências. A metodologia organiza-se em quatro pilares complementares: (i) painéis de monitoramento para diagnóstico do fluxo discente; (ii) uma calculadora de risco que permite antecipar o abandono; (iii) pesquisas com evadidos e egressos para compreender as causas subjetivas do fenômeno; e (iv) um portal que funciona como um hub de difusão científica e disseminação de conhecimentos sobre o tema. Os resultados demonstram que a integração dessas frentes permite que a universidade deixe de apenas registrar o desligamento para atuar preventivamente no acolhimento do estudante. O projeto destaca-se por transformar registros administrativos em decisões estratégicas e conhecimento compartilhado, utilizando ecossistemas de software*

*livre (R e Python) para garantir a soberania de dados e fortalecer as políticas de permanência.*

**Palavras-chave:** *Gestão Baseada em Evidências, Governança Digital, Evasão Escolar, Aprendizado de Máquina, Disseminação de Conhecimento.*

## 1. Introdução

A evasão no Ensino Superior é um dos desafios mais persistentes e complexos da gestão pública contemporânea [Barragán Moreno and González Támara, 2024]. Mais do que a simples interrupção de um percurso acadêmico, o abandono escolar representa uma subutilização de recursos públicos, a ociosidade de infraestrutura e, fundamentalmente, a frustração de projetos de ascensão social e desenvolvimento nacional [Sampaio et al., 2022]. No contexto das Instituições de Ensino Superior (IES) públicas, onde o investimento por aluno é integralmente custeado pelo Estado, a eficiência na retenção discente é necessária tanto sob a ótica social quanto sob a de responsabilidade fiscal.

A literatura especializada, tais como os trabalhos de Enguídanos et al. [2023] e Nurmalitasari et al. [2023], destaca que a evasão não é um evento isolado, mas o desfecho de um processo gradual de desengajamento influenciado por variáveis endógenas e exógenas. No contexto brasileiro, a formulação de políticas de permanência tem enfrentado, historicamente, a barreira da fragmentação de dados e da dificuldade de processar volumes massivos de informações acadêmicas em tempo hábil para a intervenção pedagógica. Assim, a transição para modelos de Governança Digital nas IES exige que os dados deixem de ser apenas registros burocráticos e passem a servir como ativos estratégicos para a equidade educacional.

Nos últimos anos, o cenário global de permanência estudantil foi severamente impactado pela pandemia de COVID-19 [Sharaievska et al., 2022]. Estudos apontam que a crise sanitária exacerbou vulnerabilidades socioeconômicas [Salakhova et al., 2022] e psicossociais [Lopes et al., 2024], alterando drasticamente a dinâmica de engajamento dos estudantes. Na Universidade Federal do Paraná (UFPR), este fenômeno manifestou-se de forma aguda: o biênio 2021-2022 registrou picos históricos de desligamentos, superando a marca de 4.000 evadidos anuais, o que demandou uma resposta institucional coordenada e baseada em evidências.

Tradicionalmente, a gestão da evasão nas universidades brasileiras tem sido pautada por uma postura reativa, na qual o fenômeno é analisado apenas após a sua consumação administrativa. Entretanto, o avanço das tecnologias e o fortalecimento da cultura de decisões baseadas em evidências abrem espaço para uma gestão pública preditiva. A aplicação de técnicas de Aprendizado de Máquina e Ciência de Dados permite que a instituição não apenas otimize o gasto público, mas também exerça uma função social ativa, identificando precocemente o estudante em situação de vulnerabilidade que, sem o suporte institucional, consolidaria o abandono.

Este artigo descreve a concepção e implementação do Observatório de Dados de Evasão e Baixa Procura da UFPR. A iniciativa diferencia-se por não ser apenas um repositório, mas um ecossistema que integra: (i) monitoramento em tempo real de indicadores de procura e permanência através de painéis interativos; (ii) modelagem preditiva individual via Aprendizado de Máquina integrada ao sistema de gestão acadêmica; (iii)

pesquisas qualitativas estruturadas com evadidos e egressos para humanização dos dados; e (iv) estratégias de transparência ativa e software livre.

