PROFIAP
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO PÚBLICA

UNIVERSIDADE FEDERAL DA FRONTEIRA SUL
CAMPUS CHAPECÓ



PROTOCOLO DE IMPLEMENTAÇÃO DE UM SISTEMA DE ALERTA PRECOCE DE EVASÃO ACADÊMICA NA UFFS CAMPUS CHAPECÓ

Diretrizes para o uso institucional periódico de modelos de Mineração de Dados e Machine Learning

PROTOCOLO DE IMPLEMENTAÇÃO DE UM SISTEMA DE ALERTA PRECOCE DE EVASÃO ACADÊMICA NA UFFS CAMPUS CHAPECÓ

Relatório técnico apresentado pelo mestrando Jean Carlo Rodio ao Mestrado Profissional em Administração Pública em Rede Nacional (PROFIAP), sob orientação do Prof. Dr. Moacir Francisco Deimling e coorientação do Prof. Dr. Giancarlo Dondoni Salton, como parte dos requisitos para obtenção do título de Mestre em Administração Pública.



SUMÁRIO

Contexto de aplicação da proposta 04

Público-alvo da proposta 05

Situação problema 06

Objetivo da proposta 08

PTT - Proposta de Intervenção 09

Autoria da Proposta 11

Referências 12

Protocolo de Entrega da Proposta aos Praticantes 13

CONTEXTO DE APLICAÇÃO DA PROPOSTA

A presente proposta de intervenção insere-se no contexto da Universidade Federal da Fronteira Sul (UFFS), campus Chapecó, instituição federal de ensino superior criada em 2009 com a missão de democratizar o acesso à educação superior na Mesorregião Grande Fronteira do Mercosul. Entre 2015 e 2023, a Taxa de Sucesso na Graduação da UFFS manteve-se em torno de 33%, patamar que o Relatório de Auditoria Operacional nº 02/AUDIN/UFFS/2025 classificou como tema de alto risco institucional, associado à ausência de mecanismos sistemáticos de monitoramento e identificação precoce de estudantes em risco de evasão (UFFS, 2025). No mesmo sentido, o Tribunal de Contas da União, por meio do Acórdão nº 1868/2024, recomendou às instituições federais de ensino superior a estruturação de políticas de prevenção à evasão fundamentadas em evidências e sustentadas por sistemas de monitoramento contínuo dos indicadores acadêmicos (TCU, 2024).

A proposta decorre diretamente da dissertação de mestrado intitulada "Aplicação de Técnicas de Mineração de Dados e *Machine Learning* para Predição da Evasão de Estudantes nos Cursos de Graduação de uma IFES", desenvolvida no âmbito do Mestrado Profissional em Administração Pública em Rede Nacional (PROFIAP), campus Chapecó, e posiciona-se no momento institucional em que a UFFS ainda carece de instrumento preditivo consolidado para subsidiar a gestão da permanência estudantil.

Do ponto de vista de seu alinhamento com a agenda internacional de desenvolvimento sustentável, a proposta dialoga diretamente com o Objetivo de Desenvolvimento Sustentável (ODS) 4 - Educação de Qualidade, da Organização das Nações Unidas, ao contribuir para a redução do abandono e para a garantia de conclusão dos estudos superiores. De forma complementar, ao evidenciar o papel de indicadores de engajamento institucional e de assistência estudantil - como o uso do Restaurante Universitário - na predição da evasão, a proposta também se relaciona ao ODS 10 - Redução das Desigualdades, na medida em que pode subsidiar políticas de permanência voltadas a estudantes oriundos de contextos socioeconômicos mais vulneráveis.

> a proposta dialoga diretamente com o Objetivo de Desenvolvimento Sustentável (ODS) 4 - Educação de Qualidade, da Organização das Nações Unidas, ao contribuir para a redução do abandono e para a garantia de conclusão dos estudos superiores.

PÚBLICO-ALVO DA PROPOSTA

A proposta de intervenção destina-se, em primeiro lugar, à Pró-Reitoria de Graduação da Universidade Federal da Fronteira Sul (PROGRAD/UFFS), instância institucional com competência e autonomia para deliberar sobre a incorporação do protocolo aqui descrito aos fluxos de gestão da permanência estudantil, configurando-se como praticante-líder desta proposta.

A operacionalização técnica do protocolo depende, ainda, da equipe técnica da Secretaria Especial de Tecnologia da Informação (SETI) da UFFS, responsável pela infraestrutura dos sistemas SIGAA, SCI, SAS e Pergamum, e destina-se, em seus resultados finais, às coordenações de curso e às equipes de assistência estudantil do campus Chapecó, responsáveis por interpretar as estimativas de risco e conduzir, com autonomia pedagógica, as intervenções de acompanhamento.

DADOS

PRATICANTE-LÍDER

Pró-Reitoria de Graduação da UFFS (PROGRAD) - delibera sobre a incorporação institucional do protocolo.

EXECUÇÃO TÉCNICA

Secretaria Especial de Tecnologia da Informação (SETI/UFFS) - executa o *pipeline* e o modelo sobre os dados institucionais.

USUÁRIOS FINAIS

Coordenações de curso e equipes de assistência estudantil do campus Chapecó - interpretam os resultados e conduzem as intervenções.







SITUAÇÃO PROBLEMA

Diagnóstico, análise e linha de intervenção

Diagnóstico

O diagnóstico que fundamenta esta proposta foi conduzido ao longo da dissertação de mestrado que a antecede e pode ser sintetizado em três constatações centrais. Primeiro, a UFFS convive, desde sua criação, com taxas de sucesso na graduação estruturalmente baixas: a média histórica de 33,11% entre 2015 e 2023 significa que menos de um terço dos estudantes ingressantes concluem o curso no tempo previsto (UFFS, 2025). Segundo, essa taxa não é acompanhada de mecanismos institucionais de identificação precoce: o Relatório de Auditoria Operacional nº 02/AUDIN/UFFS/2025 constatou a ausência de política formal e de sistema contínuo de monitoramento da evasão na instituição, limitação que o TCU (2024), identificou como recorrente em 51% das universidades federais brasileiras. Terceiro, a UFFS dispõe, ainda que de forma não integrada, de um volume expressivo de dados institucionais, acadêmicos, de assistência estudantil e de uso de serviços, que, até a realização deste estudo, não haviam sido explorados de forma sistemática para fins preditivos.

Análise

A análise que subsidiou a linha de intervenção seguiu a metodologia CRISP-DM (*Cross Industry Standard Process for Data Mining*) e integrou registros de quatro sistemas institucionais - SIGAA, SCI, SAS e Pergamum -, resultando em um conjunto analítico de 235.632 registros referentes a 8.749 matrículas de graduação do campus Chapecó, no período de 2014.1 a 2024.2. Sobre esse conjunto de dados foram treinados e comparados seis algoritmos de aprendizado supervisionado - Regressão Logística, Árvore de Decisão, *Random Forest*, XGBoost, SVM Linear e KNN, validados por meio do protocolo *StratifiedGroupKFold*, que assegura a separação estrita entre os conjuntos de treino e teste por matrícula e evita o vazamento de dados temporais entre observações de um mesmo estudante.

O modelo XGBoost apresentou o melhor desempenho global (AUC-ROC = 0,8878), seguido pela Regressão Logística (AUC-ROC = 0,8681) e pelo SVM Linear (AUC-ROC = 0,8680), o que demonstra que os dados já disponíveis nos sistemas da UFFS sustentam modelos com capacidade discriminativa robusta para identificar estudantes em risco de evasão. A análise de importância de variáveis revelou, ainda, que o desempenho acadêmico predomina na estrutura preditiva: a média final nas disciplinas respondeu por aproximadamente um terço do ganho informacional total do modelo, seguida por indicadores de ausência de registro acadêmico, número de faltas e área de conhecimento, ao passo que o uso do Restaurante Universitário nos primeiros semestres emergiu como preditor relevante de engajamento institucional, superando variáveis sociodemográficas como gênero e etnia.

Linha de intervenção

A linha de intervenção escolhida para converter esse diagnóstico em ação institucional consistiu na disponibilização de um *pipeline* de pré-processamento e de um modelo preditivo prontos para operacionalização periódica pela própria equipe técnica da UFFS, em vez de alternativas como a aquisição de sistemas comerciais de terceiros ou a criação de uma unidade dedicada exclusivamente à predição de evasão. Essa escolha justifica-se por três razões: primeiro, porque os recursos técnicos necessários à execução do *pipeline* (*Python* e bibliotecas de código aberto) já estão disponíveis na SETI/UFFS; segundo, porque a documentação integral do processo de tratamento de dados, codificação e modelagem, produzida ao longo da dissertação, permite a reprodução independente do processo sem custos de licenciamento; terceiro, porque a manutenção do conhecimento técnico internamente à instituição favorece a continuidade do protocolo.