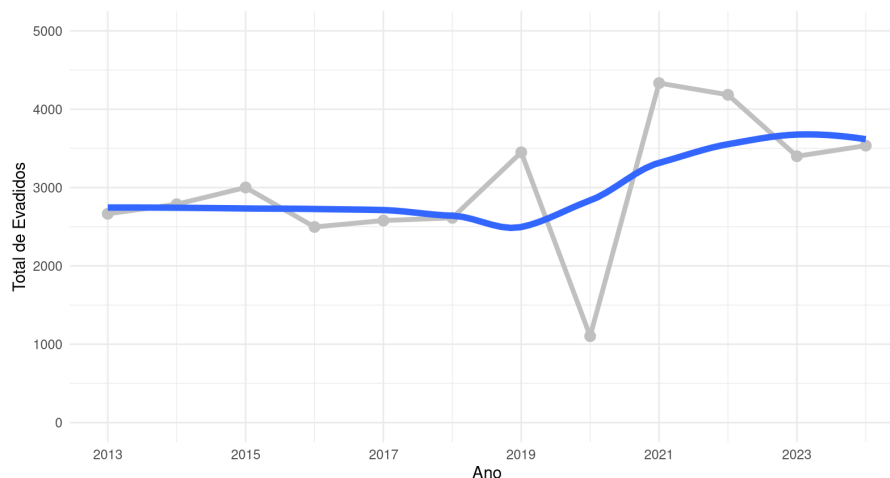
O trabalho está organizado da seguinte forma: a Seção 2 discute o problema público e o contexto institucional da crise de evasão; a Seção 3 detalha a metodologia e os quatro pilares do framework do Observatório; a Seção 4 apresenta os resultados técnicos do modelo preditivo e os impactos práticos na gestão universitária; e a Seção 5 conclui com as lições aprendidas, limitações e perspectivas de trabalhos futuros.

## 2. Contexto e Problema Público

A evasão no ensino superior é um fenômeno multifacetado, influenciado por dimensões individuais, acadêmicas, institucionais e externas [Barragán Moreno and González Támara, 2024]. No contexto das Instituições de Ensino Superior (IES) públicas brasileiras, esse problema ganha contornos críticos, uma vez que a vacância de vagas representa uma subutilização de investimentos estatais e um entrave à meta de democratização do acesso e permanência no ensino superior.

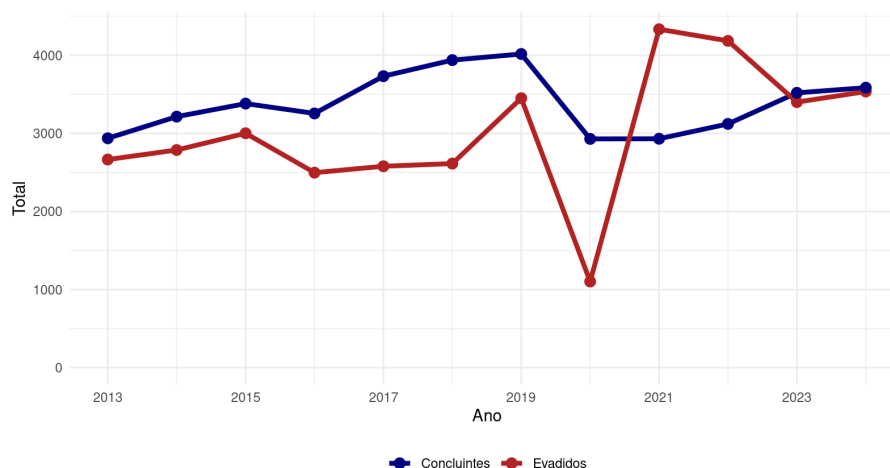
Historicamente, a UFPR apresentava taxas de evasão relativamente estáveis. Contudo, ao analisar a série histórica apresentada na Figura 1, nota-se um comportamento revelador: no período pré-pandemia, a quantidade de evasões anuais mantinha-se constante, apresentando flutuações marginais. Entretanto, um achado fundamental desta investigação é que já em 2019 existiam indícios de que a evasão subiria, com um crescimento incipiente que rompeu a estabilidade dos anos anteriores.

Com a chegada da crise sanitária, esse movimento de ascensão foi severamente catalisado. No cenário pós-pandemia, em especial nos anos de 2021 e 2022, a instituição deparou-se com a maior evasão da última década, ultrapassando a marca de 4.000 desligamentos anuais. Embora os dados mais recentes indiquem uma trajetória de melhora, os índices permanecem em um patamar superior ao observado na década passada. Essa nova realidade evidencia que os desafios de permanência tornaram-se mais complexos, exigindo uma resposta institucional que vá além da intuição administrativa e se baseie em evidências robustas.