OBJETIVO DA PROPOSTA

Esta proposta tem por objetivo dar materialidade à linha de intervenção definida a partir do diagnóstico e da análise apresentados na seção anterior, por meio da instituição de um protocolo de geração periódica de estimativas de risco de evasão para as matrículas ativas dos cursos de graduação do campus Chapecó da UFFS, a ser incorporado aos fluxos de gestão acadêmica da Pró-Reitoria de Graduação.

De modo específico, pretende-se: a) viabilizar a execução semestral do *pipeline* de pré-processamento e do modelo preditivo XGBoost sobre os dados extraídos dos sistemas SIGAA, SCI, SAS e Pergamum; b) subsidiar as coordenações de curso e as equipes de assistência estudantil com estimativas objetivas de risco de evasão por matrícula, organizadas em faixas de prioridade; c) orientar a calibração institucional dos limiares de corte que definem as faixas de risco, de acordo com a capacidade operacional de acompanhamento da instituição; d) assegurar que a implementação do protocolo observe a Lei Geral de Proteção de Dados Pessoais (LGPD), com pseudonimização dos identificadores durante o processamento técnico, acesso restrito aos profissionais autorizados e reidentificação somente quando necessária à realização do acompanhamento pedagógico ou assistencial.

> "O escore de probabilidade de evasão é uma estimativa estatística, não uma certeza, devendo ser interpretado pelas coordenações de curso e equipes pedagógicas à luz do conhecimento que detêm sobre a realidade específica de cada estudante e de cada curso."

PTT - PROPOSTA DE INTERVENÇÃO

Recursos técnicos disponibilizados

Como parte concreta desta proposta, disponibilizam-se três recursos técnicos articulados, resultantes diretamente da dissertação que a fundamenta: a) o dataset consolidado, com 235.632 registros acadêmicos referentes a 8.749 matrículas do campus Chapecó no período de 2014.1 a 2024.2, integrando informações do SIGAA, do SCI, do SAS e do Pergamum, tratado e documentado conforme os procedimentos de qualidade de dados descritos na dissertação; b) o *pipeline* de pré-processamento, implementado em Python com as bibliotecas pandas, numpy e scikit-learn, executável pela equipe técnica da SETI/UFFS sobre novos dados extraídos a cada semestre letivo, reproduzindo de forma automatizada as etapas de tratamento, codificação e engenharia de variáveis documentadas; c) o modelo XGBoost treinado e validado (AUC-ROC = 0,8878), aplicável sobre o conjunto de dados preparado para gerar, a cada semestre, uma estimativa da probabilidade de evasão de cada matrícula ativa.

Protocolo de execução periódica

Recomenda-se que a Pró-Reitoria de Graduação avalie a incorporação da execução do *pipeline* e do modelo ao calendário institucional, segundo o seguinte ciclo: ao final de cada semestre letivo, a SETI/UFFS executa o *pipeline* sobre os dados atualizados dos quatro sistemas institucionais e aplica o modelo XGBoost às matrículas ativas; os resultados são consolidados em relatório de risco por matrícula e disponibilizados às coordenações de curso e às equipes de assistência estudantil no início do semestre subsequente, de modo que as informações utilizadas reflitam sempre a trajetória acadêmica mais recente de cada estudante.

Protocolo de execução periódica

Os limiares de corte que definem a classificação de cada matrícula em alto, médio ou baixo risco devem ser calibrados pela própria instituição, em função de sua capacidade operacional de acompanhamento, podendo ser ajustados ao longo dos primeiros ciclos de uso com base nos resultados observados. Recomenda-se, adicionalmente, que a UFFS reavalie periodicamente o desempenho do modelo, sobretudo após mudanças curriculares relevantes ou alterações expressivas no perfil dos ingressantes, de modo a assegurar que os padrões históricos sobre os quais o modelo foi treinado permaneçam representativos da realidade institucional.

Protocolo de execução periódica

A aplicação do modelo preditivo deve ser compreendida exclusivamente como recurso de apoio à decisão dos gestores acadêmicos, e não como mecanismo autônomo de classificação de estudantes. O tratamento dos dados pessoais envolvidos deve observar integralmente a Lei Geral de Proteção de Dados Pessoais (LGPD), com governança transparente e mecanismos de contestação pelos próprios estudantes, conforme já indicado por Slade e Prinsloo (2013) quanto ao uso ético de analytics em contextos educacionais.

Potencial de replicação

Embora o modelo tenha sido desenvolvido e validado exclusivamente para os dados do Campus Chapecó, o pipeline proposto possui potencial de replicação em outras instituições de ensino superior, públicas ou privadas. Sua aplicação requer novo processo de treinamento e validação com dados próprios de cada instituição, uma vez que os padrões de evasão podem variar conforme o contexto acadêmico, regional e o perfil dos estudantes. Assim, a principal possibilidade de replicação está na utilização da metodologia e do pipeline desenvolvidos como referência para construção de modelos adaptados a diferentes realidades institucionais.

O quadro abaixo apresenta os modelos avaliados, seus respectivos *pipelines* de pré-processamento e os hiperparâmetros utilizados para replicação.

Modelo	*Pipeline* de pré-processamento e hiperparâmetros
Regressão Logística	SimpleImputer(median) → StandardScaler → LogisticRegression(C=0.5, class_weight='balanced', max_iter=1000, random_state=42)
Árvore de Decisão	SimpleImputer(median) → DecisionTreeClassifier(max_depth=5, min_samples_leaf=100, min_samples_split=200, ccp_alpha=0.0001, class_weight='balanced', random_state=42)
Random Forest	SimpleImputer(median) → RandomForestClassifier(n_estimators=200, max_depth=7, min_samples_leaf=50, min_samples_split=100, max_samples=0.8, oob_score=True, class_weight='balanced', n_jobs=-1, random_state=42)
XGBoost	SimpleImputer(median) → XGBClassifier(n_estimators=200, max_depth=4, learning_rate=0.1, min_child_weight=5, subsample=0.8, colsample_bytree=0.8, reg_lambda=2.0, scale_pos_weight=dinâmico, n_jobs=-1, random_state=42, eval_metric='logloss')
SVM (Linear)	SimpleImputer(median) → StandardScaler → CalibratedClassifierCV(LinearSVC(C=0.5, class_weight='balanced', max_iter=2000, random_state=42), cv=3, method='sigmoid')
KNN	SimpleImputer(median) → StandardScaler → KNeighborsClassifier(n_neighbors=31, weights='uniform', metric='manhattan', n_jobs=-1)

AUTORIA DA PROPOSTA

Discente

Jean Carlo Rodio
Servidor Técnico-Administrativo em Educação (TAE) da Universidade Federal da Fronteira Sul (UFFS) campus Chapecó e mestrando do Mestrado Profissional em Administração Pública em Rede Nacional (PROFIAP), campus Chapecó. Contato: jean.rodio@uffs.edu.br

Orientador

Prof. Dr. Moacir Francisco Deimling
Docente do Mestrado Profissional em Administração Pública em Rede Nacional (PROFIAP), Universidade Federal da Fronteira Sul (UFFS), campus Chapecó.

Coorientador

Prof. Dr. Giancarlo Dondoni Salton
Docente na Universidade Federal da Fronteira Sul (UFFS), campus Chapecó.

Data de finalização do documento: [20/07/2026]

REFERÊNCIAS

CHAPMAN, P.; CLINTON, J.; KERBER, R. et al. **CRISP-DM 1.0 Step-by-step data mining guide**. The CRISP-DM Consortium, 2000.

CHEN, T.; GUESTRIN, C. XGBoost: a Scalable Tree Boosting System. **Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16**, v. 1, n. 1, p. 785-794, 2016. Disponível em: 10.1145/2939672.2939785. Acesso em: 16 dez. 2025.

SLADE, S.; PRINSLOO, P. Learning Analytics Ethical Issues and Dilemmas. **American Behavioral Scientist**, v. 57, n. 10, p. 1510-1529, 2013. Disponível em: http://doi.org/10.1177/0002764213479366. Acesso em: 13 fev. 2026.

TCU. **Acórdão no 1868/2024 – TCU – Plenário**, 2024. Brasília, DF: Tribunal de Contas da União. Disponível em: https://pesquisa.apps.tcu.gov.br/documento/acordao-completo/*/NUMACORDAO%253A1868%2520ANOACORDAO%253A2024/DTRELEVANCIA%2520desc%252C%2520NUMACORDAOINT%2520desc/0. Acesso em: 7 nov. 2025.