**Figura 1. Série histórica do total de evadidos por ano (2013-2024) na UFPR. Linha cinza: série observada. Linha azul: tendência obtida via regressão suavizada**

Um resultado que desperta atenção especial surge ao compararmos a série de evasão em paralelo com a série de conclusão de curso. Historicamente, a universidade mantinha um saldo positivo, com um número de concluintes superior ao de evadidos. Todavia, no biênio 2021-2022, esse padrão se inverteu: pela primeira vez na década, o volume de alunos que abandonaram a instituição superou o de diplomados. Embora a curva de conclusões tenha retomado a liderança recentemente, a margem estreita sinaliza uma pressão contínua sobre a eficiência da formação pública.



**Figura 2. Série histórica do total de evadidos e concluintes por ano (2013-2024) na UFPR**

O desafio central identificado foi a natureza reativa das políticas existentes. A falta de ferramentas de monitoramento e predição limita a capacidade de intervenção oportuna. Tradicionalmente, o desligamento é processado como um evento administrativo consumado, impedindo ações de acolhimento enquanto o estudante ainda mantém vínculo, mas já apresenta sinais de vulnerabilidade.

Para enfrentar esse problema, a instituição identificou a necessidade de desenvolver mecanismos para: (i) identificação precoce de comportamentos de abandono; (ii) entendimento de motivações subjetivas não capturadas por sistemas cadastrais; e (iii) otimização do direcionamento de recursos de assistência estudantil. Nesse sentido, o problema público não é apenas a evasão, mas a carência de ferramentas de inteligência que suportem uma governança digital proativa.

### 3. Metodologia: O Framework do Observatório

A metodologia adotada para a concepção do Observatório de Dados de Evasão e Baixa Procura fundamenta-se em um *framework* composto por quatro pilares interdependentes: (i) Monitoramento; (ii) Predição; (iii) Pesquisa; e (iv) Divulgação. Esta estrutura foi desenhada para permitir que a universidade transite de uma análise meramente descritiva para uma atuação preditiva e prescritiva.

#### 3.1. Pilar 1: Monitoramento e Inteligência de Dados

O primeiro pilar do Observatório consiste na estruturação de um ecossistema de Inteligência de Negócios (BI) para a visualização dinâmica dos indicadores de fluxo discente.

O painel de monitoramento atual oferece uma visão granular através de filtros por ano, setor acadêmico e curso. A interface é composta por:

- **Indicadores Agregados (*Big Numbers*):** Visualização direta do total de matriculados, ingressantes, concluintes e evadidos para o recorte selecionado.
- **Análise por Segmentação:** Gráficos que detalham a evasão em função do turno, grau acadêmico (bacharelado, licenciatura, entre outros) e setor.
- **Ranking de Evasão:** Tabelas dinâmicas que apresentam a taxa média de evasão por curso, permitindo a identificação imediata de graduações com índices críticos.
- **Séries Temporais:** Gráficos evolutivos que mostram a variação da taxa de evasão e do número absoluto de evadidos ao longo do tempo.

Complementarmente, o monitoramento estende-se à dimensão da Baixa Procura, permitindo analisar a relação candidato/vaga por curso e sua evolução histórica. Essa funcionalidade é vital para identificar cursos que enfrentam crises de atratividade, o que frequentemente precede o aumento das taxas de evasão.

### 3.2. Pilar 2: Modelagem Preditiva e a Calculadora de Risco

Este pilar constitui o núcleo de inteligência analítica do Observatório, focado na antecipação do risco de evasão individual antes do desfecho do desligamento. A modelagem foi fundamentada em uma análise comparativa de algoritmos de Aprendizado de Máquina, a exemplo das revisões sistemáticas e estudos de caso conduzidos por Magalhães et al. [2025], de Almeida Teodoro and Kappel [2020] e Silvestri et al. [2025].

A presente seção trata do conjunto de dados utilizado, as variáveis preditoras selecionadas, e descreve concisamente os métodos aplicados.

#### 3.2.1. Dataset

Para a aplicação das metodologias, foi consolidado um conjunto de dados da UFPR abrangendo o período de 2009 a 2024, período que inclui a utilização dos sistemas acadêmicos SIE e SIGA em que havia acesso a dados digitalizados. A base de dados final, extraída e unificada, contém 107.069 observações de registros discentes anonimizados. A variável resposta *Evasão* foi definida de forma binária: valor 1 para alunos que evadiram (abandono) e 0 para alunos que concluíram o curso (formatura), desconsiderando-se observações de alunos que ainda não tiveram sua situação acadêmica definida no período.

É importante salientar que, no sistema de onde os dados foram coletados, existem várias formas de evasão cadastradas. No entanto, para o estudo em questão, apenas duas foram consideradas: Abandono, a forma mais crítica de evasão e representa mais de 40% de todas as formas; e Formatura, que indica os alunos que de fato concluíram o curso.

Devido à maioria de casos de não evasão na base original, o dataset apresentou um forte desbalanceamento de classes. Para mitigar o viés do modelo em favor da classe majoritária, foi aplicada a técnica de Subamostragem Aleatória (Random Under-Sampling - *RUS*) no conjunto de treinamento. O *RUS* consistiu em sortear aleatoriamente um número de observações da classe majoritária (Não Evadidos) igual ao número de observações da classe minoritária (Evadidos), resultando em um dataset balanceado

de 23.134 observações, divididas em 70% para treinamento e 30% para teste. Esta estratégia foi escolhida pela sua simplicidade e eficácia em reduzir o tempo de treinamento dos modelos.

Dentre uma ampla gama de possíveis variáveis a serem utilizadas na modelagem, várias foram testadas e ao final foram selecionadas apenas: IRA, PropRep, Setor, Grau e Turno.

O IRA (Índice de Rendimento Acumulado) do aluno é uma medida de desempenho acadêmico utilizada na UFPR, variando no intervalo de 0 a 1 e calculado através de uma normalização da média das notas do aluno (variando de 0 a 100) e a carga horária. Já a variável PropRep diz respeito à Proporção de Reprovações e é dada pela razão entre disciplinas reprovadas e o total de disciplinas cursadas. Quanto às variáveis identificadoras do curso, foram utilizadas: Setor (unidade acadêmica ao qual o curso pertence), turno (manhã, tarde, noite, integral) e grau (ABI, Bacharelado, Licenciatura, Tecnólogo).

Variáveis socioeconômicas (origem do questionário do vestibular) e de consumo no Restaurante Universitário foram excluídas devido à inviabilidade de compatibilização longitudinal (para socioeconômicas) ou por apresentarem alta taxa de dados faltantes (para o Restaurante Universitário), o que limitaria a generalização do modelo.

### 3.2.2. Métodos de Aprendizado de Máquina

O Aprendizado de Máquina concentra-se no desenvolvimento de algoritmos voltados à realização de previsões ou ao suporte à tomada de decisões, fundamentando-se em diversas classes de modelos consolidadas na literatura. Entre as metodologias aplicadas, a Regressão Logística (LR) destaca-se como um modelo linear que estima a probabilidade de evasão por meio da função *logit* [Morgan and Teachman, 1988]. As Árvores de Decisão (DT), por sua vez, particionaram os dados através de regras hierárquicas ajustadas por um parâmetro de complexidade de 0,01 para evitar o superajuste [Talekar, 2020]. Para aumentar a robustez preditiva, empregou-se a Floresta Aleatória (RF), um método de *ensemble* que combina 100 árvores independentes construídas sob amostragens aleatórias, conforme a técnica de Breiman et al. [1984]. Outra abordagem avaliada foi o *Boosting* de Gradiente Estocástico (GBM), configurado com profundidade de interação igual a 3 e taxa de aprendizado de 0,1 para corrigir sequencialmente os resíduos das árvores anteriores [Jun, 2021]. Adicionalmente, as Máquinas de Vetores de Suporte (SVM) buscaram um hiperplano ótimo para a separação das classes utilizando uma função *kernel* radial com parâmetros de custo e *sigma* otimizados [Vapnik, 1998, Liu, 2024]. Por fim, as Redes Neurais (NN) aplicaram modelos de Aprendizado Profundo com uma camada oculta de 5 neurônios e decaimento de 0,1 para identificação de padrões complexos [Sothe et al., 2020]. O ajuste dos modelos foi conduzido por validação cruzada de 10 dobras, utilizando a métrica AUC para otimização. Para um aprofundamento teórico sobre estas metodologias, recomenda-se a consulta à obra de Hastie et al. [2010].