UFFS. Universidade Federal da Fronteira Sul. **Relatório de Auditoria Operacional no 02/AUDIN/UFFS/2025:** evasão nos cursos de graduação da Universidade Federal da Fronteira Sul. Chapecó: UFFS, 2025. Disponível em: https://www.uffs.edu.br/uffs/conteudo/AUDITORIA%20INTERNA%20-%20AUDIN/RA%2002-AUDIN-UFFS-2025.pdf. Acesso em: 4 set. 2025.

PROTOCOLO DE ENTREGA DA PROPOSTA AOS PRATICANTES

A
Pró-Reitoria de Graduação (PROGRAD)
Universidade Federal da Fronteira Sul (UFFS)

Pelo presente, encaminhamos o produto técnico-tecnológico intitulado **"Protocolo de Implementação de um Sistema de Alerta Precoce de Evasão Acadêmica na UFFS Campus Chapecó"**, derivado da dissertação de mestrado "Aplicação de Técnicas de Mineração de Dados e Machine Learning para Predição da Evasão de Estudantes nos Cursos de Graduação de uma IFES", de autoria de Jean Carlo Rodio.

Os documentos citados foram desenvolvidos no âmbito do Mestrado Profissional em Administração Pública em Rede Nacional (Profiap), instituição associada Universidade Federal da Fronteira Sul (UFFS).

A solução técnico-tecnológica é apresentada sob a forma de relatório técnico conclusivo e seu propósito é subsidiar, por meio da geração periódica de estimativas de risco de evasão, a identificação precoce de estudantes em risco de abandono nos cursos de graduação do campus Chapecó da UFFS.

Solicitamos, por gentileza, que ações voltadas à implementação desta proposição sejam informadas à Coordenação Local do Profiap, por meio do endereço **sec.profiap@uffs.edu.br**

Chapecó, SC, 06 de agosto de 2026.

Registro de recebimento

ICP Brasil
Documento assinado digitalmente
PEDRO ADALBERTO AGUIAR CASTRO
Data: 06/08/2026 13:48:40-0300
Verifique em https://validar.iti.gov.br

Pedro Adalberto Aguiar Castro
Diretor de Registro Acadêmico
Pró-Reitoria de Graduação - UFFS

Discente: Jean Carlo Rodio, mestrando do PROFIAP
Orientador: Dr. Moacir Francisco Deimling, Professor da UFFS
Coorientador: Dr. Giancarlo Dondoni Salton, Professor da UFFS
UFFS - Universidade Federal da Fronteira Sul
Data de finalização do documento (20/07/2026)

PROFIAP
MESTRADO PROFISSIONAL EM ADMINISTRAÇÃO PÚBLICA

ATRIBUTOS QUALIFICADORES DO PTT

IMPACTO

A implementação do protocolo tem potencial de impacto direto sobre a gestão da permanência estudantil na UFFS, ao oferecer um instrumento preditivo validado (AUC-ROC = 0,8878), capaz de identificar corretamente cerca de 80% dos estudantes que efetivamente evadem (recall macro = 0,8038), com base em dados já disponíveis nos sistemas institucionais. O impacto potencial mais direto é a possibilidade de priorizar, entre milhares de matrículas ativas, os casos que demandam atenção das coordenações de curso e das equipes de assistência estudantil, tarefa inviável por meio de análise humana direta em razão do volume de dados envolvido.

EVIDENCIAÇÃO DE IMPLEMENTAÇÃO E IMPACTO DA PROPOSTA

A implementação da solução permitirá que a gestão acadêmica do campus Chapecó da UFFS identifique, de forma periódica, estudantes com maior risco de evasão. As estimativas geradas pelo modelo poderão subsidiar o planejamento de ações preventivas, como contato individualizado, orientação acadêmica, encaminhamento aos setores de assistência estudantil e acompanhamento das dificuldades acadêmicas e socioeconômicas identificadas.

O impacto potencial da proposta consiste no aprimoramento da capacidade institucional de atuação preventiva, possibilitando que as equipes responsáveis direcionem esforços e recursos aos estudantes que apresentam maior necessidade de acompanhamento. A solução também poderá contribuir para a sistematização das informações, o fortalecimento da tomada de decisão baseada em dados e o aperfeiçoamento das políticas institucionais de permanência estudantil.

Após sua implementação, os resultados poderão ser acompanhados por meio de indicadores como: número de estudantes identificados pelo modelo, quantidade de estudantes atendidos, ações de acompanhamento realizadas, permanência dos estudantes sinalizados e evolução das taxas de evasão nos cursos de graduação. Essas informações permitirão avaliar periodicamente a efetividade da solução e realizar os ajustes necessários.