### 3.2.3. Avaliação do Modelo

Para avaliar o desempenho dos modelos no cenário de evasão, foram utilizadas métricas que oferecem diferentes perspectivas quanto a acertos e erros que os modelos cometem. A Acurácia (*ACC*) expressa a taxa global de acertos do classificador, enquanto a Sensibilidade (*SENS*) mede a capacidade de identificar corretamente os alunos que de fato evadiram (Verdadeiros Positivos), o que é vital para viabilizar ações preventivas. Paralelamente, a Precisão (*PREC*) aponta a proporção de evadidos reais dentre os alertas gerados, mitigando o desperdício de recursos institucionais com alarmes falsos. Por fim, o Coeficiente de Correlação de Matthews (*MCC*) sintetiza a qualidade global da predição em uma escala de -1 a 1. As medidas de desempenho calculadas no conjunto de teste (30% dos dados não usados na fase de treinamento) encontram-se na Tabela 1:

**Tabela 1. Medidas comparativas entre as técnicas de Aprendizado de Máquina. Valores máximos em negrito**

| Modelo | ACC           | SENS          | PREC          | MCC           |
|--------|---------------|---------------|---------------|---------------|
| LR     | 0,8489        | 0,8004        | <b>0,8864</b> | 0,7011        |
| DT     | 0,8363        | 0,8871        | 0,7707        | 0,6784        |
| RF     | 0,8482        | 0,8808        | 0,8053        | 0,6989        |
| GBM    | 0,8513        | 0,8826        | 0,8104        | 0,7050        |
| SVM    | 0,8524        | <b>0,8923</b> | 0,8015        | 0,7084        |
| NN     | <b>0,8565</b> | 0,8841        | 0,8205        | <b>0,7147</b> |

A análise comparativa aponta um desempenho homogêneo entre as técnicas avaliadas. O modelo de Redes Neurais alcançou a maior Acurácia (0,8565) e *MCC* (0,7147), seguido de perto pelo *SVM*, que também se sobressaiu em Sensibilidade. Contudo, o *SVM* demandou o maior tempo de processamento, o que eleva o custo computacional para fins de implantação.

Por outro lado, a Regressão Logística obteve a maior Precisão (0,8864). Esta métrica é crítica para o cenário prático, pois garante a confiabilidade dos alertas gerados ao minimizar os alarmes falsos. Além disso, o *MCC* da Regressão Logística mostrou-se altamente competitivo se comparado aos modelos de Redes Neurais e *SVM*, considerados estruturalmente mais complexos.

Diante disso, a Regressão Logística destaca-se como a solução mais robusta para o ambiente institucional. Além do equilíbrio estatístico e da baixa carga computacional, sua interpretabilidade nativa atende ao requisito de transparência gerencial, permitindo que coordenadores e gestores compreendam os fatores exatos que colocam um estudante em risco.

A aplicação prática do modelo consolida-se na “Calculadora de Risco de Evasão”. Diferente de painéis estáticos, a ferramenta fornece uma lista priorizada e dinâmica em forma de relatório diretamente no sistema de gestão acadêmica, permitindo que a gestão universitária execute intervenções pedagógicas e de assistência estudantil proativas, otimizando a alocação de recursos públicos na retenção discente.

### 3.3. Pilar 3: Pesquisa Qualitativa e Delineamento Metodológico

O terceiro pilar reconhece que, embora os modelos preditivos identifiquem quem possui risco de evasão, as pesquisas qualitativas são fundamentais para compreender o porquê. O Observatório estruturou, baseado em Felisbino [2025], um programa de escuta ativa dividido em duas frentes complementares: a Pesquisa com Evadidos e o Acompanhamento de Egressos.

A Pesquisa com Evadidos investiga as motivações para o abandono através de um esquema conceitual que correlaciona fatores de ordem estrutural, socioeconômica, institucional e pessoal. A edição inaugural, realizada em 2024, contou com a participação de 1.783 respondentes. O instrumento de coleta é organizado em nove seções temáticas, que abrangem desde a trajetória acadêmica prévia até dimensões de saúde mental, dificuldades financeiras e percepção sobre a rigidez curricular. Esta granularidade permite identificar, por exemplo, se a evasão em determinados setores está mais vinculada a questões de infraestrutura ou a conflitos entre os horários das aulas e as exigências do mercado de trabalho.

Em paralelo, o Acompanhamento de Egressos adota um desenho longitudinal inédito na instituição, com o objetivo de mapear a eficácia da formação universitária a partir de três eixos fundamentais. No que tange ao Trabalho e Renda, o projeto realiza o monitoramento do vínculo empregatício, do tempo necessário para a inserção no primeiro emprego e da faixa salarial, adotando padrões de questionários do IBGE. Quanto à Adequação Formativa, avalia-se a percepção do egresso sobre a contribuição direta da formação acadêmica para a sua empregabilidade. Por fim, no eixo de Formação Continuada, busca-se o mapeamento do ingresso desses profissionais em programas de pós-graduação, consolidando uma visão abrangente sobre o impacto da universidade na trajetória de seus ex-alunos.

Um diferencial estratégico deste pilar é a integração tecnológica: os novos questionários são desenhados para carregar dados automaticamente via sistema de gestão acadêmica, minimizando a fadiga do respondente e garantindo maior precisão na análise cruzada entre os dados administrativos e as percepções subjetivas.

### 3.4. Pilar 4: Divulgação, Transparência e Hub de Conhecimento

O pilar final do framework materializa o compromisso com a transparência pública e a disseminação de evidências por meio de um portal *web* centralizador (<https://ufpr.br/observaevasao/>), que funciona como a face pública e o ecossistema de Governança Digital do Observatório. Este ambiente integra ações de Transparência Ativa ao disponibilizar as versões abertas dos painéis de monitoramento de evasão e baixa procura, além de atuar como um Hub de Conhecimento e Repositório Científico que reúne a produção acadêmica da instituição sobre permanência estudantil. Em alinhamento com os preceitos de Ciência Aberta, a página disponibiliza publicamente toda a documentação técnica, os materiais de divulgação (incluindo apresentações em formato PDF e vídeos explicativos) e os códigos de programação em R e Python armazenados no GitHub, assegurando a auditabilidade e a replicabilidade metodológica dos modelos preditivos por outras Instituições de Ensino Superior.

A integração contínua de todo esse ecossistema assegura que o Observatório não seja apenas uma iniciativa isolada, mas uma estratégia consolidada de gestão pública que



une dados, pessoas e transparência para fortalecer de ponta a ponta a universidade pública.

#### 4. Resultados e Discussão

A implementação do Observatório de Dados de Evasão e Baixa Procura permitiu à instituição transitar de um modelo de reporte passivo e reativo para uma estratégia de governança ativa e baseada em evidências. Os resultados demonstram que a união entre a Estatística, a Ciência de Dados e a sensibilidade humana é capaz de oferecer respostas robustas a um desafio que é, acima de tudo, estrutural e social.

Os painéis de monitoramento, primeiro pilar do Observatório, democratizaram o acesso aos dados institucionais por meio de filtros intuitivos. A ferramenta funciona como um diagnóstico contínuo que permite aos diretores de setor e coordenadores de curso identificar, em tempo real, a efetividade das políticas de permanência e os pontos que demandam ajustes urgentes.

No campo da predição, a Calculadora de Risco de Evasão, produto baseado em Aprendizado de Máquina, consolidou-se como o braço tecnológico mais inovador do projeto. O modelo de Regressão Logística, escolhido por sua transparência e facilidade de auditoria, apresentou métricas de performance consideradas robustas para o contexto de políticas públicas.

Contudo, o diferencial desta ferramenta não reside apenas na precisão algorítmica, mas na sua integração ao fluxo cotidiano da instituição por meio do sistema de gestão acadêmica. A Calculadora atua como um sinalizador proativo. Ao identificar um estudante com alto risco de abandono, o sistema alerta o coordenador de curso não para substituir o contato humano, mas para mediá-lo. O dado serve como bússola para que a coordenação possa oferecer uma orientação pedagógica, auxílio ou um acolhimento psicossocial antes que o vínculo acadêmico seja rompido.

Além disso, a soberania tecnológica foi um pilar inegociável do resultado: todo o desenvolvimento do modelo preditivo foi pautado no uso de software livre e código aberto utilizando linguagens R [R Core Team, 2024] e Python [Python Core Team, 2026]. O compromisso com o código aberto no GitHub garante que a tecnologia produzida pela universidade seja pública, transparente e auditável pela comunidade acadêmica e pela sociedade civil, permitindo sua replicabilidade em outras IES. Além disso, todo o procedimento foi pensado para atender a LGPD, desde o uso de somente dados necessários até a estratégia de controle de acesso às ferramentas.

Para além dos algoritmos, o Observatório avançou na humanização das estatísticas através do pilar de Pesquisa. Ao instituir um programa de escuta ativa com evadidos e egressos e dando continuidade ao trabalho de Felisbino [2025], a instituição começou a mapear as motivações subjetivas que os dados cadastrais são incapazes de captar. O delineamento longitudinal das pesquisas permite dar rosto e voz aos dados, transformando a frieza das planilhas em evidências sobre o descompasso entre as expectativas do ingressante e a realidade do mercado ou da vida universitária.

A pesquisa com egressos, especificamente, adiciona uma camada de análise sobre a eficácia da formação, monitorando eixos como vínculo empregatício e adequação formativa de acordo com padrões do IBGE. Esses achados permitem que as políticas de assistência estudantil ataquem barreiras reais de comunicação e burocracia que dificultam

o acesso aos auxílios existentes.

Por fim, o lançamento do Portal do Observatório materializa o compromisso com a transparência ativa e a Governança Digital. O site funciona como um *hub* de conhecimento que centraliza a produção acadêmica da instituição sobre permanência e sucesso estudantil. Ao entregar publicamente a descrição da proposta metodológica, os painéis interativos e a base científica para pesquisas de campo, o Observatório assegura que o conhecimento gerado tenha valor social e sirva de modelo replicável.

Em última análise, a experiência do Observatório demonstra que a inteligência de dados e as ferramentas digitais fornecem a direção técnica necessária para a gestão. Entretanto, é a capacidade institucional de acolher e a integração de esforços entre as diversas unidades que definem o destino de cada estudante, garantindo que a universidade pública continue sendo um espaço de transformação de trajetórias.

## 5. Conclusão

A concepção e a implementação do Observatório de Dados de Evasão e Baixa Procura representam um marco estratégico na transição da instituição em direção a uma gestão pública orientada por evidências. Mais do que o desenvolvimento de ferramentas tecnológicas isoladas, o projeto consolidou um ecossistema sociotécnico capaz de converter o volume massivo de registros administrativos em inteligência acionável. Esta iniciativa permitiu que a universidade abandonasse uma postura puramente reativa e passasse a atuar de forma proativa na proteção das trajetórias acadêmicas de seus estudantes, mitigando o desperdício de recursos públicos e a frustração de projetos de vida.

As contribuições deste trabalho sintetizam-se em três dimensões fundamentais. Na técnica, a validação da Calculadora de Risco comprovou a viabilidade de modelos de Machine Learning no setor público, priorizando o software livre para garantir soberania tecnológica e algoritmos auditáveis. Na administrativa, os painéis de monitoramento democratizaram o acesso à informação estratégica, permitindo que gestores realizem diagnósticos precisos e céleres sobre a saúde acadêmica dos cursos. Por fim, na social, o pilar de pesquisa promoveu a humanização dos dados, assegurando que motivações subjetivas e vulnerabilidades estudantis orientem, de fato, as políticas de permanência e assistência.

Apesar dos avanços significativos, o trabalho apresenta limitações que balizam as próximas etapas de evolução do ecossistema. Dentre eles, a integração da predição com os fluxos efetivos de atendimento nas coordenações carece de uma avaliação longitudinal de longo prazo para medir o impacto direto da ferramenta na redução percentual das taxas de evasão ao longo dos próximos períodos letivos.

Como perspectivas para trabalhos futuros, prevê-se a expansão do monitoramento para incluir indicadores de retenção e o cruzamento detalhado com modalidades de cotas e perfis socioeconômicos. Planeja-se, ainda, o refinamento dos modelos preditivos para incluir variáveis de sentimento extraídas de questões abertas das pesquisas qualitativas, bem como a automação de notificações personalizadas para estudantes em zona de risco.

Em suma, o Observatório demonstra que a inteligência de dados, quando aliada à sensibilidade humana e ao compromisso com a transparência, é uma aliada indispensável na defesa da universidade pública, garantindo que o acesso ao ensino superior seja acom-

panhado de condições reais para a diplomação e para a plena transformação social através da educação.

## Referências

- S. P. Barragán Moreno and L. González Támara. Complexities of student dropout in higher education: A multidimensional analysis. In *Frontiers in Education*, volume 9, page 1461650. Frontiers Media SA, 2024.
- L. Breiman, J. Friedman, R. Olshen, and C. J. Stone. *Classification and Regression Trees*. 1984.
- L. de Almeida Teodoro and M. A. A. Kappel. Aplicação de técnicas de aprendizado de máquina para predição de risco de evasão escolar em instituições públicas de ensino superior no brasil. *Revista Brasileira de Informática na Educação*, 28:838–863, 2020.
- D. Enguídanos, J. Aroztegui, M. Iglesias-Soilán, I. Sánchez-San-José, and J. Fernández. Academic emotions and regulation strategies: Interaction with higher education dropout ideation. *Education Sciences*, 13(11):1152, 2023.
- E. Felisbino. “INFELIZMENTE, A FEDERAL NÃO É PARA TODOS”: DESVELANDO DESIGUALDADES NA EDUCAÇÃO SUPERIOR A PARTIR DA PERCEPÇÃO DO ESTUDANTE EVADIDO DOS CURSOS DA UFPR. *Jornal de Políticas Educacionais*, 19(1), dez. 2025. doi: 10.5380/jpe.v19i1.98670. URL <https://revistas.ufpr.br/jpe/article/view/98670>.
- T. Hastie, R. Tibshirani, and J. Friedman. *The Elements of Statistical Learning: Data Mining, Inference, Prediction*. Springer, 2010.
- M.-J. Jun. A comparison of a gradient boosting decision tree, random forests, and artificial neural networks to model urban land use changes: the case of the Seoul metropolitan area. *International Journal of Geographical Information Science*, 35(11):2149–2167, mar 1 2021.
- X. Liu. Comparison of different machine learning models: Linear model, forest and SVM. *Applied and Computational Engineering*, 51(1):225–230, mar 25 2024.
- L. F. D. Lopes, D. J. C. da Silva, N. Kuhn, F. V. Chiapinoto, and M. P. Lima. The influence of technostress on anxiety disorder in higher education students during the covid-19 pandemic. *Knowledge Management & E-Learning*, 16(1):164–185, 2024.
- P. V. S. Magalhães, R. A. da Silva Sousa, A. C. B. Santos, F. J. M. Aguiar, and R. T. P. Milfont. Modelos de inteligência artificial para predição de evasão escolar: Uma revisão sistemática. *Revista Eletrônica de Iniciação Científica em Computação*, 23:48–54, 2025.
- S. P. Morgan and J. D. Teachman. Logistic Regression: Description, Examples, and Comparisons. *Journal of Marriage and the Family*, 50(4):929, 11 1988.
- Nurmalitasari, Z. Awang Long, and M. Faizuddin Mohd Noor. Factors influencing dropout students in higher education. *Education Research International*, 2023(1): 7704142, 2023.
- Python Core Team. *Python: A dynamic, open source programming language*. Python Software Foundation, 2026. URL <https://www.python.org/>.
- R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2024. URL <https://www.R-project.org/>.
- V. B. Salakhova, L. V. Shukshina, N. V. Belyakova, A. V. Kidinov, N. S. Morozova, and N. V. Osipova. The problems of the covid-19 pandemic in higher education. In *Frontiers in education*, volume 7, page 803700. Frontiers Media SA, 2022.

- L. M. B. Sampaio et al. Student retention policies in higher education: Reflections from a literature review for the brazilian context. *Revista de Administração Pública*, 56: 603–631, 2022.
- I. Sharaievska, O. McAnirlin, M. H. Browning, L. R. Larson, L. Mullenbach, A. Rigolon, A. D’Antonio, S. Cloutier, J. Thomsen, E. C. Metcalf, et al. “messy transitions”: Students’ perspectives on the impacts of the covid-19 pandemic on higher education. *Higher education*, pages 1–18, 2022.
- F. Silvestri, V. F. de Souza, and A. Vieira. Mineração de dados educacionais: uma análise sobre os preditores da evasão no ensino superior. *Revista Educar Mais*, 9:1–23, 2025.
- C. Sothe, D. Almeida, et al. Comparative performance of convolutional neural network, weighted and conventional support vector machine and random forest for classifying tree species using hyperspectral and photogrammetric data. *GIScience & Remote Sensing*, 57(3):369–394, jan 12 2020.
- B. Talekar. A Detailed Review on Decision Tree and Random Forest. *Bioscience Biotechnology Research Communications*, 13(14):245–248, dec 25 2020.
- V. Vapnik. *Statistical learning theory*. John Wiley Sons, 1998.